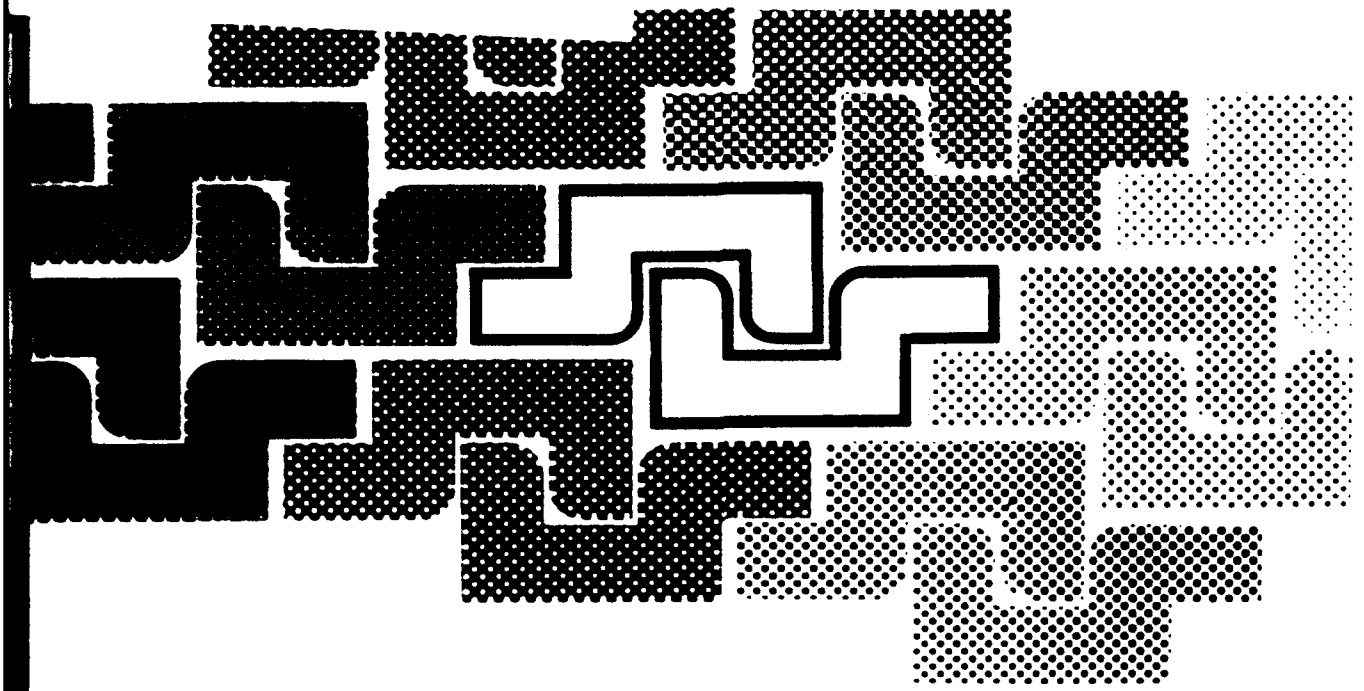
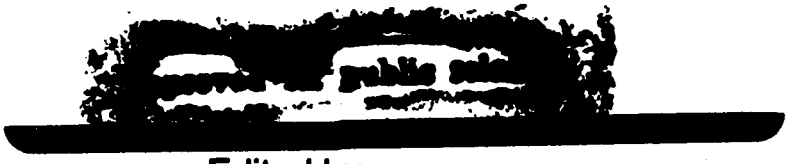


Shape in Picture

Mathematical Description of Shape in Grey-level Images



Edited by
Ying-Lie O Alexander Toet David Foster
Henk J.A.M. Heijmans Peter Meer

NATO ASI Series

Series F: Computer and Systems Sciences, Vol. 126

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	7-11 Sep 92	Final Proceedings 7-11 Sep 92

4. TITLE AND SUBTITLE
Shape in Pictures

5. FUNDING NUMBERS
CSP-92-1017

6. AUTHOR(S)
Dr Alex Toet

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NATO on Shape in Pictures
Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Sponsoring Agency:
European Office of Aerospace Research & Development
PSC 802 Box 14
FPO AE 09499-0200

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release. Distribution unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (Maximum 200 words)
This report is the Final Proceedings of the Conference. It contains all of the presentations.

14. SUBJECT TERMS

15. NUMBER OF PAGES
608

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

Shape in Picture

Mathematical Description of Shape in Grey-level Images

NATO ASI Series

Advanced Science Institutes Series

A series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The Series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A Life Sciences B Physics	Plenum Publishing Corporation London and New York
C Mathematical and Physical Sciences D Behavioural and Social Sciences E Applied Sciences	Kluwer Academic Publishers Dordrecht, Boston and London
F Computer and Systems Sciences G Ecological Sciences H Cell Biology I Global Environmental Change	Springer-Verlag Berlin Heidelberg New York London Paris Tokyo Hong Kong Barcelona Budapest

NATO-PCO DATABASE

The electronic index to the NATO ASI Series provides full bibliographical references (with keywords and/or abstracts) to more than 30000 contributions from international scientists published in all sections of the NATO ASI Series. Access to the NATO-PCO DATABASE compiled by the NATO Publication Coordination Office is possible in two ways:

- via online FILE 128 (NATO-PCO DATABASE) hosted by ESRIN, Via Galileo Galilei, I-00044 Frascati, Italy.
- via CD-ROM "NATO Science & Technology Disk" with user-friendly retrieval software in English, French and German (© WTV GmbH and DATAWARE Technologies Inc. 1992).

The CD-ROM can be ordered through any member of the Board of Publishers or through NATO-PCO, Overijse, Belgium.

Series F: Computer and Systems Sciences Vol. 126

Shape in Picture

Mathematical Description of Shape in Grey-level Images

Edited by

Ying-Lie O

Centre for Mathematics and Computer Science (CWI)
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

Alexander Toet

Netherlands Organization for Applied Scientific Research
Institute for Human Factors (TNO-IZF)
Kampweg 5, 3769 DE Soesterberg, The Netherlands

David Foster

University of Keele
Department of Communication and Neuroscience
Keele, Staffordshire ST5 5BG, UK

Henk J.A.M. Heijmans

Centre for Mathematics and Computer Science (CWI)
Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

Peter Meer

Rutgers University
Department of Electrical and Computer Engineering
Piscataway, NJ 08855-0909, USA

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo
Hong Kong Barcelona Budapest
Published in cooperation with NATO Scientific Affairs Division

Proceedings of the NATO Advanced Research Workshop "Shape in Picture", held at Driebergen, The Netherlands, September 7–11, 1992

CR Subject Classification (1991): I.3-5

ISBN 3-540-57578-2 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-57578-2 Springer-Verlag New York Berlin Heidelberg

CIP data applied for.

Typesetting: Camera ready by authors
45/3140 - 5 4 3 2 1 0 - Printed on acid-free paper

Preface

The fields of image analysis, computer vision, and artificial intelligence all make use of descriptions of shape in grey-level images. Most existing algorithms for the automatic recognition and classification of particular shapes have been developed for specific purposes, with the result that these methods are often restricted in their application. The use of advanced and theoretically well-founded mathematical methods should lead to the construction of robust shape descriptors having more general application.

Shape description can be regarded as a meeting point of vision research, mathematics, computing science, and the application fields of image analysis, computer vision, and artificial intelligence. The NATO Advanced Research Workshop "Shape in Picture" was organised with a twofold objective: first, it should provide all participants with an overview of relevant developments in these different disciplines; second, it should stimulate researchers to exchange original results and ideas across the boundaries of these disciplines.

This book comprises a widely drawn selection of papers presented at the workshop, and many contributions have been revised to reflect further progress in the field. The focus of this collection is on mathematical approaches to the construction of shape descriptions from grey-level images. The book is divided into five parts, each devoted to a different discipline. Each part contains papers that have tutorial sections; these are intended to assist the reader in becoming acquainted with the variety of approaches to the problem. It is hoped that the collection may thus be useful as a reference work and as a graduate text.

The editors would like to thank all who contributed to the production of the proceedings and to the workshop. More specifically, the editors are grateful to the authors for their essential contributions, to the numerous reviewers for their constructive comments, to the English-language editorial assistant for her precise corrections to the text, and to the participants for making the workshop successful. In particular, the editors wish to express their sincere gratitude to the NATO Scientific Affairs Division and The United States Air Force European Office of Aerospace Research and Development (EOARD) for their generous support, and to the co-sponsors for making this undertaking possible.

The editors have, as far as possible, tried to minimise errors and omissions in this mathematically oriented work. Nevertheless, they recognize the magnitude of the problem and apologize in advance for any failings.

October 1993

O Ying-Lie
Alexander Toet
David H. Foster
Henk J.A.M. Heijmans
Peter Meer

Director
Alexander Toet

Organising Committee
O Ying-Lie
David H. Foster
Henk J.A.M. Heijmans
Peter Meer

Sponsors
NATO Scientific Affairs Division
The United States Air Force European Office of Aerospace Research and Development (EOARD)

Local co-sponsors
Netherlands Organisation for Applied Scientific Research, Institute for Human Factors (TNO-IZF)
Centre for Mathematics and Computer Science (CWI)
The Foundation for Computer Science in The Netherlands (SION)

Local Organizing Committee
Peter Nacken
Frank Kooi
Antoine de Reus

English-language Editorial Assistant
Kay B. Merifield

Correspondence should be addressed to Alexander Toet.

Table of Contents

Theory of Shape Perception

Part 4 Symbolic Representation

Shape Primitives

Hierarchical Representation

Introduction

In principle, shape is that quality of an object that depends on the relative positions of all the points comprising its outline or external surface. In practice, any description of shape should reflect those attributes or features which are relevant for the intended purpose. These features may in turn be described by symbols through an appropriate function or mapping. A hierarchical shape representation may then be obtained by defining an order relation on the describing symbols, which thus allows an analysis of shape at different levels of "resolution".

A suitable shape description might therefore have the following properties:

- it is based on an underlying (continuous or discrete) topology;
- it is (semi-) continuous with respect to the topology;
- it is local, which means that the domain of influence must be restricted;
- it is invariant under certain transformations, such as translations or rotations;
- it is symbolic; that is, features can be described by symbols;
- it is hierarchical, in the sense that there exists an order relation on the describing symbols;
- it is compatible with changes in scale of both the domain and the grey-level.

In the light of these requirements, five main subject areas are considered, each forming a separate part of the book. The ordering of the parts reflects the conceptual development of shape description; that is, mathematical background, local extraction, theory of shape, symbolic representation, and evolutionary systems.

Part 1, *Mathematical Background*, introduces fundamental formal theories, *topology and geometry* and *categorical shape theory*. Topological and geometrical concepts are bound to play an important role in shape description. Although topology is a rich theory for continuous spaces, only very recently have consistent theories on topology for discrete spaces been proposed. Category theory arose from algebraic topology and provides an abstraction of structures and structure-preserving mappings in mathematics. It has found numerous applications in the field of computing science, as well as being used more recently to develop an abstract approach to shape theory.

Part 2, *Local Extraction*, presents mathematical methods that are capable of yielding relevant features for shape description. Two major topics are *mathematical morphology* and *wavelets*. In the last decade mathematical morphology has developed into an important method in the fields of image processing and computer vision. The basic idea underlying mathematical morphology, namely, to analyse the structure of an image by probing it with small test shapes, makes it eminently suited for shape description. Wavelet transformations involve decompositions of functions at different scales and positions. Recently, the theory of wavelets and its relationship to other methods based on fractal descriptions has become well established.

Part 3, *Theory of Shape*, considers aspects of *differential geometry* and the *theory of shape perception*. Differential-geometric methods are introduced here mainly to deal with the problems of the definition and estimation of differential features, and of how the invariance properties of these quantities may be investigated. Useful extensions to classical differential geometry are, in particular, discrete differential geometry and fuzzy geometry. The theory of shape perception deals with biologically and psychophysically oriented interpretations, in which invariance properties are again important. The outcome of this work may be useful in determining which processes of shape recognition and classification by human operators are relevant to more general methods of shape description.

Part 4, *Symbolic Representation*, focuses on *shape primitives* and *hierarchical representation*. It considers the formation of structured features from single features, and the representation of these features by symbols. Shape primitives are combinations of features that can be regarded as fundamental parts of shapes; hierarchical representation, on the other hand, concentrates on nested grouping of these primitives, forming layered representations. Most shape primitives are based on geometric descriptions of planar figures, and the extension of these methods to grey-level images is not always straightforward. A commonly used hierarchical representation is the pyramid, but a more general approach is based on graph theory, employing layered graphs. As a result of the latter, symbols in a single layer as well as symbols in different layers may be related to each other.

Part 5, *Evolutionary Systems*, comprises *evolutionary representation* and *multiscale description*. An evolutionary system depends continuously on a scale parameter. A well-known evolutionary representation is scale-space, derived from the linear isotropic diffusion equation; recently other types of equations have been considered. Multiscale description deals with the behaviour of shape descriptions in an evolutionary system as a function of a scale parameter.

A general mathematical description of shape necessarily draws upon ideas in all five subject areas. As a consequence, subject areas are interrelated: some papers specifically deal with these relationships; others do not fall exclusively into any one of the five areas.

Although a wide range of topics has been addressed, this book concentrates on mathematical issues. More computationally oriented aspects, for instance, have not been explicitly considered. With this caveat, the present collection is intended to be as complete as possible.

The Khalimsky Line as a Foundation for Digital Topology

*Ralph Kopperman**

Department of Mathematics, City College of New York, New York NY 10031, USA

Abstract. An object is defined from which digital spaces can be built. It combines the "one-dimensional connectedness" of intervals of reals with a "point-by-point" quality necessary for constructing algorithms, and thus serves as a foundation for digital topology. Ideas expressed in quotation marks here are given precise meanings. This study considers the *Khalimsky line*, that is, the integers, equipped with the topology in which a set is open iff whenever it contains an even integer, it also contains its adjacent integers. It is shown that this space and its interval subspaces are those satisfying the conditions mentioned previously. The Khalimsky line is used to study digital connectedness and homotopy.

Keywords: digital topology, general topology, connected ordered topological space (COTS), Alexandroff space, specialization order, Khalimsky line, digital n-space, digital Jordan n-surface, digital homotopy.

1 Introduction

A central problem of image processing is to represent regions in "continuous" Euclidean n-space, $\mathbb{R}^n$ on a finite computer screen which is manipulated using step-by-step programs. The most common approach to this problem is to approximate a portion of Euclidean space with an adjacency graph standing in for the nearness relationship between points; particularly popular instances of such graphs are the 2-dimensional (4,8), (8,4) and (6,6) adjacencies. A classical reference to such methods is [16]; [10] surveys a portion of the plentiful literature in this area.

Here we discuss a second approach: $\mathbb{R}^n$ is the product of n copies of the one-dimensional reals, and bounded portions of it may be viewed as being embedded in $[a,b]^n$ for some closed interval $[a,b] \subseteq \mathbb{R}$. Using this, we reduce the problem to finding a "one-dimensional discrete space" which combines virtues of the reals with useful properties of "discreteness". With this approach, the Jordan curve theorem holds in the 2-dimensional case, and the Jordan surface theorem holds in

* The author wishes to acknowledge comments by David Foster, Paul Meyer, and two unknown referees, which led to substantial improvements in this paper.

each dimension. Further, many of the corresponding known graph-based results can be derived from these results. Some of the literature using this and related approaches is discussed in [6-8] and [10].

2 Notation and Basic Concepts of General Topology

Except as noted, we include here only those definitions of general topology which can be found in standard textbooks on the subject, and some rudimentary discussion. Those acquainted with all of them may find a quick look at this section useful to familiarise themselves with the notation.

The reader can find further discussion of these concepts in any text on general topology. Among these texts, [17] is particularly easy to read; a text which is difficult to obtain but easy to read, and which spends some time on the finite topological spaces discussed here, is [14].

Recall that the *intersection* $\bigcap \mathcal{S}$ of a set $\mathcal{S}$ of subsets of a fixed space X (considered as a universe) is the collection of elements of X common to all elements of $\mathcal{S}$, its *union* $\bigcup \mathcal{S}$ is those elements in some element of $\mathcal{S}$.

Definition 1. A *topology* on a set X is a collection $\mathcal{T}$ of subsets of X, called the *open* sets, with the properties:
($\cap$) if $\mathcal{F} \subseteq \mathcal{T}$ is *finite* then $\bigcap \mathcal{F} \in \mathcal{T}$,
($\bigcup$) if $\mathcal{S} \subseteq \mathcal{T}$ then $\bigcup \mathcal{S} \in \mathcal{T}$.

That is, finite intersections and arbitrary unions of open sets are open (and as a result of this, $X = \bigcap \emptyset \in \mathcal{T}$ and $\emptyset = \bigcup \emptyset \in \mathcal{T}$). A $C \subseteq X$ is *closed* if its complement $X - C$, is open ($X - C \in \mathcal{T}$). We often denote $(X, \mathcal{T})$ by X, and call the pair a *topological space.*

The intersection of all closed sets containing a given subset Y is necessarily the smallest closed set containing Y, denoted $cl(Y)$; similarly there is a largest open set contained in Y, denoted $int(Y)$.

Definition 2. For topological spaces, a map $f : X \to Y$ is *continuous* if $f^{-1}[T]$ is open for each open T in Y (here, as usual, for $A \subseteq Y$, $f^{-1}[A] = \{x \mid f(x) \in A\}$, and for $A \subseteq X$, $f[A] = \{f(x) \mid x \in A\}$). It is *open* if $f[T]$ is open for each open T in X. Further, f is an *quotient map* if it is onto, and a subset of the range is open if and only if its inverse image is open; f is a *homeomorphism* if it is a one-to-one open quotient.

We are interested below only in open quotient maps. It is routine to see that a map is an open quotient iff it is open, onto and continuous. A homeomorphism is often given the equivalent definition: it is one-to-one, onto, and both it and its inverse are continuous.

Here are two standard equivalents to continuity which we use below: $f : X \to Y$ is continuous
$\Leftrightarrow f^{-1}[C]$ is closed for each closed C,
$\Leftrightarrow f[cl(A)] \subseteq cl(f[A])$ for each $A \subseteq X$.

Definition 3. By definition of topology, X and $\emptyset$ are *clopen*: simultaneously closed and open. A space $(X, \mathcal{T})$ is *connected* if its only clopen subsets are X and $\emptyset$. For $Y \subseteq X$, the *subspace topology* on Y resulting from $\mathcal{T}$ is $\{T \cap Y \mid T \in \mathcal{T}\}$ (which is easily shown to be a topology).

By the definitions of connectedness and the subspace topology, $Y \subseteq X$ is connected in the subspace topology iff there are no open sets $T, U \in \mathcal{T}$ such that $T \cap Y$ is neither $\emptyset$ nor Y, and $Y - T \cap Y = U \cap Y$. A useful equivalent way to say this is that $Y \subseteq X$ is connected in the subspace topology iff there are no open sets $T, U \in \mathcal{T}$ such that $T \cap Y, U \cap Y \neq \emptyset, T \cap U \cap Y = \emptyset$, and $Y \subseteq T \cup U$.

Recall that an *image point* of a map $f : X \to Y$ is a point of the form $f(x)$, while the *image* of f is $f[X]$, and a *continuous image* is the image of a continuous function. Further, notice that the continuous image of a connected space is connected. (For if not, we can find a clopen T in $f[X]$ containing some $f(x)$ but not another $f(y)$; then $f^{-1}[T]$ is clopen and neither X nor $\emptyset$.)

Definition 4. A connected subspace is a *(connected) component* of a topological space if it is not a proper subset of another connected subspace.

It can be shown that each connected subspace of a topological space is a subset of a component. Thus if two points are in any connected subset of a space then they are in the same component of it. In particular, if they are image points of a map from a connected space into the given space, they are in the same component.

Definition 5. For any set $\mathcal{G}$ of subsets of X, the topology *generated by* $\mathcal{G}$ is $\mathcal{T}_{\mathcal{G}} = \{T \mid \text{if } x \in T, \text{ there is a finite } \mathcal{F} \subseteq \mathcal{G} \text{ for which } x \in \bigcap \mathcal{F} \subseteq T\}$. $\mathcal{T}_{\mathcal{G}}$ is easily seen to be the smallest topology on X which contains $\mathcal{G}$. Further, if whenever $x \in G \cap G', G, G' \in \mathcal{G}$, there is an $H \in \mathcal{G}$ such that $x \in H \subseteq G \cap G'$, then $T \in \mathcal{T}_{\mathcal{G}}$ iff for each $x \in T$ there is an $H \in \mathcal{G}$ such that $x \in H \subseteq T$, and in this case, $\mathcal{G}$ is a *base* for $\mathcal{T}_{\mathcal{G}}$.

Definition 6. Given an indexed collection of topological spaces, $< (X_i, \mathcal{T}_i) \mid i \in I >$, their *Cartesian product* is $(\prod_I X_i, \prod_I \mathcal{T}_i)$, where $\prod_I X_i = \{x : I \to \bigcup\{X_i \mid i \in I\} \mid \text{for each } i \in I,\ x(i) \in X_i\}$, and $\prod_I \mathcal{T}_i$ is the topology generated by all sets of the form $T^* = \{x \in \prod_I X_i \mid x(j) \in T\}$, for $j \in I$ and $T \in \mathcal{T}_j$.

For finite I, a base for $\prod_I \mathcal{T}_i$ is $\prod_I T_i, T_i \in \mathcal{T}_i$. In the last section of this paper, we need some concepts related to compactness:

Definition 7. A topological space X is *compact* if whenever $X = \bigcup \mathcal{S}$ for some set $\mathcal{S}$ of open sets, then there is a finite $\mathcal{F} \subseteq \mathcal{S}$ for which $X = \bigcup \mathcal{F}$. It is *locally compact* if whenever $x \in T$, T open, there are U, C such that $x \in U \subseteq C \subseteq T$, U is open, and C is compact (with respect to the subspace topology).

Nomenclature is not completely standard in this area. Many texts reserve the term "compact" for those spaces satisfying the first sentence of Definition 7 which are also *Hausdorff* (if $x \neq y$, there are open, disjoint T, U such that $x \in T, y \in U$). The key spaces considered below are *not* Hausdorff.

3 Alexandroff Discreteness and the Khalimsky Line

Most of what follows can be found in [3-8], and [11]. In some cases, we prove results also shown in those papers, for convenience, and to give examples of natural methods of working with these spaces.

Below we need a central property of intervals of real numbers:

Definition 8. A *connected ordered topological space (COTS)*, is a connected topological space X such that:

if $Y \subseteq X$ contains at least three distinct points, then there is a $y \in Y$ such that $Y - \{y\}$ meets more than one component of $X - \{y\}$.

The fact that intervals of reals are COTS comes from the following well-known characterization of connected sets of reals:

Proposition 9. *The connected components of a set of real numbers are the maximal intervals it contains. A set of real numbers is connected iff it is an interval.*

By way of contrast, the deletion of a point from Euclidean n-space, $n > 1$, leaves a connected set (which thus does not have more than one component). At the end of this section we give the examples of COTS which will of interest. These spaces were systematically studied by Khalimsky, beginning in the late 1960s, and are discussed in [3] and [7]. They were independently rediscovered by Kovalevsky in [12, 13]. The "one-dimensional" nature of a COTS is emphasized by the following result of [3] (notice in its statement the notations $\uparrow, \downarrow$, which are used throughout the rest of the paper):

Theorem 10. *Each COTS X admits a total order $<$ such that for each $x \in X$ the components of $X - \{x\}$ are $\downarrow(x) = \{y \mid y < x\}$ and $\uparrow(x) = \{y \mid y > x\}$.*

The second key idea of "discreteness" was introduced even earlier, in 1937, by another Russian, Alexandroff in [1]. He even called the concept in the next definition "discrete", but we follow all modern topologists by calling the *discrete* topology that in which *all* sets are open. We also need the *indiscrete* topology: that in which only X and $\emptyset$ are open.

Definition 11. A topological space is *Alexandroff* if arbitrary intersections of open sets are open.

Thus, an Alexandroff space is one for which the law $(\cap)$ in the definition of topological space is replaced by the stronger: $(\bigcap)$ if $\mathcal{S} \subseteq \mathcal{T}$ then $\bigcap \mathcal{S} \in \mathcal{T}$.

It is useful to note that a space is Alexandroff iff $cl(\bigcup \mathcal{S}) = \bigcup\{cl[S] \mid S \in \mathcal{S}\}$. This holds because, by the de Morgan laws, a space is Alexandroff iff arbitrary unions of closed sets are closed. In particular, in an Alexandroff space $cl(Y) = \bigcup\{cl(x) \mid x \in Y\}$. Also, any subspace of an Alexandroff space is Alexandroff.

No Euclidean space is Alexandroff, since $\{x\} = \bigcap_1^\infty B_{1/n}(x)$ is non-open where $B_r(x)$ denotes the open ball $\{y \mid \|y - x\| < r\}$. On the other hand, each finite topological space is Alexandroff, as is each *locally finite* space: one in which

each element is contained in a finite open set and a finite closed set. In fact, by Lemma 12, we only need that each element be in a finite open set, for then it is in an open set with a smallest number of elements, which is necessarily minimal.

For any subset, Y, of any topological space, define $n(Y)$ to be the intersection of all open subsets containing Y (and $n(y) = n(\{y\})$). Note the analogy with the fact that $cl(Y)$ is the intersection of all closed subsets containing Y, but the fact that arbitrary intersections of closed sets is closed leads to much of the importance of this set. It is of passing interest to us that $n(Y)$ is used in computer science even when it is not open: it is the *saturation* of Y, important in the theory of continuous lattices (for more on this subject, see [2]).

But this leads to a useful characterization of Alexandroff spaces:

Lemma 12. *A topological space is Alexandroff iff each element, x, is in a smallest open set, and this set is $n(x)$.*

Proof. For Alexandroff spaces, $n(Y)$ is open, and must thus be the smallest open subset containing Y. Conversely, if each x is in a minimal open set, then this open set is necessarily $n(x)$, and T is open iff for each $x \in T$, $n(x) \subseteq T$; since an arbitrary intersection of sets containing $n(x)$ contains $n(x)$, this characterization shows that an arbitrary intersection of open sets is open. □

Notice that this last sentence showed that these minimal open sets form a base for the topology. Here is the central example of an Alexandroff COTS:

Definition 13. The *Khalimsky line* is the integers, $\mathbb{Z}$, equipped with the topology $\mathcal{K}$ generated by $\mathcal{G}_{\mathcal{K}} = \{\{2n-1, 2n, 2n+1\} \mid n \in \mathbb{Z}\}$. Two integers x, y are *adjacent* if $|x - y| = 1$. A subset I of $\mathbb{Z}$ is an *interval (of integers)* if whenever $x, y \in$ I and $x < z < y$, then $z \in$ I.

Proposition 14. *A subset of $\mathbb{Z}$ is open iff whenever it contains an even integer, it also contains its adjacent integers. It is closed iff whenever it contains an odd integer, it also contains its adjacent integers.*

Proof. Suppose $A \subseteq \mathbb{Z}$ is open and $2n \in A$. Then for some finite $\mathcal{F} \subseteq \mathcal{G}_{\mathcal{K}}$, $2n \in \bigcap \mathcal{F} \subseteq A$. But since the only element of $\mathcal{G}_{\mathcal{K}}$ containing $2n$ is $\{2n-1, 2n, 2n+1\}$, we must have $\mathcal{F} = \{\{2n-1, 2n, 2n+1\}\}$, so $\bigcap \mathcal{F} = \{2n-1, 2n, 2n+1\} \subseteq A$; this shows that the elements adjacent to $2n$ are in A.

Conversely, assume that if A contains an even integer, it also contains its adjacent integers. If $x = 2n \in A$ then, setting $\mathcal{F} = \{\{2n-1, 2n, 2n+1\}\}$, we have $x \in \bigcap \mathcal{F} = \{2n-1, 2n, 2n+1\} \subseteq A$; if $x = 2n+1 \in A$ then, setting $\mathcal{F} = \{\{2n-1, 2n, 2n+1\}, \{2(n+1)-1, 2(n+1), 2(n+1)+1\}\}$, we have $x \in \bigcap \mathcal{F} = \{2n+1\} \subseteq A$, so A is open in $\mathcal{K}$.

Now that this characterization of the Khalimsky-open sets has been established, we use it liberally throughout the rest of the proof and paper.

If $A \subseteq \mathbb{Z}$ is closed, and $x = 2n+1 \in A$, then $2n+1 \notin \mathbb{Z} - A$, an open set, so $2n, 2n+2 \notin \mathbb{Z} - A$ (since $2n+1$ is adjacent to each). Thus $2n, 2n+2 \in A$. Conversely, if whenever A contains an odd integer, then it contains the adjacent integers, then $\mathbb{Z} - A$ always contains the integers adjacent to each of its even elements, and is thus open, showing A to be closed. □

Corollary 15. *The connected components of a set of integers are the maximal intervals it contains. A set of integers is connected iff it is an interval.*

Proof. If I is an interval of integers and T is a clopen subset of I, which is neither I nor $\emptyset$, let $x \in T$, $y \notin T$, and assume $x < y$ (this involves no loss of generality, since otherwise replace T by the clopen $U = I - T$ in the argument). Proceeding by induction, using the fact that I is an interval, we can find a z between x and y so that $z \in T$, $z+1 \notin T$. But if z is even, this contradicts the openness of T since it does not contain the adjacent odd $z+1$ and if z is odd, the openness of U is similarly contradicted. If $A \subseteq \mathbb{Z}$ is not an interval, find $x < y < z$ such that $x, z \in A$ but $y \notin A$. If y is even, set $T = \downarrow(y)$, $U = \uparrow(y)$, and notice that T, U are open, $x \in T \cap A$, $y \in U \cap A$, $A \subseteq T \cup U$, and $A \cap T \cap U = \emptyset$; if y is odd, the same is true of $T = \downarrow(y+1)$, $U = \uparrow(y-1)$. Thus A is not connected. □

It is not difficult to show that the results in Proposition 9 and Corollary 15 hold for arbitrary COTS (with respect to the order given by Theorem 10). But here we reverse the process to show that $\mathbb{Z}$ is truly a COTS in the following theorem. The second assertion of the theorem is of theoretical importance, showing the central place of the Khalimsky line.

Theorem 16. *Each interval in the Khalimsky line is a locally finite COTS. A topological space is an Alexandroff COTS iff it is (homeomorphic to) an interval in $\mathbb{Z}$ or the indiscrete space with exactly two points.*

Proof. If x is an odd integer, then by Proposition 14, $\{x\}$ is open, and $\{x-1, x, x+1\}$ is closed. If x is even, then similarly, $\{x\}$ is closed and $\{x-1, x, x+1\}$ is open. Thus each x is contained in a finite open set and a finite closed set, so $\mathbb{Z}$ is locally finite.

Each interval I of $\mathbb{Z}$ is connected by Corollary 15 and if Y $\subseteq$ I contains 3 elements, it contains one, y, between two other elements of Y, $x < y < z$, $x, z \in$ Y. Again by Corollary 15, the components of $I - \{y\}$ are $\downarrow(y) \cap I$ and $\uparrow(y) \cap I$ and these both meet $Y - \{y\}$.

The proof of the second sentence must be postponed to the next section. □

Figure 1 represents part of the Khalimsky line. In it, odd numbers "look like" the open intervals between their adjacent numbers.

Fig. 1. A portion of the Khalimsky line

Here a set is open if it "looks open" to our eye, which is trained to think of the real line. The Khalimsky line is interpreted in it as an open quotient of the real line, via the map $k : \mathbb{R} \to \mathbb{Z}$ defined by: $k(x) = x$ if x is an even integer, $= y$ if $x \in (y-1, y+1), y$ an odd integer. It is routine to check that k is onto, continuous, and open.

4 The Alexandroff Specialisation Order and its Graph

Also in his 1937 paper, Alexandroff defined the following relation on an arbitrary topological space, the *(Alexandroff) specialization order*:

$$x \preceq y \text{ if } x \in cl(y).$$

A *preorder on X* is a *reflexive* (i. e. $x \leq x$) *transitive* ($x \leq z$ *and* $y \leq z \Rightarrow x \leq z$) relation on X; a *partial order on X* is a preorder on X which is also *antisymmetric* ($x \leq y$ *and* $y \leq x \Rightarrow x = y$). That $\preceq$ is reflexive is clear, and $\preceq$ is also transitive: assume $x \preceq y$ and $y \preceq z$. Then $y \in cl(z)$, so the latter is a closed set containing y, and thus is a superset of the smallest closed set containing y, $cl(y)$. But then $x \in cl(y) \subseteq cl(z)$ so $x \preceq z$.

Lemma 17. *For any topological space, $x \notin cl(y)$ iff there is an open T such that $x \in T$ and $y \notin T$. Thus, $x \preceq y$ iff $y \in n(x)$ iff $x \in cl(y)$.*

Proof. Indeed: $x \notin cl(A) \Leftrightarrow$ for some closed C, $A \subseteq C$ and $x \notin C$
$\Leftrightarrow$ for some open $T(= X - C)$, $A \cap T = \emptyset$ and $x \in T$.
But as a result, using contrapositives: $x \in cl(y)$
$\Leftrightarrow$ for every open $T \ni x, y \in T \Leftrightarrow y \in n(x)$. □

An often-ignored separation axiom is equivalent to the antisymmetry of $\preceq$, thus to the assertion that it is a partial order. A space X is T_0 if whenever $x \neq y$, then there is an open T containing exactly one of x, y. The fact that the two are equivalent is straightforward from the first assertion of Lemma 17.

The *specialization order* is itself often ignored, due to the fact that a stronger separation axiom which holds for spaces usually considered by topologists is equivalent to the assertion that $\preceq$ is equality: A space X is T_1 if whenever $x \neq y$, then there is an open T containing x but not y, and proof of this equivalence again uses the Lemma.

However, *all topological properties of an Alexandroff space are described by its specialization order, and the graph of this order is the vehicle for computer applications of these spaces.* Much of the rest of this paper is devoted to special cases of this principle.

Proposition 18. *If X is an Alexandroff space and $A \subseteq X$, then A is open iff $A = \preceq [A]$ $(= \{y \mid$ for some $x \in A, x \preceq y\})$, A is closed iff $A = \succeq [A]$. Further, for arbitrary $A \subseteq X$, $n(A) = \preceq [A]$ and $cl(A) = \succeq [A]$.*

Proof. Suppose first that A is open and $x \preceq y$. Then if $x \in A$, we must have $y \in n(x) \subseteq A$. By the arbitrary nature of x, y, this shows $A = \preceq [A]$.

Conversely, assume $A = \preceq [A]$, and $x \in A$. To show A open, it will suffice to show $n(x) \subseteq A$, but this holds since if $y \in n(x)$, we have $x \preceq y$, so $y \in A$.

The result for closed sets follows from the routinely shown fact that if $\leq$ is *any* binary relation on X, with inverse $\geq$, then $A = \leq [A]$ iff $X - A = \geq [X - A]$.

Finally, the results in the second sentence come from the fact that if $\leq$ is any partial order on X and $A \subseteq X$ then $\leq [A]$ is the smallest set containing A and closed under $\leq$. □

This theorem says that for Alexandroff spaces, we can tell which sets are open by simply checking the specialization order. Thus for these spaces, the specialization order tells us the topology. It should also tell us which functions are continuous:

Proposition 19. *For an Alexandroff space X, an $f : X \to Y$ is continuous iff it is specialization-preserving (i. e., whenever $x \preceq y$ then $f(x) \preceq f(y)$).*

Proof. Indeed, the fact that continuous maps are specialization-preserving holds for arbitrary topological spaces, for if f is continuous and $x \preceq y$ then $x \in cl(y)$ so $f(x) \in f[cl(y)] \subseteq cl(f(y))$. Conversely, suppose X is Alexandroff and f is specialization-preserving, and let $y \in f[cl(A)]$. Then for some $x \in cl(A)$, $y = f(x)$. Thus for some $z \in A, x \preceq z$ and $y = f(x) \preceq f(z)$, showing $y \in cl(f[A])$. □

Of particular interest in image processing is the issue of connectivity. This can also be settled using the specialization order graph. Two distinct elements of X, x, y are *adjacent* if $\{x, y\}$ is connected. We also let $\mathcal{A}(x) = \{y \mid \{x, y\}$ is connected and $x \neq y\}$, the set of points adjacent to x. Of course, $\mathcal{A}(x)$ depends on the space and if the context does not make the space clear, we may use a subscript. For example, if Y is a subspace of X, then notice that $\mathcal{A}_Y(x) = \mathcal{A}_X(x) \cap Y$. A *(digital) path in Y from x to y* is a sequence $x_0, \ldots, x_n$ of elements of Y such that $x = x_0, y = x_n$, and for each $i < n$, $\{x_i, x_{i+1}\}$ is connected. A set Y is *(digital) path-connected* if for each $x, y \in Y$ there is a path in Y from x to y. A *(digital) path-component of X* is a maximal path-connected set. For each element x of a topological space, $C_x = \{z \mid$ there is a path from x to $z\}$. Below we do not use the modifier "digital", but the reader should note that the usual meaning of the term "path" is the image of a continuous function whose domain is [0,1]; it is not difficult to show that a digital path is the image of a continuous function whose domain is a finite interval in $\mathbb{Z}$.

Lemma 20. *(a) In any topological space, $\{x, y\}$ is connected iff $x \preceq y$ or $y \preceq x$. Thus $\mathcal{A}(x) = (cl(x) \cup n(x)) - \{x\}$. Also, this notion of adjacency coincides with that in Definition 13 for $\mathbb{Z}$.*

(b) A subset of an Alexandroff space is connected iff it is path-connected. Its path-components are its components, and these are the sets of the form C_x. Further, the C_x are clopen.

(c) If $x \in X$, a connected Alexandroff space, then $\mathcal{A}(x)$ meets each component of $X - \{x\}$.

Proof. (a) Recall from Sect. 2 that Y is connected in the subspace topology iff there are no open sets $T, U \in \mathcal{T}$ such that $T \cap Y, U \cap Y \neq \emptyset$, $T \cap U \cap Y = \emptyset$, and $Y \subseteq T \cup U$. Thus in particular, $\{x, y\}$ is *not* connected iff there are open T, U such that $x \in T, y \in U$, $x \notin U$, and $y \notin T$. Thus, $\{x, y\}$ is *not* connected iff $x \notin cl(y)$ and $y \notin cl(x)$, or, by the contrapositive, $\{x, y\}$ is connected iff $x \preceq y$ or $y \preceq x$. This shows the first assertion, and the second results by applying the characterization of $\preceq$ in Lemma 17. For the last, if $x \in \mathbb{Z}$ is odd, then $(cl(x) \cup n(x)) - \{x\} = (\{x-1, x, x+1\} \cup \{x\}) - \{x\} = \{x-1, x+1\}$ and if $x \in \mathbb{Z}$ is even, then $(cl(x) \cup n(x)) - \{x\} = (\{x\} \cup \{x-1, x, x+1\}) - \{x\} = \{x-1, x+1\}$.

(b) Since subspaces of Alexandroff spaces are Alexandroff, it will suffice to show the above for the entire space, X. Thus, suppose X is path-connected, and, by way of contradiction let T be a clopen set, $x, y \in X$ be such that $x \in T, y \notin T$. Find a path in X from x to y, $x_0, \ldots, x_n$; since $x_0 \in T$ and $x_n \notin T$, the last i for which $x_i \in T$ has the following properties, which are contradictory:

$\{x_i, x_{i+1}\}$ is connected,

$x_i \in T$ and $x_{i+1} \notin T$, and T is clopen.

Conversely, suppose X is connected, but, by way of contradiction, not path-connected. Thus for some $x, y \in X$ there is no path from x to y.

C_x is open: if $z \in C_x$ and $z \preceq w$, then there is a path $x_0, \ldots, x_n$ from x to z; also since $x_n = z \preceq w$, $\{x_n, w\}$ is connected, thus $x_0, \ldots, x_n, w$ is a path from x to w, so $w \in C_x$. That C_x is closed is shown similarly, replacing $\preceq$ by $\succeq$. But then $C_x \neq \emptyset, X$, and is clopen, contradicting the connectedness of X.

C_x is path-connected, since if $y, z \in C_x$, there are paths $y_0, \ldots, y_n$, from x to y, $x_0, \ldots, x_m$ from x to z, so $y_n, y_{n-1}, \ldots, y_0, z_1, \ldots, z_m$ is one from y to z.

Since the C_xs are connected and no set properly containing one of them can be connected (both by previous parts of this proof), they are components, and since each $x \in X$ is in C_x, they are all the components.

(c) If $y \in X$ then there is a path $x_0, \ldots, x_n$ from x to y. Let j be least so that if $j \leq k$ then $x_k \neq x$, and notice that $\{x_j, \ldots, x_n\}$ is a connected subset of $X - \{x\}$ containing y, and must therefore be a subset of C_y. But also, by our choice, $x = x_{j-1} \neq x_j$, and $\{x_{j-1}, x_j\}$ is connected, so $x_j \in \mathcal{A}(x) \cap C_y$. □

As a result of Lemma 20 (a), $\mathcal{A}(x) = cl(x) - \{x\}$ for open x, and $\mathcal{A}(x) = n(x) - \{x\}$ for closed x. We are now ready to finish the proof of Theorem 16:

Proof. The reader should check that (because it has no subsets containing 3 distinct elements) the indiscrete two-point space is a COTS. However, we need the following result (2.8 of [3]): Each point in a COTS with at least three distinct points is either open or closed (but not both, by connectedness). Also, if $\{x, y\}$ is a connected subset of such a COTS, then exactly one of x, y is open, and the other is closed.

Now let X be an Alexandroff COTS. Notice that for each element $x \in X$, $n(x)$ has no more than 3 elements. For otherwise, let $w, y, z \in n(x)$ be distinct from x; we show that there is no point in $\{w, y, z\}$ whose deletion leaves the others in distinct components of the remainder. For without loss of generality, suppose y, z are in distinct components of $X - \{w\}$. Then there must be a clopen T containing y but not z. But if $x \in T$ then $z \in n(x) \subseteq T$, and if $x \notin T$ then $y \in n(x) \subseteq X - T$ (since both T and $X - T$ are open sets), and these are both contradictions. Similarly, $cl(x)$ has no more than 3 elements.

Thus let x be a closed point. We begin to define $f : X \to \mathbb{Z}$ recursively, with $f(0) = x$. Since $\{f(0)\}$ is closed, and thus not open, choose $f(1) \in n(f(0)) - \{f(0)\}$, i. e. so that $f(0) \preceq f(1)$ (just as $0 \preceq 1$). Assume that we have recursively defined a specialization-preserving one-to-one map f from $\{0, \ldots, 2n-1\}$ with the topology inherited from $\mathbb{Z}$, and:

if $cl(\{f(2n-1)\}) - \{f(2n-2), f(2n-1)\} \neq \emptyset$, choose $f(2n)$ in this set, and then,

if $n(\{f(2n)\}) - \{f(2n-1), f(2n)\} \neq \emptyset$, choose $f(2n+1)$ in this set.

As long as this choice is possible, f remains specialization-preserving, thus continuous. Also note that for odd or even k, we are choosing $f(k+1)$ to be the remaining element of $\mathcal{A}(\{f(k)\}) - \{f(k-1)\}$. As a result, f is also one-to-one, for if there is some $j < k-1$ such that $f(j) \in \mathcal{A}(f(k))$, then there is no element among the three distinct $\{f(j), f(k-1), f(k)\}$ whose deletion leaves the remaining two in separate components of the remainder, contradicting that X is a COTS.

Of course there may be a third element, z in $n(f(0))$; if so, an argument like that of the previous paragraph shows that it cannot be $f(k)$ for any positive k since it is always in the same component of $X - \{f(k-1)\}$ as is $f(0)$. Thus let $f(-1) = z$, and define f recursively on the negatives as it was defined on the positives above (of course, replacing $2n-1, 2n-2$, respectively, by $2n+1, 2n+2$). The result is a one-to-one $f : I \to X$ which is specialization-preserving, thus continuous for some interval I in $\mathbb{Z}$.

It remains to show that f^{-1} is continuous. By construction and the previous paragraph, if $j < k-1$ then $\{f(j), f(k)\}$ is never connected, so if $f(j) \preceq f(k)$ then j, k are adjacent; further, we have $f(j) \in cl(f(k))$, and by our choice, that could only happen if j were even, showing $j \preceq k$. □

As a result of Lemma 20, standard graph-based algorithms can be used to find the components of a finite topological space. For example, for a finite topological space $X = \{x_1, \ldots, x_n\}$, if $p(x_k)$ denotes the length of the shortest path from a fixed x_j to x_k, then the fact that:

if $(x_k \preceq x_i$ or $x_i \preceq x_k)$ and $p(x_i) = m$ and $p(x_k) \not\preceq m$, then $p(x_k) = m+1$,

leads to a simple algorithm to find $p(x_k)$, thus the connected components of X.

5 Digital n-space

It is not difficult to verify that if $Y_i \subseteq X_i$ for each $i \in I$ then $cl(\prod_I Y_i) = \prod_I cl(Y_i)$, and $n(\prod_I Y_i) = \prod_I n(Y_i)$. This has two important consequences:

for $x, y \in \prod_I X_i$, $x \preceq y$ iff for each $i \in I$, $x_i \preceq y_i$, and
the product of a finite number of Alexandroff spaces is Alexandroff.

Indeed, the first of these statements is merely another way of saying that $cl(\{x\}) = \prod_I cl(x_i)$, a special case of the above.

For the other: If $I = \{1, \ldots, n\}$, then by the above, for $x \in \prod_I X_i$, $n(x) = \prod_I n(x_i)$. The last sentence in the section on basic concepts implies that this is open and each basic open set containing x is of the form $\prod_1^n T_i$, with $x_i \in T_i$; it thus contains this product, showing that each open set containing x includes $\prod_1^n n_i(x_i)$.)

Of particular interest to us is *digital n-space*, the product of n copies of the Khalimsky line, $\mathbb{Z}^n$. Figure 2 shows digital 2-space (*the digital plane*) and digital 3-space. The former looks like graph paper, with "line crossings" emphasized. Its points are the dots at the line crossings, the line intervals between crossings, and the open boxes, and the 3-space is handled similarly. Notice that all the

coordinates of a point are odd iff that point is open, and that they are all even iff it is closed, thus the remaining points are neither open nor closed. We have called points whose coordinates are all even or all odd *pure*, and others *mixed* in [3] and elsewhere.

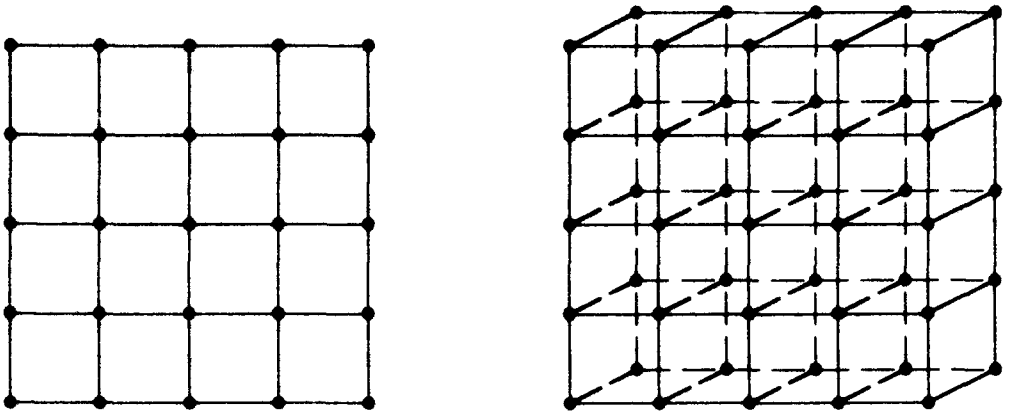


Fig. 2. Part of the digital plane Part of digital 3-space

This is our analogue of Euclidean n-space, the product of n copies of the real line, $\mathbb{R}^n$. The key difference, of course, is that $\mathbb{Z}^n$ can be analysed using the specialization order and resulting adjacency relation. Further, by the first paragraph of this section, these are easy to find; for example, if $x = (x_1, \ldots, x_n)$, with $x_1, \ldots, x_m$ odd and $x_{m+1}, \ldots, x_n$ even, then $n(x) = \{x_1 - 1, x_1, x_1 + 1\} \times \ldots \times \{x_m - 1, x_m, x_m + 1\} \times \{x_{m+1}\} \times \ldots \times \{x_n\}$, (and has 3^m elements) $cl(x) = \{x_1\} \times \ldots \times \{x_m\} \times \{x_{m+1} - 1, x_{m+1}, x_{m+1} + 1\} \times \ldots \times \{x_n - 1, x_n, x_n + 1\}$ (and has 3^{n-m} elements), while $\mathcal{A}(x) = (cl(x) \cup n(x)) - \{x\}$ and thus has $3^m + 3^{n-m} - 2$ elements. In particular, for pure points, $\mathcal{A}(x)$ contains the $3^n - 1$ elements $\neq x$, each of whose coordinates differ from the corresponding coordinate of x by at most 1 (and is thus the $3^n - 1$ adjacency with which computer scientists are familiar). For mixed points when $n = 2$, $\mathcal{A}(x)$ is the 4 points which differ from x by 1 on *one* coordinate, and is thus the familiar 4-adjacency. However, this differs from the (4,8) and (8,4) adjacency usually considered, since the adjacency graph about a point depends on the position of the point rather than on whether it lies on the foreground or background. In three dimensions, for mixed points $\mathcal{A}(x)$ contains 10 points (those which differ from x by 1 just on one fixed coordinate or agree on that coordinate but differ by 1 on at least one of the other two coordinates). Further, in dimensions over 3, not all mixed points have adjacency sets of the same size (for example, in dimension 4, points with 1 or 3 open coordinates have 28 adjacent points, while those with 2 have 16). Figure 3 shows typical such $\mathcal{A}(x)$s in dimensions 2 and 3.

Among the key similarities between $\mathbb{R}^n$ and $\mathbb{Z}^n$ is the Jordan Surface Theorem. For Alexandroff spaces, Jordan curves and surfaces can easily be defined:

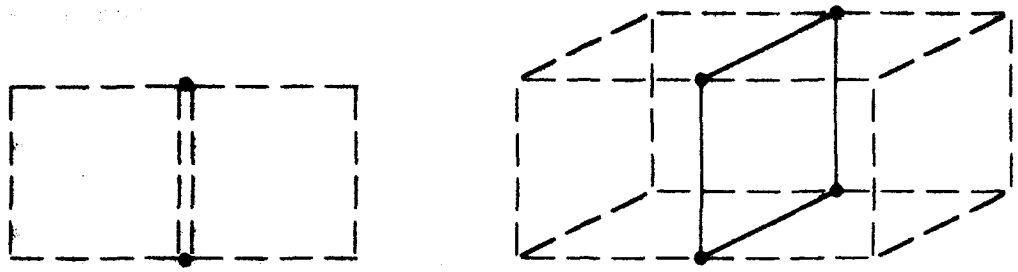


Fig. 3. $\mathcal{A}(2,1)$ $\mathcal{A}(1,2,1)$

Definition 21. A *digital Jordan curve* is a finite, nonempty, connected topological space J such that if $j \in J$ then $\mathcal{A}(j)$ is a two-point discrete space.

We then have:

Theorem 22. *If $J \subseteq \mathbb{Z}^2$ is a Jordan curve, then $\mathbb{Z}^2 - J$ consists of two connected components, one finite, $I(J)$, called the* inside, *the other, $E(J)$, the* outside. *Further, for each $j \in J$, $\mathcal{A}_X(j)$ meets both $I(J)$, and $E(J)$.*

The above theorem requires the full notation $\mathcal{A}_X(j)$ (since $\mathcal{A}_J(j) \subseteq J$, thus fails to meet either $I(J)$ or $E(J)$). A proof of the digital Jordan curve theorem is in [3]. This proof uses only point-by-point methods using the specialization order used in the preceding proofs involving Alexandroff spaces. While the definition of Jordan curve given there is different, its equivalence with that in Definition 21 is shown there. Further, the above definition can be extended to an inductive definition of Jordan n-surface for arbitrary finite n:

Definition 23. A *Jordan 0-surface* is a 2-point discrete space. For $n > 0$, a *Jordan n-surface* is a finite, nonempty, connected set J such that for each $j \in J$, $\mathcal{A}(j)$ is a Jordan $(n-1)$-surface.

A *Jordan surface* is a Jordan 2-surface.

Theorem 24. *If $J \subseteq \mathbb{Z}^{n+1}$ is a Jordan n-surface, $n > 0$, then $\mathbb{Z}^{n+1} - J$ consists of two connected components, one finite, $I(J)$, called the* inside, *the other, $E(J)$, the* outside. *Further, for each $j \in J$, $\mathcal{A}(j)$ meets both $I(J)$, and $E(J)$.*

However, the proof in three dimensions (in [11]) and in higher dimensions (a consequence of [5]) is done by using the open quotient from $\mathbb{R}^n$ to $\mathbb{Z}^n$ and the Euclidean Jordan surface theorem. A proof using the digital techniques appropriate to these spaces remains to be found.

Figure 4 shows a Jordan curve and a subset of $\mathbb{Z}^2$ which is not a Jordan curve.

A particularly appealing way of looking at a portion of $\mathbb{Z}^2$ as representing the pixels on a computer screen is to view them as the (odd,odd) points in a

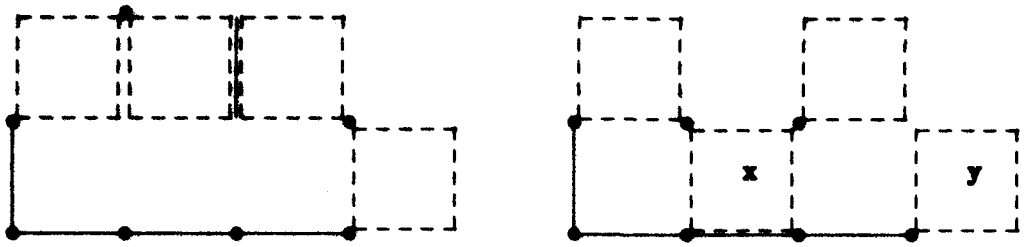


Fig. 4. Jordan curve Not one (x,y not adjacent to exactly two points)

product of two finite intervals, and see the other points as memory locations not corresponding to pixels (but rather, between them, for use as boundary points). The resulting representation has the property that the discrete subspace of pixels almost fills the space, making boundary points invisible, as they are in reality. The representation, with the subspace of pixels called the *open screen*, was discussed first in [4].

But some other embeddings of the computer screen in $\mathbb{Z}^2$ are also useful. For example, in [4], the screen is embedded in $\mathbb{Z}^2$ via a 45° *slant map*, and the result gives an easy proof of the (4,8) and (8,4) graph-theoretic Jordan curve theorems from the digital Jordan curve Theorem 22, although the same result is shown in [7] using the open screen embedding. Here we show the remaining two-dimensional graph-theoretic Jordan curve theorem from the digital Jordan curve Theorem, using another embedding. First recall that many of our definitions for the topological graph in digital n-space are special cases of general graph-theoretic definitions. Recall that a *graph* is a set with a binary relation, which we call *adjacency*.

Definition 25. A graph is a *Jordan curve* if each of its elements is adjacent to precisely two others. A *path in a graph Y from x to y* is a sequence $x_0, \ldots, x_n$ of elements of Y such that $x = x_0, y = x_n$, and for each $i < n, x_i, x_{i+1}$ are adjacent. A set Y is *path-connected* if for each $x, y \in Y$ there is a path in Y from x to y. A *path-component* of X is a maximal path-connected set. For each vertex x of a graph, $C_x = \{z \mid$ there is a path from x to $z\}$.

Also, let H^2 denote the hexagonal tiling of Euclidean 2-space in which the hexagons all have a horizontal major diagonal. In H^4, 2-adjacency is that in which a hexagon is adjacent to the 6 others which have a side common with it. We use (Jordan) 6-curve, 6-adjacent, etc., to specialize graph-theoretic ideas to this case (and continue to use the unadorned notations for the topological case).

Since it does not affect the graph and makes explanation of the following proof easier, we replace H^2 by a collection Hex of subsets of $\mathbb{Z}^2$: the *hexagons* and *edges*. The following sentences should be read with Fig. 5 in mind.

A *hexagon*, h consists of three points: its *diagonal*, $d(h) = (a, b)$, where we require that a be odd and $b \equiv (a-1) \bmod 4$, and its (*upper* and *lower*) *open*

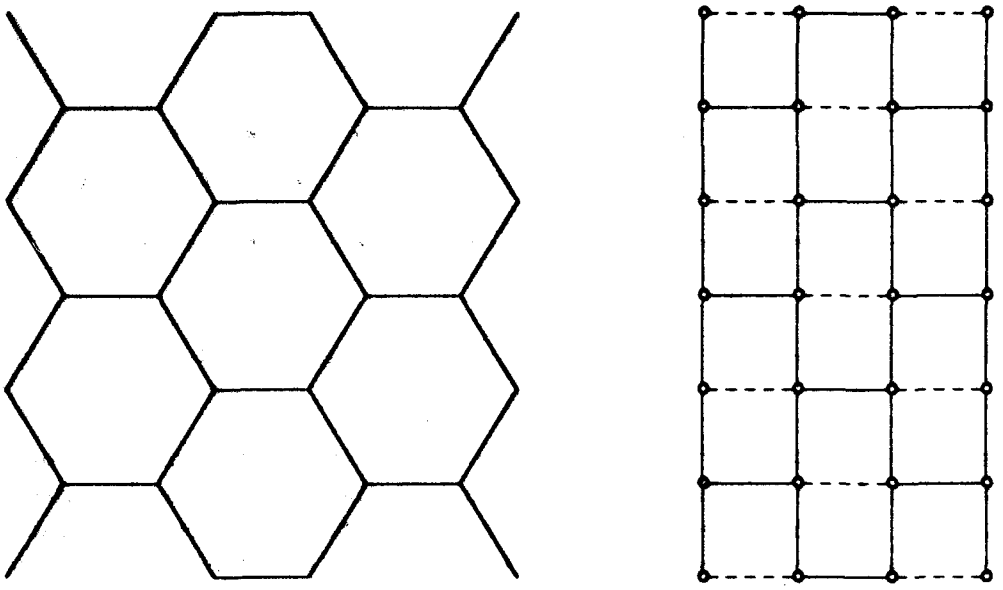


Fig. 5. Corresponding portions of H^2 and of Hex (missing closed points, edges = solid lines, diagonals = dashed lines).

elements, $o^+(h) = (a, b+1)$ and $o^-(h) = (a, b-1)$. Two hexagons h and k are 6-adjacent iff $cl(h) \cap cl(k) \neq \emptyset$, and in this case the unique element whose closure is $cl(h) \cap cl(k)$ is the midpoint $(\frac{a+a'}{2}, \frac{b+b'}{2})$; we call this the *edge* between h and k, denoted $e(h, k)$. The two remaining elements of $n(e(h, k))$ are open; one is in h, which we call $o(h, k)$, the other is in k, and called $o(k, h)$.

The edges embody the adjacency relation: A subset $A \subseteq H^2$ is 6-connected iff $\{x \mid x$ is in a hexagon in A or is the edge between two adjacent hexagons in $A\}$ is connected in $\mathbb{Z}^2$. We let H^e denote $\{x \mid x$ is in a hexagon or is the edge between two adjacent hexagons$\}$; then H^e is all of $\mathbb{Z}^2$ except its closed elements.

Given a 6-path, $A = \{h_1, \ldots, h_n\}$, we define $A^* = \{x_1, \ldots, x_m\}$ recursively: $\{h_1\}^* = \emptyset$ and assuming $\{h_1, \ldots, h_n\}^* = \{x_1, \ldots, x_m\}$ we construct $\{h_1, \ldots, h_{n+1}\}^*$ for the only three cases which may occur:
if $h_{n+1} = h_n$, $\{h_1, \ldots, h_{n+1}\}^* = \{x_1, \ldots, x_m\}$; if $h_{n+1} \neq h_n$, then
if $x_m = o(h_n, h_{n+1})$, $\{h_1, \ldots, h_{n+1}\}^* = \{x_1, \ldots, x_m, e(h_n, h_{n+1}), o(h_{n+1}, h_n)\}$,
if not, $\{h_1, \ldots, h_{n+1}\}^* = \{x_1, \ldots, x_m, d(h_n), o(h_n, h_{n+1}), e(h_n, h_{n+1}), o(h_{n+1}, h_n)\}$.

The reader should inductively check that if $A = \{h_1, \ldots, h_n\}$ is a 6-path then A^* is a path. If it is a Jordan 6-curve then $\{h_1, \ldots, h_n, h_1, h_2\}^*$ is a Jordan curve (with some points repeated): a diagonal or edge x is in A^* if its two adjacent points in H^e (those in $n(x) - \{x\}$) are in A^*. For each open x, $\mathcal{A}_{H^e}(x)$ is a 4 point discrete space, containing the diagonal of the hexagon in which x lies, and its edges with 3 other hexagons. But 2 hexagons in a Jordan 6-curve A are 6-adjacent to a given one; but by construction, $x \in A^*$ if and only if exactly two of these edges are in A^*, and the diagonal is not, or if exactly one edge and the diagonal are in A^*. Thus each element of A^* is adjacent to exactly two others.

Theorem 26. *If $J \subsetneq H^2$ is a Jordan 6-curve, then $H^2 - J$ consists of two 6-connected components, one finite, $I_6(J)$, called the* inside, *the other, $E_6(J)$, the* outside. *Further, for each $j \in J$, $\mathcal{A}_6(j)$ meets both $I_6(J)$, and $E_6(J)$.*

Proof. Suppose that J is a Jordan 6-curve in H^2. Since J^* is a Jordan curve, $\mathbb{Z}^2 - J^*$ consists of two connected components, one finite, $I(J^*)$, the other $E(J^*)$. We define $I_6(J) = \{h \in Hex \mid h \subseteq I(J^*)\}$, $E_6(J) = \{h \in Hex \mid h \subseteq E(J^*)\}$. $I_6(J)$ is surely finite.

If $h \in J$ then $\mathcal{A}_6(h)$ meets $I_6(J)$: By the construction of J^*, there is an open $x \in h \cap J^*$, and by the Jordan curve theorem in $\mathbb{Z}^2$, $\mathcal{A}(x) \cap I(J^*) \neq \emptyset$; also, by Lemma 20, $cl(x) - \{x\} = \mathcal{A}(x)$. Thus suppose $y \in \mathrm{cl}(x) \cap I(J^*)$; by Lemma 17, y is not open, and we now find, in either of the two other cases, an open $z \in n(y)$ contained in a hexagon not in J:

If y is mixed, then $n(y) = \{x, y, z\}$, and z is open, thus in a hexagon, k. If $k \in J$ then by construction $y \in J_1 \subseteq J^*$, a contradiction.

If y is closed, there are 4 open elements of $n(y)$. By definition of Hex, these lie in 3 hexagons, and then by definition of Jordan 6-curve, at most two of them can be in J. Let z be in the third.

Since $y \in \mathrm{n}(z)$, $\{y, z\}$ is connected, thus z is in the same component of $\mathbb{Z}^2 - J^*$ as is y, namely $I(J^*)$. Also, $z \in k \notin J$, and k is connected, thus $k \subseteq I(J^*)$, whence $k \in I_6(J)$.

It remains to be shown that $I_6(J), E_6(J)$ are 6-connected; the proofs are the same, so we do the former only. Given $h, k \in I_6(J)$, there is a path $\{x_1, \ldots, x_n\}$ in $I(J^*)$ from an element of h to one of k (since $I(J^*)$ is connected). Let $x_j \in cl(h_j)$; then $\{h_1, \ldots, h_n\}$ a 6-path from h to k in $I_6(J) \cup J$. If (by way of contradiction) $I_6(J)$ were not connected, then by the above, there would be some $h, k \in I_6(J)$ such that each 6-path in $I_6(J) \cup J$ from h to k meets J, and let $i > 0$ be the smallest number of elements in such an intersection. Thus we have a 6-path $\{h_1, \ldots, h_n\}$ from h to k whose intersection with J has exactly i elements, and let $h_m (\neq h, k)$ be the first element of this intersection. Exactly two elements of J, (h_{m-1}, h_{m+1}) are 6-adjacent to h_m, thus the elements of $I_6(J) \cup J$ 6-adjacent to h_m form a 6-connected set, which therefore contains a 6-path $\{g_1, \ldots, g_r\}$ from h_{m-1} to h_{m+1}. But then $\{h_1, \ldots, h_{m-1}, g_2, \ldots, g_{r-1}, h_{m+1}, \ldots, h_n\}$ is a 6-path in $I_6(J) \cup J$ with $i - 1$ elements in its intersection with J, our contradiction. □

It remains open whether the remaining known graph-theoretic digital Jordan surface theorems (other than the two-dimensional (4,8) and (8,4) results of [4, 7]) can be shown by use of Theorems 22 and 24. Of particular interest are the three-dimensional (6,18), (6,24), (18,6) and (24,6) results.

6 Alexandroff Homotopy Theory

We next define homotopy and digital homotopy. A straightforward discussion of homotopy (on [0,1]) is found in [15].

Definition 27. Given topological spaces X, Y, and continuous maps $f, g : X \to Y$, a *homotopy from f to g* is a continuous $F : X \times [0,1] \to Y$ such that for each $x \in X$, $F(x,0) = f(x)$ and $F(x,1) = g(x)$. If a homotopy from f to g exists, then f and g are called *homotopic* mappings. A *digital homotopy from f to g* is a continuous $F : X \times \mathbb{Z} \to Y$ such that for some positive integer n, each $x \in X$, $F(x,m) = f(x)$ if $m < -n$ and $F(x,m) = g(x)$ if $m > n$. If a digital homotopy from f to g exists, then f and g are called *digitally homotopic*.

Also, let Y^X denote the set of continuous functions from X to Y.

The two definitions are more similar than they look: For $F : X \times [0,1] \to Y$ and $k > 0$ define $G : X \times \mathbb{R} \to Y$, by $G(x,t) = F(x, \frac{k+t}{2k})$ if $t \in [-k,k]$, $= F(x,0)$ if $t < -k$, $= F(x,1)$ if $t > k$. This replaces a homotopy by a *real homotopy* (= a continuous $F : X \times \mathbb{R} \to Y$ such that for some positive integer n, each $x \in X$, $F(x,m) = f(x)$ if $m < -n$ and $F(x,m) = g(x)$ if $m > n$), which like a digital homotopy, is constant sufficiently far from 0. For Alexandroff spaces, particularly the finite spaces which often interest us, it is preferable to use digital homotopy, rather than homotopy, for reasons discussed below. To unify the discussion there, let $W = [0,1], \mathbb{Z}$, or $\mathbb{R}$.

Given a map $F : X \times W \to Y$, there is an induced map $\hat{F} : W \to Y^X$, defined by $(\hat{F}(w))(x) = F(x,w)$. Under certain conditions on X and W, there is a natural topology on Y^X such that $F : W \times X \to Y$ is continuous iff $\hat{F} : W \to Y^X$ is continuous, thus a homotopy F between a pair of maps is essentially a path $\hat{F}$ between the two:

Definition 28. Suppose each topological space X, is associated with a set of subsets $\mathcal{Q}_X$ of X. Then:

X *is locally* $\mathcal{Q}$ if for each $x \in T$, T open, there are $Q \in \mathcal{Q}_X$, U open such that $x \in U \subseteq Q \subseteq T$. W is a *$\mathcal{Q}$-space* if whenever $Q \times \{z\} \subseteq T \in \mathcal{T}_{X \times W}$, X any topological space, $Q \in \mathcal{Q}_X$, there is a $U \in \mathcal{T}_Y$ with $z \in U$ and $Q \times U \subseteq T$.

The *evaluation map* is $ev : X \times Y^X \to Y$ is defined by $ev(x,f) = f(x)$. Given a topological space X, $i_X : X \to X$ denotes the (continuous) *identity* map on X defined by $i_X(x) = x$.

Given two topological spaces X, Y, the *$\mathcal{Q}$-open topology* on Y^X is that generated by $\{S(Q,T) \mid Q \in \mathcal{Q}_X,\ T \in \mathcal{T}_Y\}$, where $S(Q,T) = \{g \in Y^X \mid x \in Q \Rightarrow g(x) \in T\}$. In particular, the compact-open topology is the $\mathcal{Q}$-open topology in the case that $\mathcal{Q}$ is the collection of compact sets, and the all-open topology is the $\mathcal{Q}$-open topology in the case that $\mathcal{Q}$ is the collection of all sets.

Theorem 29. *If X is locally $\mathcal{Q}$ and W is a $\mathcal{Q}$-space, then $F : X \times W \to Y$ is continuous iff $\hat{F} : W \to Y^X$ is continuous.*

If each Q in each $\mathcal{Q}_X$ is compact, then each topological space is a $\mathcal{Q}$-space. Thus if X is locally compact then for arbitrary W, $F : X \times W \to Y$ is continuous iff $\hat{F} : W \to Y^X$ is continuous with respect to the compact-open topology on Y^X.

If W is Alexandroff, then W is a $\mathcal{Q}$-space for arbitrary $\mathcal{Q}$. Thus if X and W are Alexandroff then for arbitrary Y, $F : X \times W \to Y$ is continuous iff $\hat{F} : W \to Y^X$ is continuous with respect to the all-open topology on Y^X.

If X is finite, both of the above topologies are restrictions of the product topology to $Y^X \subseteq \prod_X Y$.

Proof. We first show that ev is continuous if X is locally $\mathcal{Q}$: If $(x, f) \in X \times X^Y$ and an open $T \subseteq Y$ such that $f(x) = ev(x, f) \in T$, then by the fact that f is continuous and X is locally $\mathcal{Q}$, find an open $U \subseteq X$ and a $Q \in \mathcal{Q}_X$ such that $x \in U \subseteq Q \subseteq f^{-1}[T]$. Then consider the open $U \times S(Q, T) \subseteq X \times X^Y$; certainly $(x, f) \in S(Q, T)$ and if $(y, g) \in S(Q, T)$ then $ev(y, g) = g(y) \in T$.

Thus if X is locally $\mathcal{Q}$ and $\hat{F} : W \to Y^X$ is continuous then $F : X \times W \to Y$ is continuous, since $F = ev \circ (i_X, \hat{F})$, a composition of continuous maps.

We next show that for Y a $\mathcal{Q}$-space, if $F : X \times W \to Y$ is continuous then so is $\hat{F} : W \to Y^X$. We must find, if $\hat{F}(w) \in T$, T open in X^Y, an open $U \ni w$ in W such that $\hat{F}[U] \subseteq T$. But $\hat{F}(w) \in T \Rightarrow F[X \times \{w\}] \subseteq T \Rightarrow$ (since Y is a $\mathcal{Q}$-space) for some $U \ni w$ open in W, $X \times U \subseteq F^{-1}[T]$. But then, $\hat{F}[U] = F[X \times U] \subseteq T$, as required.

Now assume each Q in each $\mathcal{Q}_X$ is compact, or Y is Alexandroff. The last two assertions of the theorem will be shown if for arbitrary $Q \in \mathcal{Q}_X$, $z \in Y$, if $Q \times \{z\} \subseteq T \in \mathcal{T}_{X \times Y}$, Y, there is a $U \in \mathcal{T}_Y$ with $z \in U$ and $Q \times U \subseteq T$. Because T is open in the product topology, for each $x \in Q$ there are $V_x \ni x, U_x \ni z$ open such that $V_x \times U_x \subseteq T$. If Q is compact, there is by Definition 7 a finite $R \subseteq Q$ such that $Q \subseteq \bigcup_{x \in R} V_x$; otherwise, let $R = Q$, and in either case let $U = \bigcap_{x \in R} U_x$. If R is finite or Y is Alexandroff, U is open, and if $(x, y) \in Q \times U$ then for some $w, x \in V_w$, and then $y \in U_w$, so $(x, y) \in T$.

Indeed, if $f \in Y^X$, $x \in X$, $f(x) \in T$, T open, then $\{g \in Y^X \mid g(x) \in T\}$ is in either of the above topologies, so each is at least as rich as the product topology. On the other hand, for any subset Z of X, $S(Z, T) = \bigcap_{z \in Z} \{g \mid g(z) \in T\}$, a finite intersection of sets open in the product topology, thus in the product topology. □

Thus for finite X, Y, maps are homotopic iff they are in the same component of Y^X with the restricted product topology. Now consider a path between two such maps. This is, in fact, an $\hat{F} : \{0, \ldots, n\} \to Y^X$ for some $n \geq 0$. The continuity of $\hat{F}$ is by the definition of product, and the discussion of products of Alexandroff spaces at the beginning of the section on digital n-space, equivalent to the fact that for even i, $\hat{F}(i) \in cl(\hat{F}(i-1))$, and for odd i, $\hat{F}(i-1) \in cl(\hat{F}(i))$. Thus to analyse when maps are homotopic in this situation, it suffices to determine when $f \in cl(g)$. But by definition of the product topology and the discussion preceding, this will hold iff for each x, $f(x) \in cl(g(x))$.

References

1. Alexandroff, P. S. (1937). Diskrete Räume, Mat. Sb. 1, pp. 501–519.
2. Gierz, G., Hofmann, K. H., Keimel, K., Lawson, J. D., Mislove M. and Scott, D. S. (1980). A Compendium of Continuous Lattices, Springer-Verlag, Berlin.
3. Khalimsky, E. D., Kopperman, R. D., Meyer, P. R. (1990). Computer graphics and connected topologies on finite ordered sets, Topology and Appl. 36, pp. 1–17.

4. Khalimsky, E. D., Kopperman, R. D., Meyer, P. R. (1990). Boundaries in digital planes, J. of Applied Mathematics and Stochastic Analysis 3, pp. 27–55.
5. Kong, T. Y., Khalimsky, E. D. (1990). Polyhedral analogs of locally finite topological spaces, In: R. M. Shortt (ed.), General Topology and Applications, Proc. 1988 Northeast Conference, Marcel Dekker, NY, pp. 153–164.
6. Kong, T. Y., Kopperman, R. D., Meyer, P. R. (1991). Using general topology in image processing, Geometric Problems of Image Processing (Vol. 4, Research in Informatics) Akademie Verlag, Berlin, pp. 66–71.
7. Kong, T. Y., Kopperman, R. D., Meyer, P. R. (1991). A topological approach to digital topology, Am. Math. Monthly 98, pp. 901–917.
8. Kong, T. Y., Kopperman, R. D., Meyer, P. R. (1991). Which Spaces have Metric Analogs?, Gen. Top. and Appl., Lecture Notes 134, Marcel Dekker, pp. 209–216.
9. Kong, T. Y., Kopperman, R. D., Meyer, P. R. (eds.) (1992). Special issue of Topology and its Applications 46 (3), pp. 173–180.
10. Kong, T. Y., Rosenfeld, A. (1989). Digital Topology: Introduction and Survey, Computer Vision, Graphics, and Image Processing 48, pp. 357–393.
11. Kopperman, R. D., Meyer, P. R., Wilson, R. G. (1991). A Jordan surface theorem for three-dimensional digital spaces, Discrete and Computational Geometry 6, pp. 155–162.
12. Kovalevsky, V. A. (1986). On the Topology of Discrete Spaces. Studientexte, Digitale Bildverarbeitung, Heft 93/86, Technische Universität Dresden.
13. Kovalevsky, V. A. (1989). Finite topology as applied to image analysis, Computer Vision, Graphics and Image Processing 46, pp. 141–161.
14. Morris, S. A., Topology Without Tears, available from author, Dean of Informatics, University of Wollongong, Wollongong, NSW, 2500, Australia.
15. Munkres, J. R. (1975). Topology: A First Course, Prentice-Hall, Englewood Cliffs, NJ.
16. Rosenfeld, A. (1979). Picture Languages, Academic Press, NY.
17. Simmons, G. F. (1983). Introduction to Topology and Modern Analysis, Krieger, Malabar, FL.

Topological Foundations of Shape Analysis

Vladimir A. Kovalevsky

Technische Fachhochschule Berlin, Luxemburger Str. 10, 13353 Berlin, Germany

Abstract. A concept of *finite topological spaces* is presented based on combinatorial topology and on the notion of abstract *cell complexes*. It is shown how to apply the concept to image processing and especially to shape analysis. Topologically consistent solutions of the following problems are presented: connectivity of subsets, labelling connected components, tracking boundaries in two- and three-dimensional images, filling interiors of curves, and determining the genus of surfaces.

Keywords: abstract cell complexes, boundary, connectivity, filling, finite topological space, genus of a surface, labelling connected components, membership rule, tracking.

1 Introduction

Topological notions like boundaries, connectivity of subsets, and genus of a surface, as well as various geometrical notions, are important for shape analysis. For this reason it is necessary to look for possibilities of implementing basic topological concepts in digital picture analysis.

The overwhelming majority of the topological literature is concerned with infinite sets. Digitised pictures, however, are defined on finite sets, for example, on arrays of pixels or voxels. The problem of transferring the topological knowledge concerned with infinite sets into the world of digital pictures is by no means a trivial one. Some basic ideas of general topology are not applicable to finite set as, for example, the concept, that any small neighbourhood of a point contains infinitely many other points.

Publications covering finite topological spaces appeared at least 50 years ago [1], but this knowledge was weakly represented in topological text books. That is why specialists in image analysis were forced to look for their own solution of the problem. Thus Rosenfeld [12] introduced the notion of *adjacency graphs* representing a digital image as a graph whose vertices are the pixels or voxels. An edge of the graph corresponds to each pair of adjacent pixels (voxels). It became possible to introduce the notion of connected subsets as corresponding

to connected subgraphs. The notion of a neighbourhood, which is necessary to define boundaries, was introduced as the set of adjacent graph vertices.

However, attempts to develop a consistent topology based on adjacency graphs have failed due to the well-known connectivity paradox [10] and certain great difficulties in defining the boundary of a subset. More about these difficulties and their solutions may be found in [7, 8].

Intuitive attempts to overcome the difficulties with the boundary were often reported in the literature. About 15 years ago the notion of "cracks" was introduced [13]. A crack is a short line segment separating two adjacent pixels, the latter being considered as squares. Herman and Webster [4] define the boundary surface of a three-dimensional region as a set of "faces": space elements separating two adjacent voxels from each other. These ideas may serve as evidence that image processing specialists have a strong intuition that a consistent topological concept for digital images must include space elements of *various nature*.

This intuition will be verified in Sect. 2 where it is shown that the solution of these problems consists of considering the digital plane as a finite topological space in full accordance with topological axioms. It is shown that the most suitable for practical purposes is the particular case of a finite topological space known as abstract *cell complex*. A topologically consistent definition of connectivity is given in Sect. 3 with applications to labelling connected components. Boundaries in cell complexes are considered in Sect. 4. Applications to tracking boundaries in two-dimensional images, filling their interiors, and reconstructing subsets from their boundaries are presented here. Section 5 is devoted to boundaries (surfaces) in three-dimensional images.

2 Finite Topology

As is well known, a topological space T is a set E of abstract space elements, usually called *points*, with a system SY of some singled out subsets of T declared to be the *open* subsets. The system SY must satisfy the axioms:

A1 The union of any family of sets of SY belongs to SY.

A2 The intersection of any finite family of sets of SY belongs to SY.

A topological space $T = (E, SY)$ is called *finite* if the set E contains finitely many elements.

As an example of a finite topological space consider the surface of a polyhedron (Fig. 1). It consists of three kinds of space elements: faces, edges, and vertices. An edge l bounds two faces, say f' and f''.

The edge l is bounded by two vertices v' and v''. These two vertices are also said to bound the faces f' and f''. Let us declare as *open* any subset S of faces, edges, and vertices, such that for every element e of S all elements of the surface which are bounded by e are also in S. According to this declaration a face is an open subset. An edge l with the two faces f' and f'' bounded by it, also compose an open subset. So does a vertex united with all edges and faces bounded by it. It is easy to see that the open subsets thus defined satisfy the axioms. Hence,

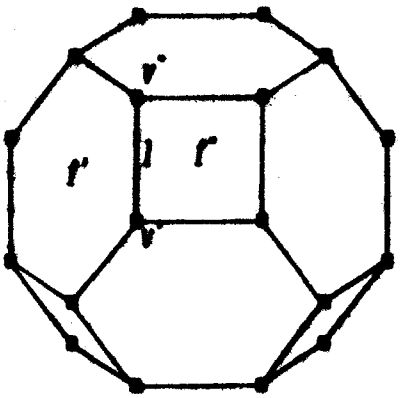


Fig. 1. The surface of a polyhedron considered as a topological space

a topological space is defined. It is finite if the surface of the polyhedron has a finite number of elements.

The elements of such a space are not topologically equivalent: for example, a face f' bounded by the edge l, belongs to all open subsets containing l, but l does not belong to the set $\{f'\}$ which is an open subset containing f'. One can see that this kind of order relation between the elements corresponds to the bounding relation.

Further, it is possible to assign numbers to the space elements in such a way that elements with lower numbers are bounding those with higher numbers. The numbers are called *dimensions* of the space elements. Thus vertices which are not bounded by other elements get the lowest dimension, say 0; the edges get the dimension 1, and the faces the dimension 2. Structures of this kind are known as abstract cell complexes [15].

2.1 Cell Complexes

Definition 1. An *abstract cell complex* (ACC) $C = (E, B, \dim)$ is a set E of abstract elements provided with an antisymmetric, irreflexive, and transitive binary relation $B \subset E \times E$ called the *bounding relation*, and with a dimension function $\dim : E \to I$ from E into the set I of non-negative integers such that $\dim(e') < \dim(e'')$ for all pairs $(e', e'') \in B$.

Elements of E are called *abstract cells*. It is important to stress that abstract cells should *not* be regarded as point sets in a Euclidean space. That is why ACCs and their cells are called abstract. Considering cells as abstract space elements makes it possible to develop the topology of ACCs as a *self-contained theory which is independent of the topology of Euclidean spaces.*

If the dimension $\dim(e')$ of a cell e' is equal to d then e' is called a *d-dimensional* cell or a *d-cell*. An ACC is called *k-dimensional* or a *k-complex* if the dimensions of all its cells are less or equal to k. If $(e', e'') \in B$ then e' is said *to bound* e''.

Examples of ACCs are shown in Fig. 2. Here and in the sequel the following graphical notations (similar to that of Fig. 1) are used: 0-cells are denoted by small circles or squares representing points, 1-cells are denoted by line segments, 2-cells by interiors of rectangles, 3-cells by interiors of polyhedrons. The bounding relation in these examples is defined in a natural way: a 1-cell represented in the figure by a line segment is bounded by the 0-cells represented by its end points, a 2-cell represented by the interior of a square is bounded by the 0- and 1-cells composing its boundary etc.

The notion of a pixel which is widely used in computer graphics and image processing should be identified with that of a 2-cell (elementary area) rather than with a point, since a pixel is thought of as a carrier of a grey value which can be physically measured only if the pixel has a non-zero area. On the other hand, we are used to thinking of a point as an entity with a zero area. Similarly, a voxel is a three-dimensional cell.

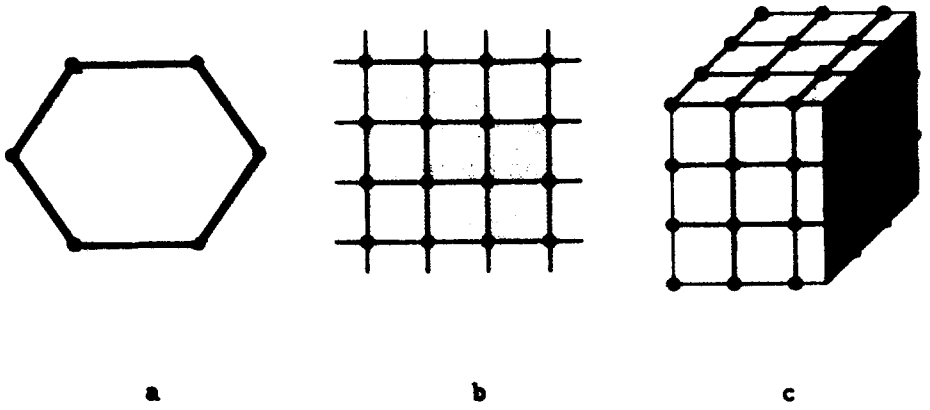


Fig. 2. Examples of ACCs: (a) 1-dimensional, (b) 2-dimensional, (c) 3-dimensional

The topological structure of an ACC is defined by

Definition 2. A subset S of E is called *open in* C if for every element e' of S, all elements of C which are bounded by e' are also in S.

It has been shown [7] that for any finite topological space there exists an ACC having an equivalent topological structure. A particular feature of ACCs is, however, the presence of the dimension function. Due to this property ACCs are attractive for applications: dimensions make the concept descriptive and comprehensible for non- topologists. It is possible to make drawings of ACCs to demonstrate topological evidence (e.g. Figs. 2 and 3), a possibility lost, unfortunately, during the modern phase of topological development. ACCs invented many years ago are being discussed more and more [3, 5] because of their attractive features. Therefore we shall restrict ourselves to considering ACCs as representatives of finite topological spaces.

Definition 3. A *subcomplex* $S = (E', B', \dim')$ of a given ACC $C = (E, B, \dim)$ is an ACC whose set E' is a subset of E and the relation B' is an intersection of B with $E' \times E'$. The dimension $\dim'$ is equal to dim for all cells of E'.

This definition makes clear that to define a subcomplex S of $C = (E, B, \dim)$ it suffices to define a subset E' of the elements of E. Thus it is possible to speak of a subcomplex $E' \subset E$ while understanding the subcomplex $S = (E', B', \dim')$. All subcomplexes of C may be regarded as subsets of C and thus it is possible to use the common formulae of the set theory to define intersections, unions, and complements of subcomplexes of an ACC C.

Definition 2, defining the notion of open subsets, simultaneously defines the open subcomplexes of a given ACC. According to the axioms of topology any intersection of a finite number of open subsets is open. In a finite space there is only a finite number of subsets. Therefore in a finite ACC the intersection of all open subcomplexes containing a given cell c is an open subcomplex. It is called the *smallest open neighbourhood of* c in the given ACC C and will be denoted by SON(c). Notice that there is no such notion for a connected Hausdorff space.

It is easy to see that SON(c) consists of the cell c itself and of all cells of C bounded by c. Figure 3 shows some examples of the SONs of cells of different

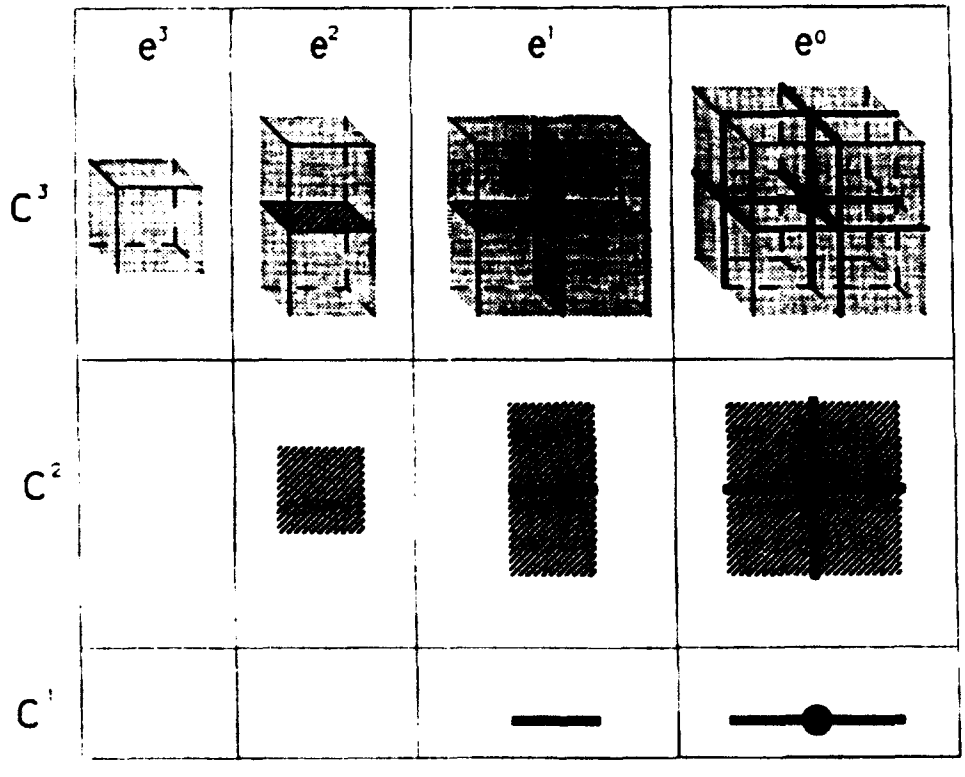


Fig. 3. Smallest open neighbourhoods of k-dimensional cells e^k in d-dimensional ACCs C^d, $d = 1, 2, 3$

dimensions in different ACCs. Notice that in any case the SON of a cell of the highest dimension is the cell itself. More about ACCs may be found in [7, 8].

2.2 Multidimensional Manifolds

There are spaces with some especially simple structures. They are called manifolds. The notion of manifolds in the Hausdorff topology is defined in a rather complex way. In the finite topology a manifold may be defined in a rather simple way: a finite manifold is a connected nonbranching finite space. To make this notion more precise let us introduce:

Definition 4. Two cells e' and e of an ACC C are called *incident with each other in* C iff either $e' = e''$, or e' bounds e'', or e'' bounds e'.

Definition 5. Two ACCs are called B-isomorphic to each other if there exists a one-to-one correspondence between their cells which retains the bounding relation.

Definition 6. An *n-dimensional finite manifold* M_n is an n-dimensional ACC satisfying the following conditions:

(1) a 0-dimensional manifold M_0 consists of two cells with no bounding relation between them;

(2) an n-dimensional manifold M_n with $n > 0$ is connected;

(3) for any cell c of M_n the subcomplex of all cells different from c and incident with c is B-isomorphic to an $(n-1)$-dimensional manifold (nonbranching condition).

This definition is an attempt to generalize the well-known definition of pseudomanifolds [14].

Topological properties of two-dimensional manifolds are well known. They are defined by the *genus* which in turn is defined by the Euler polyhedron formula:

$$N_2 - N_1 + N_0 = 2\,(1 - G)\ .$$

Here N_2, N_1, and N_0 are the numbers of 2-, 1-, and 0- dimensional cells respectively; G is the genus.

The notion of the genus can be illustrated by the following remarks: a manifold of genus 0 looks like a sphere (subdivided into cells), a manifold of genus 1 looks like a torus. A manifold of genus equal to G looks like a sphere with G handles.

Properties of manifolds of higher dimensions are still not sufficiently investigated. On the other hand, they may be of great interest for our understanding of the universe since there are reasons to believe that our physical space is a four-dimensional manifold. Topological properties of the space may be of great importance for the theory of elementary particles. Since ACCs of any dimension may be easily represented by computers there is a possibility of investigating the properties of finite manifolds of dimensions greater than two by means of computers.

3 Connectivity

Consider now the transitive closure of the incidence relation according to Definition 4. (The transitive closure of a binary relation R in E is the intersection of all transitive relations in E containing R). This new relation will be declared as the *connectedness relation*. As any transitive closure it must be defined recursively:

Definition 7. Two cells e' and e'' of an ACC C are called *connected to each other in* C iff either e' is incident with e'', or there exists in C a cell c which is connected to both e' and e''.

It may be easily shown that the connectedness relation according to Definition 7 is an equivalence relation (reflexive, symmetric, and transitive). Thus it defines a partition of an ACC C into equivalence classes called the *components of* C.

Definition 8. An ACC C consisting of a single component is called *connected*.

It is easy to see that Definitions 7 and 8 are directly applicable to subsets of an ACC C: any subset is, according to Definition 3, a subcomplex of C , and is again an ACC. It is, however, important to stress that all intermediate cells c mentioned in Definition 7 must belong to the subset under consideration. Therefore it is reasonable to regard an equivalent definition of connected ACCs:

Definition 9. A sequence of cells of an ACC C beginning with c', and finishing with c'' is called a *path in* C from c' to c'' if every two cells which are adjacent in the sequence are incident.

Definition 10. An ACC C is called *path-connected* if for any two cells c', and c'' of C there exists a path in C from c' to c''.

Kong *et al.* [6] have shown that Definitions 8 and 10 are equivalent. As shown by the author [7, 8] these definitions are in full accordance with classical topology and free of paradoxes.

3.1 Membership Rules

An n-dimensional image ($n = 2$ or 3) is defined by assigning numbers (grey values or densities) to the n-dimensional cells of an n-dimensional ACC. There is no need to assign grey values or densities to cells of lower dimensions. Such an assignment would be unnatural since a grey value may be physically determined only for a finite area. We have agreed to interpret 2-cells in a two-dimensional ACC as elementary areas. Cells of lower dimensions have area equal to zero. Similarly, a density may be physically determined only for a finite volume which is represented in a three-dimensional ACC by a 3-cell.

However, when considering the connectivity of a subset (subcomplex) of an n-dimensional ACC, the membership in a subset under consideration must be specified for cells of *all* dimensions. Under this condition the connectivity of the

subset is consistently specified by Definition 8 or 10. It is important to stress that the connectivity is determined by means of the lower dimensional cells which are serving as some kind of "glue" joining n-dimensional cells. A set consisting of only n-dimensional cells is always disconnected.

Generally a partition of the ACC in disjoint subsets must be considered, and each cell of the ACC must be assigned to exactly one subset of the partition. Every cell gets the identification number of a subset as its membership label.

The membership of cells of lower dimensions cannot be specified in the same way as that of the cells of highest dimension (n-cells) since the lower dimensional cells have no grey values. This must be done by using certain a priori knowledge about the image under consideration. The membership of a lower dimensional cell may be specified as a function of the membership labels, and grey values of the n-cells bounded by it by means of the *membership rules*. Consider an example of such a rule.

Maximum Value Rule: In an n-dimensional ACC every cell c of dimension less than n gets the membership label of that n-cell which has the maximum grey value (density) among all n-cells bounded by c.

It is possible to formulate a similar Minimum Value Rule. The connectivity of a binary image is similar in both cases to that obtained according to a widely used idea of an 8- adjacency for objects, and a 4-adjacency for the background [13]. An important advantage of the Maximum (Minimum) Value Rule is the possibility of using it for multi-valued images. A slightly more complicated, and also practically useful rule may be found in [7]. Also situations in which an explicit specification of the membership labels may be useful, are discussed there.

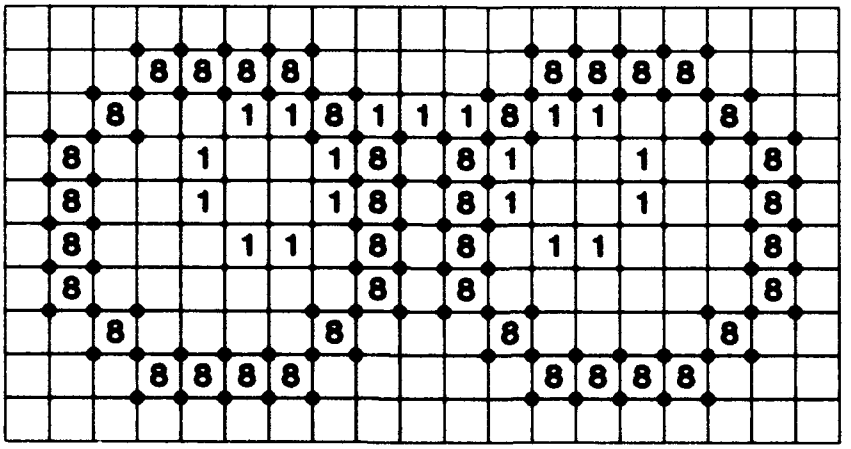


Fig. 4. An image with three grey values (empty pixels have the value 0, explanation in text)

The advantages of the Maximum Value Rule may be seen in the example of Fig. 4. The image has three grey values: 8, 1, and 0 (empty boxes in Fig. 4).

When applying the Maximum Value Rule, 0-cells shown as dark circles obtain their membership from the pixels with the grey value 8, 0-cells shown as dots belong to the sets with the grey-level 1. Correspondingly, the image has 2 components with the value 8; 3 components with the value 1; and 5 components with the value 0. This is in accordance with our intuitive idea of connected components.

Compare these results with that obtained with adjacency relations. When applying different kinds of adjacency for pixels with different grey values, one obtains numbers of components shown in Table 1. The notation in the table may be explained by the following example: $GV = 1$ and $Ad = 8$ means that all pixels in Fig. 4 with grey value 1 have the 8- adjacency. $NC = 1$ means that the set of such pixels consists of 1 component.

Table 1. Number of components in the image of Fig. 4 under different adjacencies GV - grey value, Ad - adjacency, NC - number of components

Variant 1			Variant 2			Variant 3		
GV	Ad	NC	GV	Ad	NC	GV	Ad	NC
8	8	2	8	8	2	8	8	2
1	8	1	1	8	1	1	4	9
0	8	1	0	4	5	0	4	5

It is easy to see that all variants contradict our intuition.

3.2 Labelling and Counting Connected Components

Definition 7 may be directly used to label connected components of a segmented image. A digital image is given as a two- or three-dimensional array with grey values (densities) assigned to each pixel (voxel). Results of the segmentation of the image into quasi-homogeneous segments are also given as segment labels assigned to each pixel (voxel).
The problem of labelling connected components consists of assigning to each pixel (voxel) of the image the identification number of the component to which it belongs.

The well-known solution for two-dimensional binary images [13] is as follows. The image is scanned row by row. For each pixel P the following set S of pixels is defined: a pixel belongs to S if it is adjacent to P, is already visited, and has the same segmentation label as P. If S is empty, then P is given a new component number. If all components of S have the same component number, then P has this number. If S consists of more than one component and the components of S have different component numbers, then P is given one of the numbers and all the numbers are recorded as being equivalent. When the whole image has been scanned in this way, the records must be investigated and the classes of equivalent numbers determined. The image must then be rescanned and the old numbers replaced by the numbers of equivalence classes.

The algorithm based on ACCs is similar to that just described. The main difference consists of the following. The set S of adjacent pixels is replaced by the set C of incident cells of lower dimensions which are simultaneously incident to some already visited pixels. Each cell of C is given its segment label according to a membership rule as explained in the previous section. The cell is also given the corresponding component number from one of the already visited pixels. Now the subset C' of cells of C having the same segmentation label as P is investigated in the same way as the set S. The advantage of this procedure is that it may be used for non-binary images while avoiding wrong decisions demonstrated in Table 1.

Unfortunately, in most publications (including [13]) there is no description of an efficient procedure for finding the equivalence classes. The few procedures described in the literature need either much computation time or an *additional* memory space greater than the output image containing the component labels. The author has found a component labelling algorithm which needs no additional memory. The processing time is twice the time of scanning the image. The algorithm cannot be presented here for reasons of space. The algorithm is *applicable for three- dimensional non-binary images.*

The problem of *counting* the components is much simpler than that of labelling them since no equivalence classes need to be determined. The subset C' must be determined in the same way as before. The component counter is first incremented for each pixel P. Then the counter must be decremented by the number of components of C'.

4 Boundaries

The theory of ACCs leads to a topologically consistent definition of the boundary of a subset of an image. The notion of a boundary remains the same as in general topology:

Definition 11. The *boundary* (frontier) of a subcomplex S of an ACC C relative to C is the subcomplex $\mathrm{Fr}(S, C)$ consisting of all cells c of C such that the $\mathrm{SON}(c)$ contains cells both of S and of its complement $C - S$.

Figure 5a shows an example of a subcomplex S of a two- dimensional ACC, Fig. 5b its boundary according to Definition 11, and Fig. 5c the "inner" and "outer" boundaries of S under 8-adjacency [13].

Consider their properties. The boundary $\mathrm{Fr}(S, C)$ in an n- dimensional ACC C contains no n-dimensional cells since n is the highest dimension and hence an n-cell is bounding no cells of C. Therefore the SON of such a cell consists of a single cell which is the cell itself. Hence the SON cannot contain cells of both S and its complement, and the cell cannot belong to the boundary. Consequently, the boundary of S is a subcomplex of a lower dimension equal to $n-1$. Thus the boundary of a region (a connected open subcomplex) in a two-dimensional ACC contains no pixels and consists of 0- and 1-cells. It looks like a closed polygon or several polygons if the region has some holes in it. The boundaries thus defined

are analogous to the "(C, D)-borders" or sets of "cracks" briefly mentioned in [13, (second edition)].

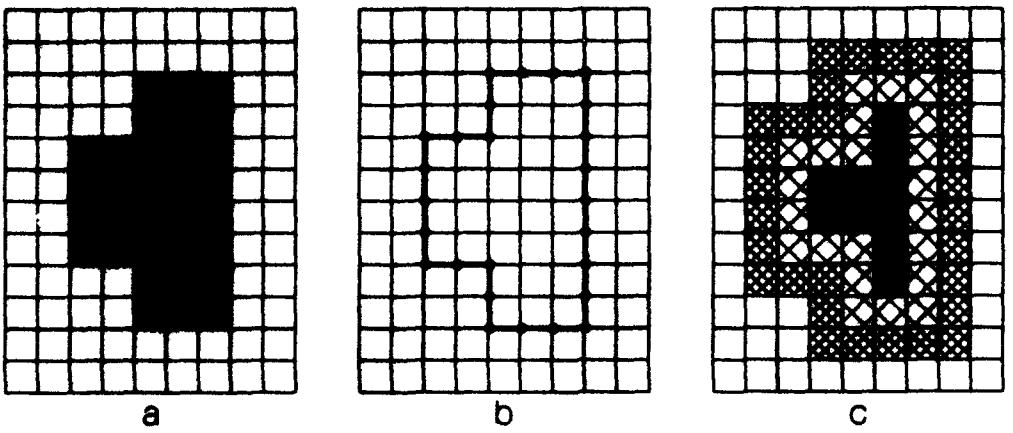


Fig. 5. (a) A subset, (b) its boundary according to Definition 11, (c) and its "inner" and "outer" boundaries under 8-adjacency

Similarly, the boundary of a region in a three-dimensional ACC contains no voxels and consists of 0-, 1-, and 2-cells. It looks like a closed surface of a polyhedron (or several surfaces if the region has some holes). A 2-cell of a boundary separates a voxel of the region from a voxel of its complement. Thus the 2-cells of the boundary are the "faces" considered in [4]. We may see now that the theory of the ACCs brings many intuitively introduced notions together in a consistent and topologically well-founded concept.

The boundaries $\mathrm{Fr}(S, C)$ in two-dimensional images have a zero area and boundaries in three-dimensional images a zero volume, which is not the case for boundaries in adjacency graphs (see Fig. 5c).

The next peculiarity of the boundary $\mathrm{Fr}(S, C)$ is that it is unique: there is no need (and no possibility!) of distinguishing between the inner and outer boundary as they were defined, for example, by Pavlidis [10] or between the "D-border of C" and "C-Border of D" [13]. A boundary according to Definition 11 is the same for a subset and for its complement, since Definition 11 is symmetric with respect to both subsets. This is not the case for boundaries in adjacency graphs.

The boundary $\mathrm{Fr}(S, C)$ depends neither on the kind of adjacency (which notion is no longer used) nor on the membership rules as defined in Sect. 3. The proof of the last assertion may be found in [8].

4.1 Tracking Boundaries in Two-Dimensional Images

The tracking algorithm described next is identical with "crack following" [13]. The present description is given in terms of ACCs which has the advantage that it is topologically justified and more comprehensible. The tracking goes from one 0-cell to the next, step by step, in such a direction that the region with the chosen label (the object) always remains on the right hand side of the direction. These moves travel along the 1-cells which in a two-dimensional Cartesian ACC [9] are either horizontal or vertical. Thus there are only *four* possible directions as shown in Fig. 6.

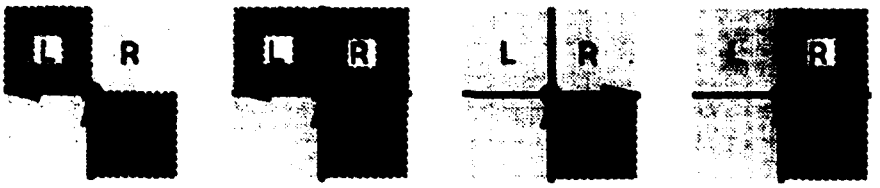


Fig. 6. Turn rules for boundary tracking

Having only four directions, rather than eight as is usual when tracking "boundaries" in adjacency graphs, already makes the algorithm simpler.
When arriving at the next 0-cell p, the direction of the last step having led to p is known. Thus it is known that the 2-cell lying to the right of this direction belongs to the object and that lying to the left belongs to the background (Fig. 6). In this way the membership of two pixels of SON(p) is already known. It is only necessary to test the labels of the remaining two pixels of SON(p) lying ahead: one to the right and one to the left of the direction of the last step (L and R in Fig. 6). Consider the case when the object has a greater grey-level than the background and accept, for example, the Maximum Value Rule to determine the membership of the 0-cells. Then the actual 0-cell p (denoted by a circle in Fig. 6) always belongs to the object, because p is a boundary cell and, according to Definition 10, there must be in the SON(p) at least one object pixel. This pixel having the maximum grey-level determines the membership of p.

The direction of the next step depends upon the labels of L and R in the following way: if L is in the object then turn left, else if R is in the background turn right, else retain the old direction. This decision rule is the kernel of the tracking algorithm. The reminder consists of some obvious procedures for calculating the necessary coordinates. The whole procedure contains about 20 Pascal instructions. Tracking algorithms which do not use the concept of cell complexes [11, 13] are much more complicated and less comprehensible.

When tracking the boundary of an object, it is possible to encode the boundary by the *crack-code*, which is the well-known chain code (Freeman code) with four directions. Together with the coordinates of the starting point, it gives the possibility to reconstruct the boundary and the object itself (see next section).

4.2 Filling Interiors of Curves

Consider now the problem of filling the interior of a closed curve. The problem is obviously equivalent to that of deciding if a pixel is inside or outside the curve: the inner pixels must be filled, the outer must not. The decision is based on the fact that a ray which starts at a given point and goes to some point at the border of the image crosses the given curve an odd number of times if the point is inside the curve, and an even number of times otherwise. Difficulty arises in discriminating between crossing and tangency.

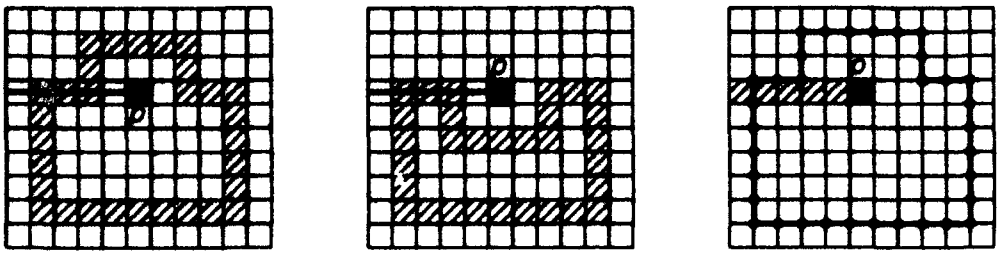


Fig. 7. Recognising inner pixels in adjacency graphs (a,b) and in an ACC (c)

It may be seen in Fig. 7a and 7b that when describing curves as sets of pixels, situations may occur in which it is impossible to decide correctly whether a pixel p is in the interior of the curve when analysing only the line containing p: the lines containing p are identical in Fig. 7a and 7b whereas p is inside the curve in Fig. 7a but outside in Fig. 7b.

Algorithms not based on the concept of ACCs (e.g. [11]) are rather complicated since they need to test three adjacent lines to decide between crossing and tangency. In the case of an ACC, the ray is replaced by a horizontal open strip consisting of alternating 2-cells and vertical 1-cells, all lying in a horizontal row of the raster containing the pixel p (Fig. 7c). The curve is represented as a 1-dimensional subcomplex consisting of alternating 0- and 1-cells. There arises no problem of tangency since a horizontal strip does not contain horizontal 1-cells. Crossings with the curve are only possible on vertical 1-cells. Therefore the filling is reduced to scanning the image with the given curve horizontally, row by row, and counting in each row the encountered vertical 1-cells of the curve. Counting must start with 0 at the left side of each row. For each pixel in the row the number of vertical 1-cells counted since the start of the row must be tested: if the count is odd then the pixel must be filled, otherwise not. In other words, filling of subsequent pixels in a row must be started whenever the count becomes odd, and stopped whenever it becomes even. For example, in the 4th row of Fig. 7c the count becomes equal to 1 in the second column. Thus the pixels in columns 2 to 10 must be filled. In the 11th column the count becomes 2 and the filling must be stopped. A similar algorithm again based on the notion

of "cracks" is described in [13].

Discriminating between inner and outer pixels of a closed curve is important for shape analysis since it gives us the possibility of determining which of a set of objects with holes are interior to other objects. It is possible to construct in this way *tree structures* completely describing the topological structure of a complex object.

5 Surfaces in Three-Dimensional Images

An important means of determining the shape of objects in three-dimensional images consists of analysing their surfaces. A surface of an object is its boundary according to Definition 11. The problem consist in detecting connected components of the boundaries and in determining their properties, both topological and geometrical. One of the possible ways of realising this consists of tracking the surfaces. Tracking goes from a cell to another cell connected to it. Hence, the process of uninterrupted tracking always relates to a connected component of a surface.

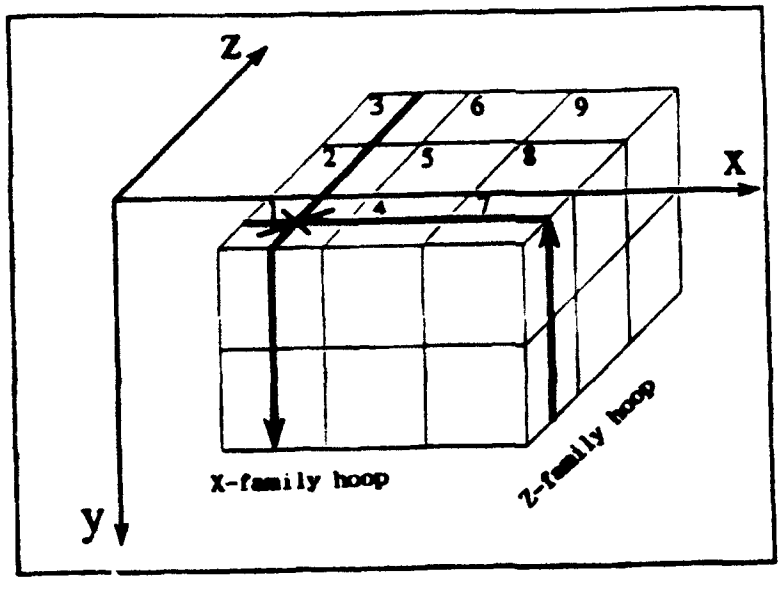


Fig. 8. Tracking a surface in a three-dimensional ACC

Tracking surfaces is much more complicated than tracking curves. Only in the simplest cases can it be organized as scanning rows and columns. However, it may be shown that any surface in a Cartesian ACC [9] may be represented as a union of "hoops" (Fig. 8), each lying in a slice with one of the three coordinates being constant. A *hoop* is a closed sequence of alternating mutually incident two- and one- dimensional cells having one of the three coordinates constant.

There are three families of hoops each corresponding to one of the coordinate axes: a hoop of one family circulates about an axis parallel to the corresponding coordinate axis. The corresponding coordinate remains constant. Hoops of two such families are shown in Fig. 8.

Gordon and Udupa have shown [2] that, to track a component of a surface, it suffices to track all hoops of only two families. A hoop of the X-family and a hoop of the Z-family have at least two common faces (2-cells) whose normal is parallel to the Y-axis (for example, face 1 labelled with a cross in Fig. 8). One of the normal orientations must be chosen as the basic orientation.

The hoops are tracked by means of the algorithm described in Sect. 4.1. for two-dimensional images since a hoop lies in a two-dimensional slice. All faces with the basic orientation encountered during the tracking are recorded in a list, accompanied by a label of the hoop family.

When the tracking of a hoop finishes, a basic face is extracted from the list; the tracking along a hoop of the other family is started; and the face is deleted from the list. In this way any basic face is visited twice. Faces with other orientations are visited only once.

The surface may be economically encoded by means of the crack-codes of all the hoops being tracked. This code may be used for the approximation of the surface by planar patches, which is important for analysing geometrical properties of the surface.

Topological properties of a surface may be described by its genus as explained in Sect. 2.2. It may be shown that each component of the boundary of an object in a three- dimensional image is (under the usual conditions) a two- dimensional manifold, as defined in Sect. 2.2. Any such manifold is topologically equivalent to a sphere with G handles, G being the genus of the manifold.

To calculate the genus of a boundary component, it is necessary to count the 0-, 1- and 2-cells in the component. It is incorrect to count the cells directly in the crack-codes of hoops, since some cells are repeated in many hoops. The correct way consists of using the crack-codes to determine the faces (2-cells) belonging to the desired component while labelling the already counted cells in a three-dimensional array representing the image. Each element of the array must have 7 bits for the labels of one 0-cell, three 1-cells and three 2-cells assigned to the corresponding voxel, for example, by the "nearest-to-the-origin" membership rule. At the beginning of the count all the labels (bits) must be set to zero. Each boundary face is bounded by four 0-cells and four 1-cells. All nine cells belong to the desired boundary component. The bit corresponding to each of the nine cells must be found in the array and, if it is zero, the corresponding cell must be counted with the proper sign and the bit must be set to one.
The genus of a boundary component provides important information about the object.

6 Conclusion

This study has presented some concepts and algorithms for determining topological properties of subsets of two- and three-dimensional images. These properties

are considered to be important for analysing shapes. Geometrical properties of objects in digital images belong to the area of digital geometry which is founded on the basis of finite topology presented here. Elements of digital geometry with applications to shape analysis are presented in [9].

References

1. Alexandroff, P. (1937). Diskrete topologische Räume, Matematicheskii Sbornik 2 (44), Moscow, pp. 501–519.
2. Gordon, D., Udupa, J.K. (1989). Fast surface tracking in three-dimensional binary images, Computer Vision, Graphics and Image Processing 45, pp. 196–214.
3. Herman, G.T. (1990). On topology as applied to image analysis, Computer Vision, Graphics and Image Processing 52, pp. 409–415.
4. Herman, G.T., Webster, D. (1983). A topological proof of a surface tracking algorithm, Computer Vision, Graphics and Image Processing 23, pp. 162–177.
5. Kong, T.Y., Rosenfeld, A. (1991). Digital topology: a comparison of the graph-based and topological approaches. In: Reed, G.M., Roscoe, A.W., Wachter, R.F. (eds.), Topology and Category Theory in Computer Science, Oxford University Press, Oxford, U.K., pp. 273–289.
6. Kong, T.Y., Kopperman, R., Meyer, P.R. (1991). A topological approach to digital topology, American Mathematical Monthly 98, pp. 901–917.
7. Kovalevsky, V.A. (1989). Finite topology as applied to image analysis, Computer Vision, Graphics and Image Processing 46, pp. 141–161.
8. Kovalevsky, V.A. (1992). Finite topology and image analysis. In: Hawkes, P. (ed.), Advances in Electronics and Electron Physics, Academic Press, Vol. 84, pp. 197–259.
9. Kovalevsky, V.A. (1993). A new concept for digital geometry, this volume, pp. 37–51.
10. Pavlidis, T. (1977). Structural Pattern Recognition. Springer-Verlag, Berlin.
11. Pavlidis, T. (1982). Algorithms for Graphics and Image Processing. Computer Science Press, Rockville, USA.
12. Rosenfeld, A. (1970). Connectivity in digital pictures, Journal of the ACM 17, pp. 146–160.
13. Rosenfeld, A., Kak, A.C. (1982). Digital Picture Processing. Academic Press, New York, 1976 (Second Edition, 1982).
14. Seifert, H., Threlfall, W. (1980). A Textbook of Topology. Academic Press, New York.
15. Steinitz, E. (1908). Beiträge zur Analysis, Sitzungsbericht Berliner Mathematischer Gesellschaft 7, pp. 29–49.

A New Concept for Digital Geometry

Vladimir A. Kovalevsky

Technische Fachhochschule Berlin, Luxemburger Str. 10, 13353 Berlin, Germany

Abstract. A concept for geometry in a topological space with finitely many elements without the use of infinitesimals is presented. The notions of congruence, collinearity, convexity, digital lines, perimeter, area, volume, etc. are defined. The classical notion of continuous mappings is transferred (without changes) onto finite spaces. A slightly more general notion of connectivity preserving mappings is introduced. Applications for shape analysis are demonstrated.

Keywords: topological coordinates, continuous mapping, connectivity-preserving mapping, n-isomorphism, digital half-plane, digital straight-line segment, digital circular arc, perimeter, area, volume, cell list, polygon matching.

1 Introduction

Researchers in the areas of image processing and computer graphics have recently been placed in a strange situation. There is an increasing need to produce practically applicable results in the absense of adequate theory: there are many problems in image analysis which cannot be solved on the basis of classical Euclidean geometry. Consider as an example the problem of measuring the curvature of lines in digital images. All the knowledge of differential geometry turns out to be useless in this case. Another example is drawing digital polygons with some acute angles and filling their interiors: some vertices disappear, others induce stripes running through the whole image. Similar situations occur each time that some fine details of the image must be processed. The reason is that classical geometry is developed for working with point sets having infinitely many elements. According to the topological foundations of classical geometry, even the smallest neighbourhood of a point contains infinitely many other points. Therefore, classical geometry has no tools for working with single space elements, which is highly important in analyzing digital images. In such cases one gets the feeling that Euclidean geometry gives only an approximate description of geometric figures in the digital space, namely with a precision of plus or minus a few space elements. The feeling contradicts the common belief that classical geometry gives a precise description of figures, while every numeric description is an

approximate one. This is, of course, true as far as classical geometry is applied to a space of infinitely small space elements: an error of "a few infinitesimals" is less than an error of a single finite element. However, when applying both the classical and the digital approach to a space of finite elements, the digital approach gives a higher precision.

Some elementary concepts of self-contained digital geometry in a two- and three-dimensional space is presented. The concepts are kept, on the one hand, as close as possible to the practical demands of image processing and computer graphics and, on the other hand, as close as possible to the "macro-results" of Euclidean geometry which has been proven to describe adequately the real macro world. Therefore they are dissociated from such approaches to digital geometry which, for example, consider a digital line as a disconnected set of remote points, or use a non-Euclidean metric, or permit rotation of the space only by a multiple of 90° [5].

The present approach is based on the topology of abstract cell complexes, which is a special case of a finite T_0-topology in the classical sense of this notion. The theory is independent of Euclidean geometry as well as of Hausdorff topology: all the geometric notions are introduced anew and are based only on the notions of finite topology. Therefore geometrical figures in the digital space are not defined as results of digitizing some Euclidean figures. Digitization is considered as a transfer from a space with finer space elements to a coarser space. The theory of cell complexes will not be repeated here since it may be found in this volume [13].

In Sect. 2 the notion of a Cartesian finite space is introduced. This provides the possibility of defining "topological" coordinates *before* introducing a metric and the notion of a straight line. Section 3 introduces the notions of a half-plane, a digital straight-line segment, and collinearity. An algorithm for drawing curves as boundaries of regions defined by inequalities is presented. Section 4 is devoted to metric, circles, and spheres. Also the notion of congruence is introduced there. Section 5 describes mappings among finite spaces. It is shown here that it is impossible to describe all mappings important for applications by functions. The notions of continuous multivalued correspondence, connectivity-preserving mapping, and n-isomorphism are introduced. The notions are used to analyse the properties of digital geometric transformations. Section 6 presents methods of calculating perimeter, area, and volume. Section 7 describes some applications to shape analysis.

2 Finite Cartesian Spaces

Digital geometry must be developed in a finite topological space. As explained in [13], it is expedient to accept an *abstract cell complex* (ACC) as such a space. All properties of ACCs important for this presentation may be found in [13]. To analyse shapes in digital images, it is important to have coordinates in the corresponding space. A natural way of introducing coordinates in ACCs consists of constructing ACCs with some special simple structure as explained below.

First consider the finite number line as a one-dimensional ACC. There must be a linear order in the set of its cells and hence no branches in the ACC. (See [13, Definition 6] for the nonbranching condition). ACCs without branches are manifolds [13]. Thus, what we need is a connected subset of a one-dimensional manifold: it is a connected ACC in which each 0-cell, except two of them, has exactly two incident 1-cells. Such an ACC looks like a polygonal line whose vertices are the 0-cells and whose edges are the 1-cells (A_1 and A_2 in Fig. 1).

It is possible to assign subsequent integer numbers (in addition to dimensions) to the cells in such a way that a cell with the number x is incident with cells having the numbers $x-1$ and $x+1$. These numbers are considered as *coordinates* of cells in a one-dimensional space. ACCs of greater dimensions are defined as Cartesian products of such one-dimensional ACCs. A product ACC is called a *Cartesian* ACC. The set of cells of an n-dimensional Cartesian ACC C_n is the Cartesian product of n sets of cells of one-dimensional ACCs. These one-dimensional ACCs are the *coordinate axes* of the n-dimensional space. They will be denoted by A_i, $i = 1, 2, \ldots, n$. A cell of the n-dimensional Cartesian ACC C_n is an n-tupel $(a_1, a_2, \ldots, a_n)$ of cells a_i of the corresponding axes: $a_i \in A_i$. The bounding relation of the n-dimensional ACC C_n is defined as follows: the n-tupel $(a_1, a_2, \ldots, a_n)$ is bounding another distinct n-tupel $(b_1, b_2, \ldots, b_n)$ iff for all $i = 1, 2, \ldots, n$ the cell a_i is incident with b_i in A_i and $\dim(a_i) \leq \dim(b_i)$ in A_i. The dimension of a product cell is defined as the sum of dimensions of the factor cells in their one-dimensional spaces. Coordinates of a product cell are defined by the vector whose components are the coordinates of the factor cells in their one-dimensional spaces.

Consider the two-dimensional product ACC in Fig. 1. The 1-cell with coor-

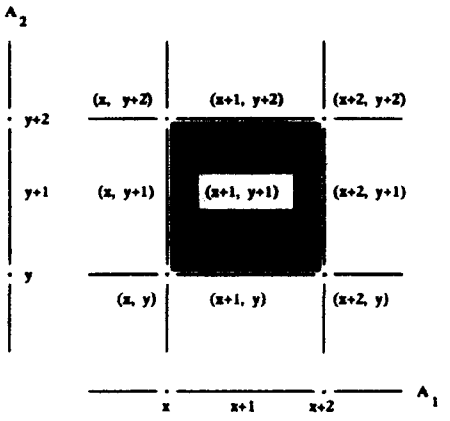


Fig. 1. Composition of a two-dimensional Cartesian ACC

dinates $(x+1, y+2)$ is a pair consisting of the 1-cell $x+1$ of the one-dimensional ACC A_1 and the 0-cell $y+2$ of the ACC A_2. The 2-cell $(x+1, y+1)$ of the product ACC consists of the 1-cells $x+1$ of A_1 and $y+1$ of A_2 etc.

Notice, that coordinates have been introduced without having introduced a metric, or the notion of a straight line, or the scalar product. Therefore it is

correct to call the coordinates *topological* ones. Similar spaces without regarding dimensions of space elements were considered by Khalimsky [6] (see also [7]). It is easy to see that a Cartesian ACC represents a finite analogue of a Cartesian Euclidean space.

The coordinate notation used by Khalimsky has the disadvantage that the size of a pixel, which is equal to the difference of the coordinate of the sides of the corresponding square, is equal to 2 rather than to 1 as is usual in image processing. There are two possibilities for overcoming this drawback. One of them consists of assigning the same integer to a 0-cell and to the next incident 1-cell of an axis. Dimensions of cells must then be coded by additional labels. This notation gives no possibility of expressing the fine difference in the location of a pixel and of one of the 0-cells incident with it. This sometimes leads to an undesired asymmetry of figures described by inequalities. For example, a digital circle, defined as a set of pixels whose distance to a point, that is, to a 0-cell, is limited by the given radius, is asymmetric with respect to the point.

The second possibility is to assign subsequent rational numbers with denominator 2 to subsequent cells of an axis. The size of a pixel is then equal to 1 and cells of different dimensions always have different coordinates. Under this notation the coordinates of a pixel in a two-dimensional space and, generally, of an n-cell c in an n-dimensional space, are equal to the arithmetic mean of the coordinates of all cells bounding c. Hence, fractional coordinates of an n-cell may be interpreted as coordinates of its "middle point". This prevents the imprecise definition of figures by inequalities. In the general case, coordinates may be *rational numbers* with any constant denominator, or floating point numbers while even mantissae correspond to 0-cells and odd mantissae to 1- cells. It is possible to achieve with this notation any required precision in determining the coordinates while preserving the possibility in recognizing the dimension of a cell from its coordinates. Let us consider in the sequel coordinates of cells of the axes as rational numbers with denominator 2. Then dimensions of cells may be recognized in the following way: the coordinates of 0- cells of the axes are integers and those of 1-cells are fractions. All n coordinates of a 0-cell of C^n are integers. All coordinates of an n-cell are fractions. A d-dimensional cell of C^n has d fractional and $n-d$ integer coordinates. The recognition of dimensions in the general case of an arbitrary denominator is similar to that just explained.

3 Linear Inequalities in the Two-Dimensional Space

For convenience, let us call the 0-cells of the space "points", the 1-cells "cracks" and the 2-cells "pixels". Some definitions are now introduced which are important for subsequent development.

Definition 1. A *region* is an open connected subset of the space. A region R of an n-dimensional ACC C^n is called *solid* if every cell $c \in C^n$ which is not in R is incident with an n-cell of the complement $C^n - R$.

Definition 2. A *digital half-plane* is a solid region containing all pixels of the space, whose coordinates satisfy a linear inequality.

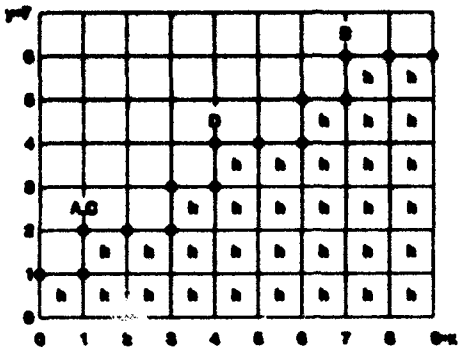


Fig. 2. Examples of a half-plane and a DSS

For example, Fig. 2 shows the half-plane defined by $2x - 3y + 2 > 0$. All pixels of the half-plane are labelled "h".

Definition 3. A non-empty intersection of digital half-planes is called a *digital convex subset* of the space.

Definition 4. A *digital straight-line segment* (DSS) is any connected subset of the boundary of a half-plane.

In Fig. 2 the cracks of the DSS composing the boundary of the half-plane "h" are drawn as thick lines and the points as black circles.

Definition 5. A point (0-cell) C is said to be *strictly collinear* with two other points A and B if

$$(x_c - x_b) \cdot (y_b - y_a) - (y_c - y_b) \cdot (x_b - x_a) = 0 \ .$$

It is said to lie *to the right* of the ordered pair of points A and B if

$$(x_c - x_b) \cdot (y_b - y_a) - (y_c - y_b) \cdot (x_b - x_a) > 0 \ .$$

It lies *to the left* of A and B if

$$(x_c - x_b) \cdot (y_b - y_a) - (y_c - y_b) \cdot (x_b - x_a) < 0 \ .$$

Consider all ordered pairs of points of a DSS, such that all other points of the DSS do not lie to the left of the pair. Choose the pair (A, B) with the greatest absolute difference of the coordinates $x_b - x_a$ or $y_b - y_a$ (Fig. 2). If there are points of the DSS which are strictly collinear with A and B, choose the pair of such points which are closest to each other. Denote the points C and D. This point pair is called the *right base of the DSS*. The left base may be defined similarly. The slope M/N of the base is defined by two integers:

$$M = y_d - y_c \qquad \text{and} \qquad N = x_d - x_c \ .$$

In the example of Fig. 2, $M = 2$, $N = 3$. Owing to the choice of points that are closest and strictly collinear with A and B, the fraction M/N is irreducible. From the definition of the DSS as a boundary of a half-plane (Definition 4) and the definition of the boundary [13], it follows that every point (x, y) of the DSS satisfies the following inequalities:

$$0 \le (x - x_c) \cdot M - (y - y_c) \cdot N \le |M| + |N| - 1 \ . \tag{1}$$

Note that x, y, x_c, y_c, M, and N are all integers. These inequalities are used for the fast recognition of DSSs [10].

Definition 6. A two-dimensional vector with integer components (x, y) is called *right semi-collinear* with another integer vector (n, m) if the following inequalities hold:

$$0 \le (x \cdot M - y \cdot N) \le |M| + |N| - 1 \ ,$$

where M and N are numerator and denominator of the irreducible fraction $M/N = m/n$.

The notion of left semi-collinear vectors may be defined similarly.

By means of this definition, a DSS with a given base (C, D) may be defined as a digital curve K (connected subset of a one-dimensional manifold, see [13]) such that each point P of K composes with one of the end points of the right base (say, C) a vector $(P - C)$ left semi-collinear with the vector $(D - C)$ of the right base. A similar definition is possible when using the left base.

One of the simplest methods to draw a DSS in a two-dimensional ACC consists of tracking the linear inequality defining the corresponding half-plane. For this purpose the tracking algorithm described in [13] may be used. To adapt the algorithm for tracking an inequality rather than an object in a binary image, the tests of the two pixels L and R for their membership in the object must be replaced by the test of whether the half-integer coordinates of the pixels satisfy the inequality. The tracking may be made faster when calculating the increments of the left side of the inequality rather than the expression itself. The calculation becomes still simpler when transforming the desired DSS to one lying in the first octant. This modification of the tracking corresponds to the famous Bresenham algorithm [2].

The tracking technique may be used to draw boundaries of regions defined by *any inequalities*, also non-linear, for example, circles, parabolas, etc.

4 Metric, Circles, and Spheres

Only the Euclidean metric may be used in digital geometry. This is necessary to obtain results as close as possible to those of classical geometry. Correspondingly, the distance $D(A, B)$ between two points (cells) A and B is declared to be equal to

$$D(A, B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \ ,$$

A_i and B_i being the ith coordinates of the corresponding points in an n-dimensional Cartesian space as defined in Sect. 2.

Having defined the distance, we may immediately specify the inequality of a digital disk in the two-dimensional space:

Definition 7. A *digital disk* is a solid region containing all pixels of the space, whose coordinates satisfy the following inequality:

$$(x - x_c)^2 + (y - y_c)^2 < R^2 \ , \tag{2}$$

where x and y are the half-integer coordinates of pixels, x_c and y_c are the coordinates of the centre, R is the radius of the disk. The values of x_c, y_c and R may be either integer or fractional.

Definition 8. A *digital circular arc* (DCA) is any connected subset of the boundary of a digital disk.

To draw a DCA, the technique of tracking the boundary of an inequality, as described in the previous section, may be used. As in the case of a line, the tracking may be made faster when calculating increments of the left side of 2 rather than the expression itself and when restricting the set of possible step directions according to the known octant of the arc. This modification of tracking is wellknown in computer graphics as the Bresenham arc algorithm [3]. Recognition of DCA is described in [10].

In a similar way, digital balls and spheres (as boundaries of balls) may be defined in the three-dimensional space. Tracking surfaces in three-dimensional binary images is described in [13]. The same technique may be used to track the surface of an arbitrary body defined by an inequality.

The notions of distance and collinearity may be used to introduce that of congruence:

Definition 9. The distance d between two points is declared *digitally equal* to a number n, if the absolute difference between d and n is less than or equal to length of a pixel's diagonal ($\sqrt{2}$ under the accepted notation).

Definition 10. The *value of semi-collinearity* of a point C relative to an ordered pair of points A and B is declared to be 0 if C is semi-collinear with (A, B). If it is not semi-collinear, then the value is declared to be -1 or $+1$ depending on whether C lies to the left or to the right of (A, B) according to Definition 5.

Definition 11. Two figures F and G are called *congruent* with each other iff there exists such a mapping from F to G that the distance between any two cells of G is digitally equal to the distance of their pre-images in F, and the value of semi-collinearity of any three points of G is the same as that of their pre-images in F.

The mapping is not necessarily a bijection. The class of considerable mappings called CPM is described in the next section.

5 Mappings Among Finite Spaces

Mappings among finite spaces are rather different from those among infinite spaces. Consider the simplest example of mapping a one- dimensional finite space X onto another such space Y by a function. A function must assign one cell of Y to each cell of X. Consider a function F and a subset S of Y consisting of two incident cells of Y having the coordinates y and $y + 1/2$ [13]. The pre-image $F^{-1}(S)$ must consist of at least two different cells, since the function is single-valued. The difference D_y between the values of y is equal to $1/2$, while the difference D_x between the extreme values of x in $F^{-1}(S)$ is greater than or equal to $1/2$. Hence the average slope D_y/D_x of F cannot be greater than 1. Thus a problem arises: functions mapping one finite space into another such space cannot have a slope greater than 1. If we decide to restrict ourselves to such functions, the problem is still unsolved, since there is no possibility of considering inverse functions, which in this case must have a slope greater than or equal to 1.

The only possible solution is to consider more general correspondences between X and Y, assigning to each cell of X a *subset of* Y rather than a single cell.

5.1 Connectivity-Preserving Correspondences

A correspondence between X and Y or a multi-valued mapping of X into Y is a subset F of ordered pairs (x, y) containing all cells $x \in X$ and some cells $y \in Y$. There is a difference between a correspondence and a binary relation: in the case of a relation the sets X and Y must be identical. A function is a special case of a correspondence: a correspondence is a function if any value of $x \in X$ is encountered in exactly one pair (x, y) of F. Given a correspondence F, the set of all y encountered in pairs of F containing a fixed x is called the *image* of x. The set of all x encountered in pairs of F with a fixed y is called the *pre-image* of y. The union of the images of all x of a subset SX of X is called the image of SX. Similarly, the union of the pre-images of all y of a subset SY of Y is called the pre-image of SY.

A correspondence may be continuous in the classical sense of the notion if the pre-image of any open subset of Y is open (for example, G in Fig. 3). Coordinates in Fig. 3 are denoted by their numerators, to make the notation simpler. However, in finite mathematics another class of correspondence is important.

Definition 12. A correspondence between X and Y is called a *connectivity-preserving mapping* (CPM) if the image of any connected subset of X is connected.

An example of a CPM is F in Fig. 3. It is easy to see that every continuous correspondence is a CPM but not vice versa.

Let us denote by $V(x, y)$ the connected component of $F(x)$ containing y, and by $H(x, y)$ the connected component of $F^{-1}(y)$ containing x.

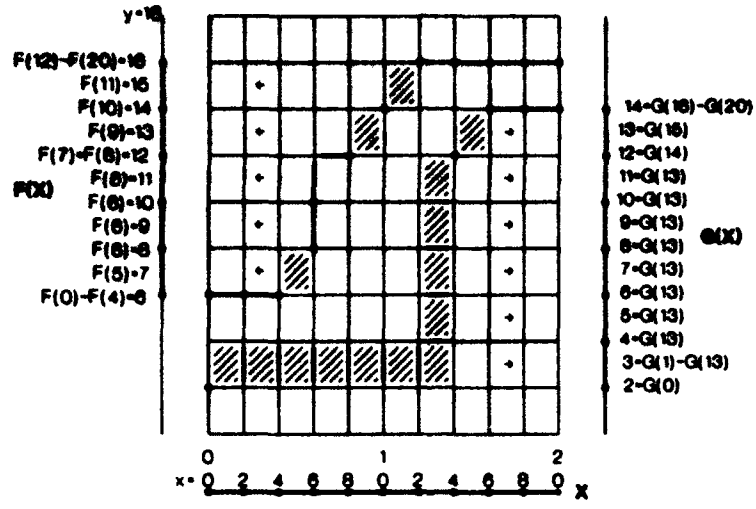


Fig. 3. Examples of correspondences: F is connectivity-preserving, simple, and not continuous; G is continuous and not simple.

Definition 13. A correspondence F is called *simple* if for each pair $(x, y) \in F$ at most one of the sets $V(x, y)$ and $H(x, y)$ contains more than one element.

For example, F in Fig. 3 is a simple CPM, while G is not simple since for the pair $x = 13$, $y = 3$ both $V(x, y)$ and $H(x, y)$ contain more than one cell. In the sequel, we shall consider mainly simple CPMs which are the substitutes of continuous mappings in finite spaces.

Consider some more examples. The translation $y = x + a$ with integer constant a maps a subset of X onto a subset of Y in such a way that a 0-cell is mapped onto a 0-cell and a 1-cell onto a 1-cell. Thus the bounding relation [13] is preserved. Such a mapping is an isomorphism. However, if we consider a magnification, say by a factor of two, we cannot describe it as $y = 2x$, since this transformation maps the cells of X onto each second cell of Y while the other cells of Y remain uncovered by the image of X.

To perform a true magnification, each cell of X must be mapped onto several cells of Y. To magnify a two-dimensional picture X by a factor of M, each pixel of X must be mapped onto a solid region containing $M \times M$ pixels of the picture Y. Thus magnification must be a multi-valued mapping. On the other hand, a reduction by the factor M maps a solid region of $M \times M$ pixels of X onto a single pixel of Y. Thus reduction is a contractive mapping.

Consider now the rotation of a two-dimensional image with pixel coordinates x and y by an arbitrary angle α. The simplest version of the rotation is defined by the wellknown formulae:

$$x' = \text{Round}\,(x \cdot \cos\alpha - y \cdot \sin\alpha)\ ,$$
$$y' = \text{Round}\,(x \cdot \sin\alpha + y \cdot \cos\alpha)\ ;$$

where the rounding-off operation "Round" is necessary to convert the trans-

formed coordinates x', y' to coordinates of pixels, that is, half-integers. It may easily be shown, that this transformation maps some pairs of adjacent pixels of the input image onto one pixel of the output. Thus it is a contractive mapping. When using the more perfect "anti-aliasing" rotation, a grey value of an output pixel is calculated as a function of the grey values of four adjacent input pixels. Such a rotation must be considered as a mapping which is simultaneously contractive and multi-valued. In any case it is not an isomorphism. However, it is *approximately* an isomorphism. Let us give this assertion a precise meaning.

5.2 The Notion of n-Isomorphism

The notion of the smallest open neighbourhood (SON) of a cell in an ACC was presented in [13]. Two more notions which are needed to define the n-isomorphism are now introduced.

Definition 14. The *closed hull* (closure) $\mathrm{Cl}(S)$ of a subset S of an ACC C is the smallest closed subset of C containing S.

Definition 15. The *open hull* $\mathrm{Op}(S)$ of a subset S of an ACC C is the smallest open subset of C containing S.

Examples are shown in Fig. 4.

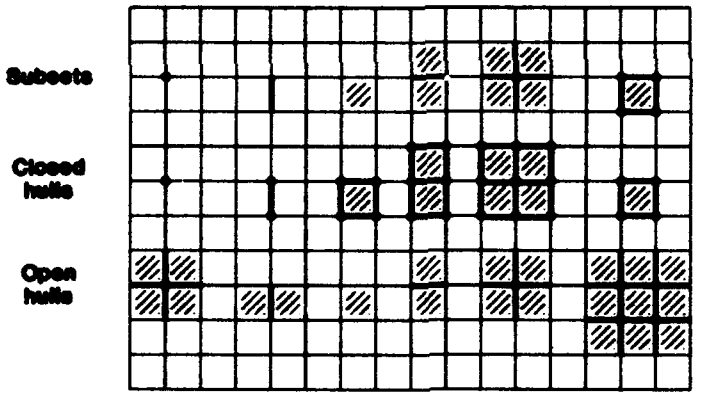


Fig. 4. Examples of some subsets, their closed and open hulls

Definition 16. The n-neighbourhood $U_n(c)$ of a cell $c \in C$ is an open subset of C satisfying the following conditions:

(1) $U_0(c) = \mathrm{Op}(c) = \mathrm{SON}(c)$– the smallest open neighbourhood of c;

(2) $U_{n+1}(c) = \mathrm{Op}(\mathrm{Cl}(U_n(c)))$.

Examples are given in Fig. 5.

The notion is now introduced of an n-isomorphism as a multi-valued mapping that approximately preserves the bounding relation of the cells in an ACC: it

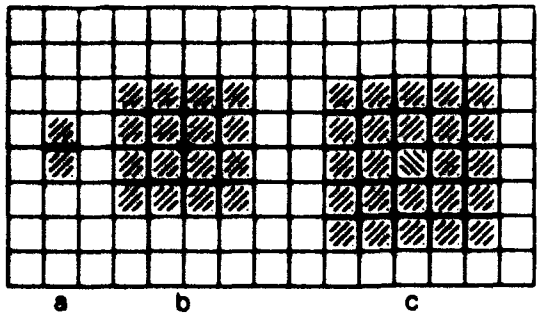


Fig. 5. Examples of n-neighbourhoods: (a) a 0-neighbourhood of a 1-cell, (b) a 1-neighbourhood of a 0-cell, and (c) a 2-neighbourhood of a 2-cell.

maps two *incident* cells onto cells that are *not too far away* from each other. In contrast, two cells that are *far away* from each other must *not* be mapped onto *incident* cells. Note that the bounding relation in ACCs may be expressed in terms of SONs: if a cell c_1 bounds another cell c_2 then $c_2 \in \mathrm{SON}(c_1)$. The cell c_1 "approximately bounds" the cell c_2 if c_2 is in a greater neighbourhood of c_1. Thus we introduce

Definition 17. A multi-valued mapping $F : X \to Y$ from a finite space X into a finite space Y is called *n-isomorphism* if for any two cells x_1, x_2 of X and for any cells of the images of them $y_1 \in F(x_1)$, $y_2 \in F(x_2)$ the following two conditions are satisfied:

(1) $x_2 \in U_0(x_1) \Rightarrow y_2 \in U_n(y_1)$;
(2) $x_2 \notin U_n(x_1) \Rightarrow y_2 \notin U_0(y_1)$.

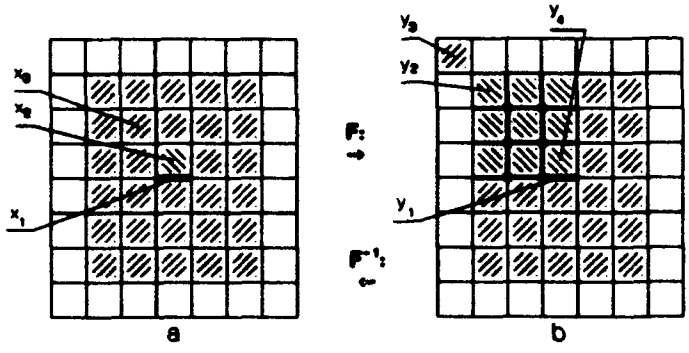


Fig. 6. Illustration to Definition 17: a triple magnification $F : a \to b$ and a triple reduction $F^{-1} : b \to a$.

Figure 6 illustrates these conditions for the cases of a triple magnification and triple reduction of a two-dimensional space (these transformations were defined in Sect. 5.1). The cell y_2 in Fig. 6b is an element of the image $F(x_2)$ (dark shaded area). Similarly, the cell y_1 is an element of $F(x_1)$. The cell x_2 is bounded by

x_1, that is, x_2 belongs to $U_0(x_1)$ (compare with Fig. 5a). Correspondingly, y_2 belongs to $U_2(y_1)$ which is represented by the light shaded area in Fig. 6b.

Figure 6 simultaneously illustrates a triple reduction F^{-1} as a mapping from Fig. 6b into Fig. 6a. The first condition of Definition 17 is illustrated by the cells y_1 and y_4: $y_4 \in U_0(y_1)$ and correspondingly, $x_2 = F^{-1}(y_4) \in U_2(x_1)$ with $x_1 = F^{-1}(y_1)$. The second condition is illustrated by the cells y_1 and y_3: $y_3 \notin U_2(y_1)$ and correspondingly, $x_3 = F^{-1}(y_3) \notin U_0(x_1)$ with $x_1 = F^{-1}(y_1)$.

It may be shown that a magnification and a reduction with the factor M are both $(M-1)$-isomorphisms. A rotation by an arbitrary angle is a 1-isomorphism. The notion of n-isomorphism gives us the possibility of quantitatively estimating topological distortions caused by various mappings. Thus, for example, a rotated digital straight line is no more a straight line but its deviation from a digital straight line does not exceed 1 pixel, since rotation is a 1-isomorphism.

6 Metrical Properties of Figures

Properties such as area, volume, perimeter must be independent of translations and rotations of a figure. Consider first the two-dimensional space. The commonly used measure of the area of a region in a two-dimensional space is the number of pixels. It may be demonstrated that this measure may slightly vary under rotation. However, the difference of the areas before and after rotation increases linearly with the scale, while the area itself increases quadratically. Therefore the relative change tends to zero when the pixel size becomes smaller and smaller relative to the size of the area.

Different behaviour is demonstrated by the commonly used measures of the perimeter [15]. It was demonstrated in [11], both theoretically and experimentally, that all perimeter measures known from the literature contain systematic errors depending on the rotation of the figure, which do not disappear when the pixel size decreases. It was also demonstrated that the following perimeter definition is free from this imperfection.

Definition 18. The *perimeter* of a region R in a two-dimensional finite space is the sum of the lengths of subsequent DSSs obtained by subdividing the boundary of R into as few as possible DSSs.

It was shown in [11] that this perimeter estimate is invariant with respect to rotation: the absolute difference between the perimeters before and after rotation by an arbitrary angle tends to zero when the size of the pixels (relative to the diameter of the region) decreases.

In a three-dimensional finite space the perimeter of a closed digital curve, may be defined in the same way. Properties of DSSs in a three-dimensional space are described in [1]. Estimation of the area of a two-dimensional surface in a three-dimensional space is a problem still more difficult than that of the perimeter. The author supposes that a surface must be dissolved into maximum patches of digital planes and the areas of the patches must be added. However, defining the area of a subset of a plane in the three-dimensional space is itself a

non-trivial problem. By no means must the area be determined as the number of facets (2-cells) in the subset. Such an estimate would have the same imperfection as the perimeter estimate by the number of cracks in a two-dimensional space: the estimate would not be rotation invariant. The area of a plane patch in space must be determined as the area of a plane polygon by means of the coordinates of its vertices. The coordinates must be determined by means of a procedure for recognising digital plane patches in space, similar to the recognition of the DSSs. Unfortunately, no such algorithm is known to the author.

On the other hand, the estimate of the *volume* of a three-dimensional region in a three-dimensional space as the number of voxels in the region is supposed to be rotation invariant: the error tends to zero as the size of voxels decreases. Probably, this is the property of the measure of an n-dimensional subset of an n-dimensional space for any n.

7 Application to Shape Analysis

7.1 The Cell List Data Structure

Finite topology suggests a new means for an efficient coding of images: the *cell list* [8, 12]. This data structure makes the calculation of geometrical features and topological relations of objects of interest fast and easy. A segmented two-dimensional image is described in the cell list by a collection of sublists: the sublist of regions, of boundary curves, and of branching points. Branching points are the locations where three or more regions meet. Each element of a sublist is provided with pointers indicating which other elements are bounding it or are bounded by it. The bounding relation is similar to that of the ACCs [13] and is based on the theory of block complexes familiar in topological literature. This topological part of the structure may be regarded as a *generalization of the well-known region adjacency graph* [16]. The topological part is augmented by metric data consisting of coordinates of the branching points and some intermediate points of the boundary curves which define the location of the curves. Intermediate points are determined by dissolving the curves into DSSs of maximum length [4, 10] or by approximating the curves by *polygonal lines* with a desired tolerance. Thus each boundary curve is represented as a digital *polygonal line.*

7.2 Polygon Matching

One of the methods useful for analyzing shapes consists of describing the objects of interest by polygons and in comparing the polygons with some prepared prototype polygons. The corresponding problem statement is as follows:
Given: A polygon to be recognized and some prototype polygons.
Find: A prototype polygon and a set of transformation parameters, such that the Hausdorff distance between the given polygon and the transformed prototype polygon be minimal.

The difference of two polygons is defined as the minimum squared distance of their vertices. The polygons must be converted in such a form that they have the

same number of vertices: otherwise a one-to-one mapping of the sets of vertices is impossible. The squared distance is defined as

$$SD = \sum_{i=1}^{n} \left((X_i' - X_i'')^2 + (Y_i' - Y_i'')^2 \right) \ ;$$

where (X_i', Y_i') and (X_i'', Y_i'') are the coordinates of the ith vertex of the first and second polygons respectively; n is the common number of vertices.

The squared distance SD must be minimized over all geometrical transformations of the prototype polygon. The transformations are: translation, rotation, and magnification. The minimization is performed by the classical method of least squares. The problem of making the number of vertices equal was solved by some heuristics. The solution of this problem and of a more complex modification with non-quadratic functionals has been successfully used for the recognition of hand-written characters, analyzing hand-made and technical drawings, digitizing topographical maps, and others [9, 12].

Consider the case when the image to be analysed contains overlapping objects that originate from orthogonal transformations of known prototypes. The statement of the problem is as follows:
Given: One scene polygon SP and many prototype polygons $PP_1, PP_2, \ldots, PP_m$.
Find: The minimum number of prototype polygons and the parameter of their geometrical transformations such that the "overlap" of the transformed prototype polygons form a polygon close to SP.

"Overlap" of polygons should be understood as the boundary of the union of their interiors (Fig. 7). "Close" means that the Hausdorff distance should be less than a given tolerance.

Fig. 7. The tolerance tube of a scene polygon and the optimally transformed prototypes found by the program

An approximate solution of this problem [14] is based on dissolving SP and the PPs into pieces which contain no concavities. Certain geometrical features of the pieces are used to find prototype pieces similar to the scene pieces. Transformation parameters are found by the least-squares method while minimizing the squared distance between two pieces. Then the whole prototype is transformed by the found transformation and compared with SP : its parts which are outside of SP must lie in the tolerance tube around SP (the shaded area in

Fig. 7). If this is the case, the prototype with its transformation is accepted as a candidate. The minimum number of candidates completely covering SP is found by dynamic programming. Figure 7 shows a scene composed of two overlapping objects (figures from a children's game "the blind cow") and the transformed prototypes as polygons inside the tolerance tube.

References

1. Andersen T.A., Kim C.E. (1985). Representation of digital line segments and their preimages, Computer Vision, Graphics and Image Processing 30 (3), pp. 279–288.
2. Bresenham J.E. (1965). Algorithm for computer control of a digital plotter, IBM Systems Journal 4 (1), pp. 25–30.
3. Bresenham J.E. (1977). A linear algorithm for incremental digital display of circular arcs, Communication of the ACM 20 (2), pp. 100–106.
4. Freeman H. (1974). Computer processing of line-drawing images, Comput. Surv. 6, pp. 57–97.
5. Huebler A. (1991). Diskrete Geometrie fuer die digitale Bildverarbeitung, Dissertation, University of Jena, Germany.
6. Khalimsky E. (1977). Ordered Topological Spaces (in Russian). Naukova Dumka, Kiev.
7. Kopperman R. (1993). The Khalimsky line as a foundation for digital topology, this volume, pp. 3–20.
8. Kovalevsky V.A. (1989). Finite topology as applied to image analysis, Computer Vision, Graphics and Image Processing 46, pp. 141–161.
9. Kovalevsky V.A. (1989). Zellenkomplexe in der Kartografie, Bild und Ton (9, 10) Germany, pp. 278–280, 312–314.
10. Kovalevsky V.A. (1990). New definition and fast recognition of digital straight segments and arcs, Proc. 10th Int. Conf. on Pattern Recognition, Atlantic City, June 17–21, IEEE Press, Vol. II, pp. 31–34.
11. Kovalevsky V.A., Fuchs S. (1992). Theoretical and experimental analysis of the accuracy of perimeter estimates. In: Förster, Ruwiedel (eds.), Robust Computer Vision, Wichmann Karlsruhe, pp. 218–242.
12. Kovalevsky V.A. (1992). Finite topology and image analysis. In: Hawkes, P. (ed.), Advances in Electronics and Electron Physics, Academic Press, Vol. 84, pp. 197–259.
13. Kovalevsky V.A. (1993). Topological foundations of shape analysis, this volume, pp. 21–36.
14. Reinecke M. (1991). Object Recognition in Two-Dimensional Binary Images, Graduation thesis, Technical College Berlin (TFH).
15. Rosenfeld A., Kak A.C. (1982). Digital Picture Processing. Academic Press, New York San Francisco London.
16. Strong J.P., Rosenfeld A. (1973). Region adjacency graphs, Communications of the American Computer Machinery 4, pp. 237–246.

Theoretical Approaches to N-Dimensional Digital Objects

Klaus Voss

Friedrich-Schiller-University Jena, Department of Mathematics and Informatics, UHH 17.OG, 07743, Jena, Germany

Abstract. The paper presents a unified and general theory of objects in n-dimensional orthogonal lattices as used in image processing. In contrast to set-theoretical topology (cellular complexes), the theory of incidence structures (see Beutelspacher, Einführung in die endliche Geometrie I, Wissenschaftsverlag, Mannheim, 1982) is developed consistently. New object quantities beside the Euler number are introduced, some inequalities between these quantities are derived, and an effective algorithm for surface detection is presented.

Keywords: digital topology, shape description, incidence structure, surface detection, similarity of digital objects, Euler number.

1 Introduction

As already mentioned in 1983 by Klette, the theory of discrete spaces useful in image processing can be developed using two basic notions [6]. The first is the theory of n-dimensional cellular spaces. The n-dimensional cells are defined by the n-dimensional unit cubes which build up the space.

The second notion is that of the n-dimensional *grid point space* (the n-dimensional *cubic lattice*). Here, the lattice points are the basic elements of the theory. This approach is preferred because the number theory offers many results which are useful or at least of interest for image processing problems.

But both approaches are possible ways of building up a theory of image processing. They are dual theories as the following table shows:

	cellular space C^3	grid point space $\mathbb{Z}^3$
basic elements	unit cubes	lattice points
boundaries of basic elements	faces	edges
boundaries of boundaries	edges	faces
constructed elements	points	cubes

Definition 1. An *incidence structure* $\Sigma = [E, I]$ is given by a set $E = \bigcup_{i=0}^{n} E_i$ of elements and a reflexive and symmetrical *incidence relation* $I \subset \bigcup_{i,k} E_i \times E_k$. The sets E_i are pairwise disjoint. The elements of set E_i are called i-dimensional. The number n is the *dimension* of the incidence structure.

Definition 2. In a finite incidence structure, there are non-negative numbers $b_{kl}(e)$ for each k-dimensional element $e \in E_k$ which describe the number of l-dimensional elements $e' \in E_l$ with $(e, e') \in I$. The numbers b_{kl} are called *structure constants*.

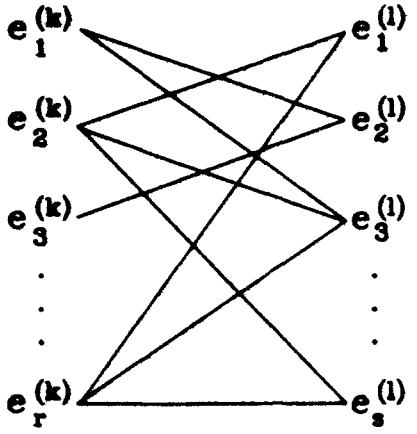


Fig. 1. Incidence relations between k-dimensional elements and l-dimensional elements

The connection between k-dimensional elements and l-dimensional elements given by the incidence relation I is represented in Fig. 1. The line between a k-dimensional element $e(k)$ and an l-dimensional element $e(l)$ symbolizes that $(e(k), e(l)) \in I$.

This representation is useful also for $k = l$. But it must be taken into account that, because of the reflexivity of relation I, there are loop-lines from each element $e(k)$ to itself. Therefore, we always set $b_{kk} = 1$. Counting lines in Fig. 1, we obtain

$$\sum_{e \in E_k} b_{kl}(e) = \sum_{e' \in E_l} b_{lk}(e')$$

for any k and l with $0 \leq k, l \leq n$ with n as the dimension of the incidence structure. This formula is called a *matching theorem*. With respect to image processing, this is the only important formula for all n-dimensional incidence structures. It should be pointed out that the notion of incidence structure sometimes also has another meaning. In [1], the matching formula is denoted as the "principle of double counting", and it is used for general one-dimensional incidence structures which are called "designs". For one-dimensional structures, the *node theorem*

$$\sum_{p \in P} \nu(p) = \sum_{e \in E_0} b_{01}(e) = \sum_{e \in E_1} b_{10}(e') = \kappa \tag{1}$$

of graph theory follows with $\nu(p)$ as the number of directed edges outgoing from points $p \in P$, and κ as the number of all undirected edges of an undirected graph.

2 Homogeneous Incidence Structures

Essential new insights are obtained only if we restrict ourselves to homogeneous structures:

Definition 3. An incidence structure is called *homogeneous* if $b_{kl}(e) = b_{kl} =$ const for all elements $e \in E_k$ and all k, l with $0 \leq k, l \leq n$.

For finite n-dimensional incidence structures, $a_k = card(E_k)$ is the number of k-dimensional elements of the structure. Then, we obtain from the matching theorem the *matching formulae*

$$a_k b_{kl} = a_l b_{lk} \quad \text{for} \quad 0 \leq k, l \leq n \ . \tag{2}$$

Given the number a_0 of points, we obtain $a_k = a_0 b_{0k}/b_{k0}$. Because a_0 and a_k must be integer numbers, not all integers a_0 are allowed in general. Therefore the above relationships are a nonlinear diophantic equation system. Using the formulae $a_k = a_m b_{mk}/b_{km}$ and $a_l = a_m b_{ml}/b_{lm}$, it follows that $a_m b_{mk} b_{kl}/b_{km} = a_m b_{ml} b_{lk}/b_{lm}$; therefore with $a_m > 0$ the following relationships are obtained:

$$\frac{b_{kl} b_{lm}}{b_{ml} b_{lk}} = \frac{b_{km}}{b_{mk}} \quad \text{for} \quad 0 \leq k, l, m \leq n \ . \tag{3}$$

These formulae and the generalizations

$$\frac{b_{kr} b_{rs} \ldots b_{tl}}{b_{lt} \ldots b_{sr} b_{rk}} = \frac{b_{kl}}{b_{lk}}$$

are combinatorial laws for the structure constants which are based only on the matching theorem and on the requirement of homogeneity.

It is certainly meaningful to represent explicitly some structures to illustrate the notion of a structure and to give certainty that there are incidence structures in fact. One-dimensional incidence structures are represented in Fig. 2 (see also 1). We will characterize such structures by tuples $(a_0, a_1, b_{00}, b_{01}, b_{10}, b_{11})$. The matching formula requires that $a_0 b_{01} = a_1 b_{10}$ or $\varepsilon.\nu = \kappa.2$ if we assume $b_{10} = 2$ because of geometrical reasons (ε is the number of points, κ is the number of edges, and ν is the number of edges incident to one point). Now, if the tuple $(\varepsilon, \nu, \kappa)$ is a solution of this diophantic equation then $(c\varepsilon, \nu, c\kappa)$ and $(\varepsilon, c\nu, c\kappa)$ with $c > 0$ are also solutions. In the first case, we get a c-fold repe[illegible]on of the structure $(\varepsilon, \nu, \kappa)$ which is only a meaningless and tedious construction. In the second case, we get graphs with multiple edges—also a solution without interest for the theory of neighbourhood graphs and image processing.

The complete graphs $(\varepsilon, \nu, \kappa) = (c, c-1, c(c-1)/2)$ are also incidence structures because $c(c-1)$ is always divisible by 2. But we will not search here for all possible solutions. It is important solely for image processing that the homogeneous incidence structures $(\varepsilon, \nu, \kappa) = (c, 2, c)$ with any large values of c are the base for one-dimensional image processing.

Two-dimensional incidence structures are characterized by a 12-tuple of numbers a_0, a_1, a_2 and $b_{00}, b_{01}, b_{02}, b_{10}, b_{11}, b_{12}, b_{20}, b_{21}, b_{22}$. In Fig. 3, these numbers are given for two examples—an octahedron and a cube (hexahedron).

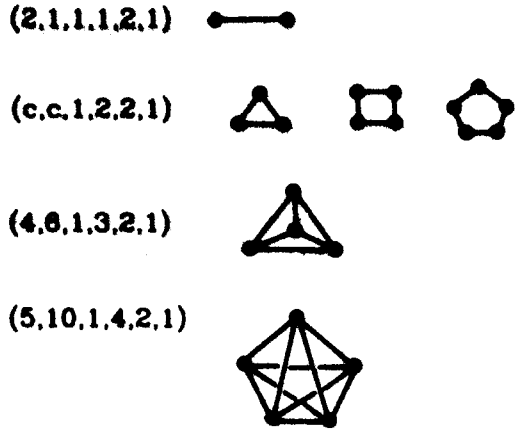


Fig. 2. One-dimensional incidence structures

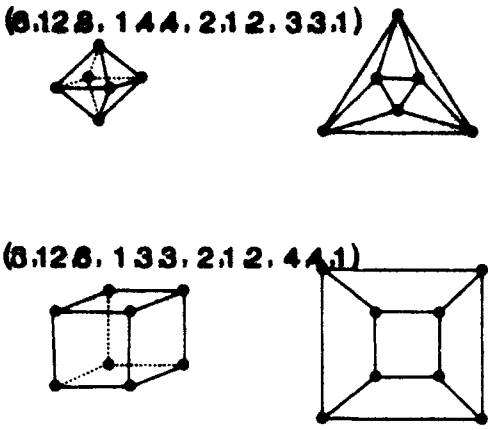


Fig. 3. Two-dimensional incidence structures

In the left of Fig. 3, pseudo-three-dimensional representations of these polyhedrons are shown where the polygonal sides of the polyhedrons are the two-dimensional elements of the incidence structures. The right-hand diagrams show the corresponding planar oriented neighbourhood structures where the meshes are the two-dimensional elements.

Using the following geometrica' y motivated relations $b_{10} = b_{12} = 2$, $b_{01} = b_{02}$, $b_{20} = b_{21}$, $b_{00} = b_{11} = b_{22} = 1$, we can characterize all two-dimensional incidence structures by tuples $(a_0, a_1, a_2, b_{00}, \ldots, b_{22}) = (\varepsilon, \kappa, \omega, 1, \nu, \nu, 2, 1, 1, \lambda, \lambda, 1)$, where ω is the number of meshes, and λ is the length of meshes. Using the five quantities ε, κ, ω, ν and λ), we obtain three relationships given by the matching formulae:

$$a_0 b_{01} = a_1 b_{10} \Rightarrow \varepsilon.\nu = \kappa.2 \ ,$$
$$a_0 b_{02} = a_2 b_{20} \Rightarrow \varepsilon.\nu = \omega.\lambda \ ,$$
$$a_1 b_{12} = a_2 b_{21} \Rightarrow \kappa.2 = \omega.\lambda \ .$$

There are many two-dimensional incidence structures of this kind (see [12, 14]). But as in the one-dimensional case, only a few of these structures are suited for image processing, namely the toroidally closed lattices with $(\nu, \lambda) = (3, 6)$, $(4, 4)$, and $(6, 3)$, respectively.

Finally, it will be shown that three-dimensional incidence structures also exist. Such structures are characterized mainly by four numbers $(a_0, a_1, a_2, a_3) = (\varepsilon, \kappa, \mu, \zeta)$ of points, edges, meshes, and cells. Further, we have 16 structure constants and 6 matching relations between them. Because these relations will be investigated in the following sections in general, here only a simple example is given for demonstration (Fig. 4). The structure constants of this example are

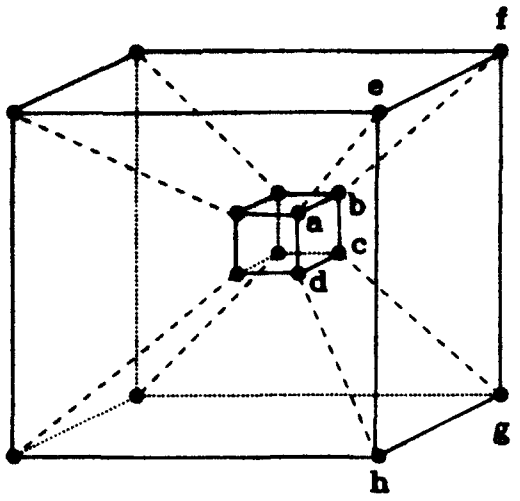


Fig. 4. Three-dimensional incidence structure

$$(b_{00}, b_{01}, b_{02}, b_{03}) = (1, 4, 6, 4) \ , \qquad (b_{10}, b_{11}, b_{12}, b_{13}) = (2, 1, 3, 3) \ ,$$
$$(b_{20}, b_{21}, b_{22}, b_{23}) = (4, 4, 1, 2) \ , \qquad (b_{30}, b_{31}, b_{32}, b_{33}) = (8, 12, 6, 1) \ .$$

This incidence structure contains $a_0 = \varepsilon = 16$ points $a, b, c, \ldots, a_1 = \kappa = 32$ edges (for example $\{a, b\}$ or $\{b, f\}$), $a_2 = \mu = 24$ meshes as for instance $< a, b, c, d >$ or $< a, b, f, e >$, and $a_3 = \zeta = 8$ cells like that built up by points a, b, c, d, e, f, g, h. Thereby the "outlying" cell is also counted, similarly to the "outlying" meshes in the case of the right-hand diagrams of Fig. 3.

3 $\mathbb{Z}^n$ as Incidence Structure

Some further general conclusions can be derived from the matching formulae. Assuming that these formulae for all values $k, l \leq n$ are valid, the formulae are also fulfilled for all $k, l \leq n-1$. Therefore, we get an $(n-1)$-dimensional incidence structure Σ_{n-1} from an n-dimensional incidence structure Σ_n by neglecting the n-dimensional elements in Σ_n.

Thus, the structure of Fig. 4 is, by neglecting the cells, a two-dimensional incidence structure with $(\varepsilon, \kappa, \mu, \nu, \lambda) = (16, 32, 24, 4, 4)$ which cannot be drawn in the plane without line crossing. Further, we get by neglecting cells and meshes a regular (non-planar) graph with $\varepsilon = 16$ points, $\kappa = 32$ edges, and with a neighbourhood degree $\nu = 4$. Therefore, we can formulate the following theorem:

Theorem 4. *Consider an n-dimensional incidence structure Σ_n. Neglecting all k-dimensional elements with $n' < k \leq n$, we get an n'-dimensional incidence structure $\Sigma_{n'}$. The new structure $\Sigma_{n'}$ is called a* skeleton *of Σ_n.*

There is another possibility of deriving a new incidence structure from the given one. If we choose a single n'-dimensional element of an n-dimensional structure, than once more a structure arises with $a'_l = b_{n'l}$ l-dimensional elements.

The new structure constants are the same as the old because a k-dimensional element is attached to b_{kl} l-dimensional elements in both structures for $l < k$.

Theorem 5. *Consider an n-dimensional incidence structure Σ_n. If we take a single n'-dimensional element with all of its attached k-dimensional elements for $k < n'$, then we obtain a new n'-dimensional incidence structure $\Sigma_{n'}$ with the same structure constants as in Σ_n. This incidence structure $\Sigma_{n'}$ contains b_{nl} l-dimensional elements with $l < n'$.*

A third method of generating new incidence structures is given by a *duality principle:*

Theorem 6. *Given an n-dimensional incidence structure Σ_n, a new incidence structure Σ'_n is obtained by replacing the numbers of elements and the structure constants by*

$$a'_k = a_{n-k} \quad \text{and} \quad b'_{kl} = b_{n-k,n-l} \ .$$

In all three cases described by Theorems 4, 5, and 6, we can prove that the matching formula and the relations between structure constants are fulfilled. In fact, therefore, new incidence structures are generated.

There is a simple and very important case where n-dimensional finite and infinite incidence structures exist. The n-dimensional number lattice is such a structure. Assuming that $n = 1$, we can close the ring-like structure so that a finite homogeneous structure of type $\mathbb{Z}^1$ arises (see the second row in Fig. 2). The three-dimensional number lattice $\mathbb{Z}^3$ with integer-valued point coordinates is an infinite incidence structure as given below:

b_{kl}	$l=0$	1	2	3
$k=0$	1	6	12	8
1	2	1	4	4
2	4	4	1	2
3	8	12	6	1

With respect to the foregoing theorems, we can restrict ourselves to the determination of the structure constants b_{0k} for $1 \leq k \leq n$. Any k-dimensional element attached to the origin as zero-dimensional element can be represented as an n-tuple of 1s and 0s in which the 1 occurs k-times. There are $n!/(k!(n-k)!)$ different tuples of this kind when we only take positive coordinate values into account. If we take into account that the 1s can have positive or negative signs, we obtain altogether

$$b_{0k}^{(n)} = 2^k \binom{n}{k} \ .$$

The number b_{kl} is given in $\mathbb{Z}^n$ for $k < l$ by all possibilities of $(n-k)$-tuples with $l-k$ signed 1s and $n-l$ 0s so that

$$b_{kl}^{(n)} = 2^{l-k} \binom{n-k}{l-k} = 2^{l-k} \binom{n-k}{n-l} \ .$$

The number b_{kl} with $k > l$ is the number of all l-dimensional elements attached to k-dimensional elements in $\mathbb{Z}^n$. This number is independent of dimension n. Because of Theorems 5 and 6, we obtain

$$b_{kl}^{(n)} = b_{n-k,n-l}^{(n)} = 2^{k-l}\binom{k}{k-l} = 2^{k-l}\binom{k}{l} .$$

We condense these results to the formula of structure constants for $\mathbb{Z}^n$:

$$b_{kl}(n) = \begin{cases} 2^{l-k}\binom{n-k}{n-l} & \text{for } k < l \\ 1 & \text{for } k = l \\ 2^{k-l}\binom{k}{l} & \text{for } k > l \end{cases} . \tag{4}$$

These expressions for the structure constants b_{kl} of the n-dimensional number lattice considered as incidence structure are fundamental both for finite toroidally closed structures and for infinite grid point spaces $\mathbb{Z}^n$ (see also [9, 5]). Using very simple proofs it can be shown that all matching formulae can be fulfilled using the structure constants of $\mathbb{Z}^n$. Therefore all axioms, laws, theorems, and conclusions of homogeneous incidence structures are valid also for $\mathbb{Z}^n$. It should be mentioned that with

$$\frac{b_{lk}}{b_{kl}} = \frac{2^{l-k}\binom{l}{k}}{2^{l-k}\binom{n-k}{n-l}} = \frac{\binom{n}{k}}{\binom{n}{l}} \quad \text{for } k \le l ,$$

$$\frac{b_{lk}}{b_{kl}} = \frac{2^{k-l}\binom{n-l}{n-k}}{2^{l-k}\binom{k}{l}} = \frac{\binom{n}{k}}{\binom{n}{l}} \quad \text{for } k \ge l ,$$

the important identity

$$\sum_{k=0}^{n}(-1)^k \frac{b_{lk}}{b_{kl}} = \frac{1}{\binom{n}{l}} \sum_{k=0}^{n}(-1)^k \binom{n}{k} = 0 \tag{5}$$

is valid for the structure constants b_{kl} of $\mathbb{Z}^n$ for $n > 0$.

Theoretical approaches to three-dimensional images have been investigated previously [2, 4, 7, 8]. But up to now there is no generally applicable theory for n-dimensional objects in grid point spaces $\mathbb{Z}^n$. However the basic notion of homogeneous incidence structures allows a fruitful generalization of the theory of two-dimensional neighbourhood structures [12] to more than two dimensions. In $\mathbb{Z}^2$, we know the relation $1/\nu + 1/\lambda = 1/2$ for the structure constants ν and λ (number of neighbours and length of meshes). In the general case of $\mathbb{Z}^n$ there are $n(n+1)$ structure constants which are, however, not independent of one another. The special case of the "orthogonal" grid space $\mathbb{Z}^n$ is the most

important model for n-dimensional homogeneous incidence structures. For $n=2$ there are three different structure models, and for $n=3$ eleven different structure models [13, 14].

The *Euler characteristic* of a finite n-dimensional incidence structure is defined by

$$\chi^{(n)} = \sum_{k=0}^{n} (-1)^k a_k \ . \tag{6}$$

For homogeneous structures and $\chi^{(n)} = 0$, we obtain with $a_k = a_l b_{lk}/b_{kl}$ the relationships

$$\frac{\chi^{(n)}}{a_l} = \sum_{k=0}^{n} (-1)^k \frac{b_{lk}}{b_{kl}} = 0 \quad \text{for} \quad 0 \le l \le n \ . \tag{7}$$

Using the relationships (3) between the structure constants, we can show that all of these $n+1$ equations are equivalent to one another.

4 Objects in $\mathbb{Z}^n$

A subset $B \subset E_0$ of grid points of $\mathbb{Z}^n$ is defined as an object. Connectivity of an object is determined by a neighbourhood relation: two points $p, q \in E_0$ are neighbours if they are attached to the same edge. Each point $p \in B$ is an *object point*. A k-dimensional element is called an *object element* if all b_{kl} attached l-dimensional elements with $l < k$ are also object elements.

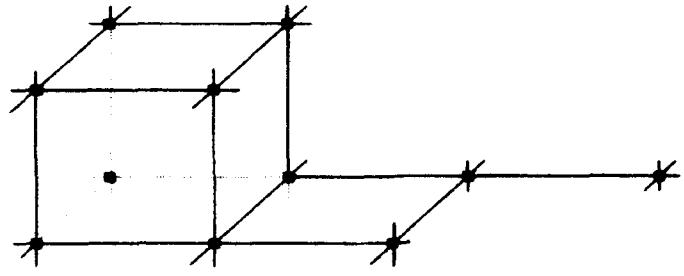


Fig. 5. This object ($\varepsilon = 11, \kappa = 16, \mu = 7, \zeta = 1$) is characterized by $c_{01} = 34, c_{12} = 26$, and $c_{23} = 8$

Definition 7. A k-dimensional element $e \in E_k$ is called an *object element* of B if all $b_{k0} = 2^k$ to e attached points $p \in E_0$ are object points, that is, belong to B. *Marginal elements* of an object B are elements of $\mathbb{Z}^n$ which are not object elements of B but are attached at least to one point of B.

The number $c_{lk}(e)$ with $l < k$ is the number of k-dimensional marginal elements which are attached to the l-dimensional object element e. The sums

$$c_{lk} = \sum_{e \in E_l} c_{lk}(e)$$

taken for all l-dimensional object elements of B are called marginal numbers (see Fig. 5). For an object B, the general matching equations have to be replaced by

$$\begin{aligned} a_l b_{lk} - c_{lk} &= a_k b_{kl} \quad \text{for} \quad k > l \ , \\ a_l b_{lk} \qquad &= a_k b_{kl} \quad \text{for} \quad k = l \ , \\ a_l b_{lk} + c_{lk} &= a_k b_{kl} \quad \text{for} \quad k < l \ . \end{aligned} \tag{8}$$

Here a_k means as the usual the number of k-dimensional object elements. The first group of equations follows because all l-dimensional elements attached to a k-dimensional element belong likewise to the object B. But there are c_{lk} k-dimensional elements among all k-dimensional elements attached to l-dimensional object elements which do not belong to the object B. Finally, the last group of equations follows by reversing k and l.

These equations—called *object matching formulae*—express the inhomogeneity of objects in opposition to the homogeneity of the underlying incidence structure. For any object B of $\mathbb{Z}^n$, we define

$$\Psi^{(n)}(B) = \sum_{k=0}^{n} (-1)^k a_k(B)$$

as the *Euler number* of the object B. Contrary to the expression (6) for an incidence structure, the Euler number of an object can be different from zero as simple examples show (in Fig. 5, $\Psi^{(3)} = \varepsilon - \kappa + \mu - \zeta = 1$, where μ is the number of "meshes" or faces, and ζ is the number of cells of the object). A motivation for this alternating sum is given in Sect. 7 where $\Psi = q_0$ is shown to be invariant with respect to object magnification. It is

$$\frac{\binom{l}{k}}{\binom{n-k}{n-l}} = \frac{\binom{n}{k}}{\binom{n}{l}} \quad \text{for} \quad k < l \ ,$$

$$\frac{\binom{n-l}{n-k}}{\binom{k}{l}} = \frac{\binom{n}{k}}{\binom{n}{l}} \quad \text{for} \quad k > l \ .$$

Therefore using the structure constants of $\mathbb{Z}^n$, we obtain

$$\begin{aligned} \sum_{k=0}^{n} (-1)^k \frac{b_{lk}}{b_{kl}} &= \sum_{k=0}^{l-1} (-1)^k \frac{b_{lk}}{b_{kl}} + (-1)^l + \sum_{k=l+1}^{n} (-1)^k \frac{b_{lk}}{b_{kl}} \\ &= \frac{1}{\binom{n}{l}} \sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 \end{aligned}$$

for any l with $0 \le l \le n$ corresponding to $\chi^{(n)} = 0$. With respect to this, the Euler number of an object B can be expressed by the object matching formulae so that

$$\Psi^{(n)}(B) = \sum_{k=0}^{l-1}(-1)^k \frac{c_{kl}(B)}{b_{kl}} - \sum_{k=l+1}^{n}(-1)^k \frac{c_{lk}(B)}{b_{kl}}$$

follows for any l with $0 \le l \le n$. This formula means that the Euler number of any object in $\mathbb{Z}^n$ can be determined by counting only the marginal elements (for $n = 3$, Lee and Rosenfeld have shown that $\Psi^{(3)}$ can be determined by using only the Gaussian curvature of the object surface [7]. Using the object matching formula, it is possible to express successively all numbers a_k by only the point number a_0 and the numbers $c_{l-1,l}$:

$$a_k = a_0 \frac{b_{0k}}{b_{k0}} - \sum_{l=1}^{k} \frac{c_{l-1,l}}{b_{l-1,l}} \cdot \frac{b_{l-1,k}}{b_{k,l-1}} .$$

This equation is correct for $k = 0$. Assuming the correctness for a given k, we have for fixed l

$$\begin{aligned} a_{k+1} &= a_k \frac{b_{k,k+1}}{b_{k+1,k}} - \frac{c_{k,k+1}}{b_{k+1,k}} \\ &= a_0 \frac{b_{0k}b_{k,k+1}}{b_{k+1,k}b_{k0}} - \sum_{l=1}^{k} \frac{c_{l-1,l}b_{l-1,k}b_{k,k+1}}{b_{l-1,l}b_{k,l-1}b_{k+1,k}} - \frac{c_{k,k+1}}{b_{k+1,k}} \qquad (9) \\ &= a_0 \frac{b_{0,k+1}}{b_{k+1,0}} - \sum_{l=1}^{k+1} \frac{c_{l-1,l}b_{l-1,k+1}}{b_{l-1,l}b_{k+1,l-1}} , \end{aligned}$$

so that the equation is also fulfilled for $k + 1$. Taking into account this general expression for a_k and the formula $\sum(-1)^k b_{0k}/b_{k0} = 0$, we obtain

$$\Psi^{(n)} = \sum_{k=0}^{n}(-1)^k a_k = \sum_{k=0}^{n}(-1)^k \left(- \sum_{l=1}^{k} \frac{c_{l-1,l}b_{l-1,k}}{b_{l-1,l}b_{k,l-1}} \right) .$$

The double sum can be rearranged:

$$\Psi^{(n)} = \sum_{l=1}^{n} \frac{c_{l-1,l}}{b_{l-1,l}} \sum_{k=l}^{n}(-1)^{k+1} \frac{b_{l-1,k}}{b_{k,l-1}} .$$

If we consider the structure constants b_{kl} for $\mathbb{Z}^n$, the formulae (9) for the a_ks, and the identity

$$- \sum_{k=m+1}^{n}(-1)^k \binom{n}{k} = \sum_{k=0}^{m}(-1)^k \binom{n}{k} = (-1)^m \binom{n-1}{m} , \qquad (10)$$

which is provable by induction from m to $m + 1$, then it follows

$$\Psi^{(n)} = \frac{1}{2n} \sum_{l=}^{n} \qquad {}^{-1} c_{l-1,l} . \qquad (11)$$

This is a very simple formula for the Euler number $\Psi(n)$ of an object in $\mathbb{Z}^n$. Independent of dimension n, the $c_{l-1,l}$ are

- the number c_{01} of all marginal edges counted for all object points,
- the number c_{12} of all marginal meshes counted for all object edges,
- the number c_{23} of all marginal cells counted for all object meshes.

All $c_{l-1,l}$ appearing here and in many further formulae can be expressed by the a_ks as the object matching formula shows:

$$c_{l-1,l} = a_{l-1}b_{l-1,l} - a_l b_{l,l-1} \ .$$

5 Similarity of Objects

A fundamental notion in geometry is that of *similarity*. Similar objects can be obtained in discrete geometry by magnification of objects or by refinement of the $\mathbb{Z}^n$-lattice. At all coordinate axis between x_l and $x_l + 1$, we introduce new values $x'_l = r \cdot x_l, r \cdot x_l + 1, r \cdot x_l + 2, \ldots, r \cdot x_l + r - 1$. Both values $r \cdot x_l$ and $r \cdot (x_l + 1)$ form the support of a "super lattice" in the new lattice (see Fig. 6).

This definition of similarity has the advantage that a point gets a point, a line gets a line, etc. The first to use this notion of magnification of a curve-like object was Freeman [3]:

> The process of expansion is performed as illustrated by the following example. Given a curve represented by (the code sequence) 012075, a curve exactly twice this size, but otherwise indistinguishable, is given by 001122007755. To expand a curve by a ratio n, each of the digits of the curve must be replaced by a set of n digits. One notes that n must be an integer. *Freeman*

By means of r-fold magnification, each old k-dimensional object element yields a k-dimensional r-cube, that is, a cube with $(r+1)^k$ object points. The l-dimensional object elements of the magnified object can be parts of different k-cubes for $k \geq l$. For instance, the points of the magnified object are caused by points, edges, meshes, ... of the old object (see Fig. 6).

We will determine the number of l-dimensional elements which are caused only by k-cubes but not by m-cubes with $l \leq m < k$. Each l-dimensional element is characterized by an origin $(x_1, \ldots, x_i, \ldots, x_n)$ and by l pairs $\{x_i, x_{i+1}\}$ which determine the 2^l points of the l-dimensional element. Within the k-cube, there are r possibilities for the values of each of the l pairs so that both points x_i and x_{i+1} lie inside the k-cube. Therefore, we obtain r^l possibilities for l-dimensional elements using l varying coordinates of the element origin and $k - l$ fixed coordinates.

The $k - l$ fixed coordinates can have only $r - 1$ different values because the "extremal values" would give l-dimensional elements at the surface of the k-cube. In total, we obtain $r^l(r-1)^{k-l}$ possibilities that a given l-dimensional element

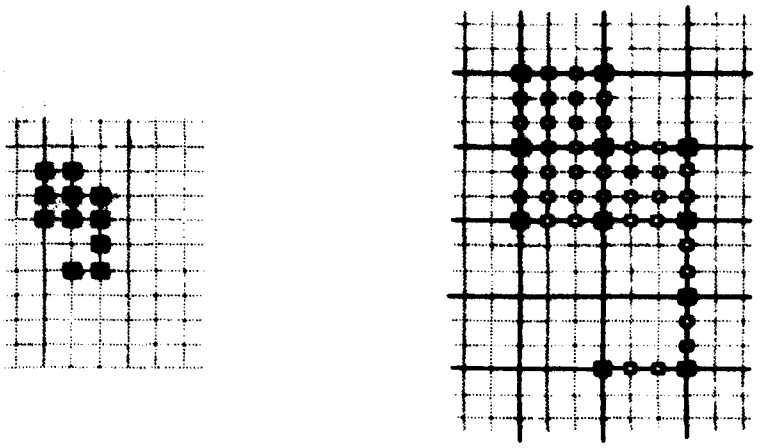


Fig. 6. Threefold refinement of a two-dimensional lattice and magnification of an object

lies inside a k-cube. Because we can select any l-tuple of k coordinates, each k-dimensional element e of the old lattice $\mathbb{Z}^n$ leads to

$$a_l'(e) = \binom{k}{l} r^l (r-1)^{k-l}$$

l-dimensional elements of the refined lattice for $l \leq k$. Therefore the total number of l-dimensional elements of the magnified object is

$$a_l' = \sum_{k=l}^{n} \sum_{e \in E_k} a_l'(e) = \sum_{k=l}^{n} \binom{k}{l} r^l (r-1)^{k-l} a_k \ .$$

Especially, we obtain for n=2:

$$\begin{aligned} \varepsilon' &= \varepsilon + (r-1)\kappa + (r-1)^2\mu \ , \\ \kappa' &= r\kappa + 2r(r-1)\mu \ , \\ \mu' &= r^2\mu \ , \end{aligned}$$

and for $n = 3$:

$$\begin{aligned} \varepsilon' &= \varepsilon + (r-1)\kappa + (r-1)^2\mu + (r-1)^3\zeta \ , \\ \kappa' &= r\kappa + 2r(r-1)\mu + 3r(r-1)^2\zeta \ , \\ \mu' &= r^2\mu + 3r^2(r-1)\zeta \ , \\ \zeta' &= r^3\zeta \ . \end{aligned}$$

Using these formulae, one can derive for $n = 2$ the relationships

$$\begin{aligned} \varepsilon' - \kappa' + \mu' &= (\varepsilon - \kappa + \mu)r^0 \ , \\ \kappa' - 2\mu' &= (\kappa - 2\mu)r^1 \ , \\ \mu' &= \mu r^2 \ . \end{aligned}$$

For n=3, the relationships

$$\begin{aligned} \varepsilon' - \kappa' + \mu' - \zeta' &= (\varepsilon - \kappa + \mu - \zeta)r^0 \ , \\ \kappa' - 2\mu' + 3\zeta' &= (\kappa - 2\mu + 3\zeta)r^1 \ , \\ \mu' - 3\zeta' &= (\mu - 3\zeta)r^2 \ , \\ \zeta' &= \zeta r^3 \ . \end{aligned}$$

follow. It can be seen that there are "dimensionless" characteristics, the Euler numbers $\Psi^{(2)} = \varepsilon - \kappa + \mu$ and $\Psi^{(3)} = \varepsilon - \kappa + \mu - \zeta$, and other characteristics which can be transformed by powers of the magnification factor r. Therefore, a general investigation should be successful.

6 General Surface Formulas

The terms in the transformation formulae for a'_l can be rearranged with respect to powers of magnification factor r:

$$a'_l = \sum_{m=l}^{n} \binom{m}{l} r^m q_m \ . \tag{12}$$

Here q_m means

$$q_m = \sum_{j=0}^{n-m} (-1)^j \binom{j+m}{j} a_{j+m} \ . \tag{13}$$

We obtain by lattice refinement or object magnification

$$q'_m = \sum_{j=0}^{n-m} (-1)^j \binom{j+m}{j} a'_{j+m} \ .$$

By consideration of the new a'_ks given in formula (12), it follows that

$$\begin{aligned} q'_m &= \sum_{j=0}^{n-m} (-1)^j \binom{j+m}{j} \sum_{k=j+m}^{n} \binom{k}{j+m} r^k q_k \\ &= \sum_{k=m}^{n} \sum_{j=0}^{k-m} (-1)^j \binom{j+m}{j} \binom{k}{j+m} r^k q_k \\ &= \sum_{k=m}^{n} \binom{k}{m} r^k q_k \sum_{j=0}^{k-m} (-1)^j \binom{k-m}{j} \ . \end{aligned}$$

The second sum has the value 1 for $k = m$ and otherwise the value 0, and we obtain the important relationship

$$q'_m = r^m q_m \ . \tag{14}$$

Because of this equation, the q_m defined above are object characteristics of *similarity degree* m. They can be determined by numbers a_k of k-dimensional elements of an object:

$$\begin{aligned}
q_0 &= a_0 - a_1 + a_2 - a_3 + a_4 - a_5 \quad \pm\cdots+(-1)^n a_n \ , \\
q_1 &= \quad a_1 - 2a_2 + 3a_3 - 4a_4 + 5a_5 \quad \mp\cdots+n(-1)^{n-1}a_n \ , \\
q_2 &= \quad\quad a_2 - 3a_3 + 6a_4 - 10a_5 \quad \pm\cdots+\binom{n}{2}(-1)^n a_n \ , \\
q_3 &= \quad\quad\quad a_3 - 4a_4 + 10a_5 \quad \mp\cdots+\binom{n}{3}(-1)^{n-1}a_n \ , \\
q_4 &= \quad\quad\quad\quad a_4 - 5a_5 \quad \pm\cdots+\binom{n}{4}(-1)^n a_n \ .
\end{aligned}$$

Now, we can take into account the relationships (9) where the a_ks are expressed by the numbers $c_{l-1,l}$ of marginal elements. Then, it follows from the a_k's as functions of c's and b's

$$\begin{aligned}
q_m = &\sum_{j=0}^{n-m}(-1)^j\binom{j+m}{j}a_0\frac{b_{0,j+m}}{b_{j+m,0}} \\
&-\sum_{j=0}^{n-m}(-1)^j\binom{j+m}{j}\sum_{l=1}^{j+m}\frac{c_{l-1,l}}{b_{l-1,l}}\cdot\frac{b_{l-1,j+m}}{b_{j+m,l-1}} \ .
\end{aligned}$$

Using the structure constants of $\mathbb{Z}^n$ (see formula (4)), we obtain

$$\begin{aligned}
q_m = &\, a_0\binom{n}{m}\sum_{j=0}^{n-m}(-1)^j\binom{n-m}{j} \\
&-\sum_{j=0}^{n-m}(-1)^j\binom{j+m}{j}\sum_{l=1}^{j+m}\frac{c_{l-1,l}}{2}\cdot\frac{\binom{n-(l-1)}{n-(j+m)}}{\binom{n-(l-1)}{n-(j+m)}\binom{j+m}{l-1}} \ .
\end{aligned}$$

The first term vanishes for $m < n$. The second term can be transformed to

$$\begin{aligned}
q_m &= \sum_{j=0}^{n-m}(-1)^j\sum_{l=1}^{j+m}\frac{c_{l-1,l}}{2}\cdot\frac{(n-l)!(l-1)!}{(n-m)!m!}\binom{n-m}{j} \\
&= -\binom{n}{m}\sum_{j=0}^{n-m}(-1)^j\binom{n-m}{j}\sum_{l=1}^{j+m}\frac{c_{l-1,l}}{2l}\frac{1}{\binom{n}{l}} \ ,
\end{aligned}$$

or by rearrangement

$$q_m = -\binom{n}{m}\sum_{l=1}^{n}\frac{c_{l-1,l}}{2l}\frac{1}{\binom{n}{l}}\sum_{j=0}^{n-m}(-1)^j\binom{n-m}{j}$$

$$+\binom{n}{m}\sum_{l=m+1}^{n}\frac{c_{l-1,l}}{2l}\frac{1}{\binom{n}{l}}\sum_{j=0}^{l-m-1}(-1)^j\binom{n-m}{j}\ .$$

Once more, the first term vanishes for $m < n$. The second term can be transformed, using the identity formula (10), to

$$q_m=\binom{n}{m}\sum_{l=m+1}^{n}\frac{c_{l-1,l}}{2l}\frac{1}{\binom{n}{l}}(-1)^{l-m-1}\binom{n-m-1}{l-m-1}\ .$$

Finally, we obtain for $0 \le m < n$ the simple but very important *geometric main formula*

$$q_m^{(n)}=\frac{1}{2(n-m)}\sum_{l=m+1}^{n}(-1)^{l-m+1}\binom{l-1}{m}c_{l-1,l}\ .$$

Especially, it follows for the quantities used in image processing:

$$q_0^{(0)}=a_0=\varepsilon$$

$$\begin{aligned} q_0^{(1)}&=\tfrac{c_{01}}{2}=\varepsilon-\kappa\ ,\\ q_1^{(1)}&=a_1=\kappa\ , \end{aligned}$$

$$\begin{aligned} q_0^{(2)}&=\tfrac{c_{01}-c_{12}}{4}=\varepsilon-\kappa+\mu\ ,\\ q_1^{(2)}&=\tfrac{c_{12}}{2}=\kappa-2\mu\ ,\\ q_2^{(2)}&=a_2=\mu\ , \end{aligned}$$

$$\begin{aligned} q_0^{(3)}&=\tfrac{c_{01}-c_{12}+c_{23}}{6}=\varepsilon-\kappa+\mu-\zeta\ ,\\ q_1^{(3)}&=\tfrac{c_{12}-2c_{23}}{4}=\kappa-2\mu+3\zeta\ ,\\ q_2^{(3)}&=\tfrac{c_{23}}{2}=\mu-3\zeta\ ,\\ q_3^{(3)}&=a_3=\zeta\ . \end{aligned}$$

7 Interpretation of Object Characteristics

We have already shown in Sect. 4 that the Euler number $\Psi^{(n)}=q_0$ of an object can be determined only by numbers $c_{l-1,l}$ of marginal elements. The marginal elements are attached uniquely to the surface of the object, and therefore $c_{l-1,l}^{(i)}$ means the number $c_{l-1,l}$ for the ith surface. With

$$\Psi^{(n,i)}=\frac{1}{2n}\sum_{l=1}^{n}(-1)^{l+1}c_{l-1,l}^{(i)}\ ,$$

it follows that

$$\Psi^{(n)}=\sum_{\text{surfaces } i}\Psi^{(n,i)}\ .$$

Similar formulae hold also for the object geometric characteristics q_m, because the geometrical main formula shows that we can define $q_0, q_1, \ldots q_{n-1}$ for single surfaces. For $n = 3$, we can determine the Euler number $\Psi^{(3,i)} = 1 - g^{(i)} = 1 - T^{(i)} \in \{1, 0, -1, \ldots\}$ depending on the genus $g^{(i)}$ or on the number $T(i)$ of "tunnels" through the i-th surface. Therefore the "phenomenological" formula for the Euler number of a three-dimensional object is

$$\Psi^{(3)} = \sum_{\text{surfaces } i} (1 - g^{(i)}) = S - T$$

where S is the number of surfaces and T is the number of tunnels through all these surfaces.

It can be seen from the formulae of Sect. 5 and 6 that there are "dimensionless" characteristics, the Euler numbers $\Psi^{(2)} = \varepsilon - \kappa + \mu$ or $\Psi^{(3)} = \varepsilon - \kappa + \mu - \zeta$. But there exist also "linear" characteristics: $\kappa - 2\mu = l/2$ for $n = 2$ with $l = c_{12}$ as contour length, $\kappa - 2\mu + 3\zeta = (l - 2m)/4$ for $n = 3$ with $l = c_{12}$ as the number of marginal meshes, and $m = c_{23}$ as the number of marginal cells. The expression $\kappa - 2\mu + 3\zeta$ for three-dimensional objects in $\mathbb{Z}^3$ corresponds to the mean curvature integral [11] for objects in E^3, and $\mu - 3\zeta = m/2$ is a "quadratic" measure for the surface content of an object in $\mathbb{Z}^3$ (Rosenfeld offers the number c_{01} of nodes of the surface graph as a measure for "surface area" [10]).

Because q_m has to be transformed by the factor r^m, we can derive many "nonlinear" dimensionless numbers. For two-dimensional objects in $\mathbb{Z}^2$, the *shape factor* or *form factor*

$$f = \frac{q_1^2}{4q_2} = \frac{(\kappa - 2\mu)^2}{4\mu} = \frac{l^2}{16\mu} \geq 1$$

exists for $\mu > 0$, which has the same value for all similar objects, and reaches its minimal value for square-shaped objects. The proof for this statement can be given in the following way:

1. Let X be any connected object with μ meshes. Let $F(X)$ be the corresponding "filled" object where the included holes are filled with points, edges, and meshes:

$$\frac{c_{12}^{(a)}}{2} = q_1(F) \leq q_1(X) = \frac{c_{12}(X)}{2} = \frac{c_{12}^{(a)} + \sum c_{12}^{(i)}}{2} ,$$
$$\mu^{(a)} = q_2(F) \geq q_2(X) = \mu^{(a)} - \sum \mu^{(i)} ,$$

   where the superscripts (a) and (i) denote values for the outer and inner boundaries, respectively. Therefore we obtain
   $f(F) = q_1^2(F)/q_2(F) \leq q_1^2(X)/q_2(X) = f(X)$.
2. Let F be any connected object without holes with μ meshes. Let $R(F)$ be the circumscribing rectangle, that is, the maximal object with the same extremal coordinate values as F:

$$c_{12}(R) = l(R) = 2q_1(R) \leq 2q_1(F) = c_{12}(F) = l(F),$$
$$\mu(R) = q_2(R) \geq q_2(F) = \mu(F) .$$

It follows that $f(R) \leq f(F)$ and therefore $f(R) \leq f(X)$.

3. Let R be any rectangular shaped object with $M \cdot N$ points. We obtain

$$\kappa = N(M-1) + M(N-1) \ ,$$
$$\mu = q_2 = (N-1)(M-1) \ ,$$
$$q_1 = \kappa - 2\mu = N + M - 2 \ .$$

Without restriction, we can assume that $M = N + k \geq N$ with $k \geq 0$. Then the inequality $(2N + k - 2)^2 \geq 4(N-1)(N+k-1)$ is fulfilled and finally

$$f(R) = \frac{(N+M-2)^2}{4(N-1)(M-1)} \geq \frac{(N+N-2)^2}{4(N-1)^2} = f(S) = 1$$

follows with $f(S)$ as the form factor of a square S. Because of that, the inequality $f(X) \geq f(S) = 1$ is fulfilled for any connected object X.

In $\mathbb{Z}^3$, there exist two independent shape factors for objects:

$$g = \frac{q_1^2}{3q_2} = \frac{(\kappa - 2\mu + 3\zeta)^2}{3(\mu - 3\zeta)} = \frac{(c_{12} - 2c_{23})^2}{24c_{23}} \geq 1 \ ,$$
$$h = \frac{q_1 q_2}{9q_3} = \frac{(\kappa - 2\mu + 3\zeta)(\mu - 3\zeta)}{9\zeta} = \frac{(c_{12} - 2c_{23})c_{23}}{72\zeta} \geq 1 \ .$$

They have the same numerical values for all similar objects assuming that $\mu > 3\zeta$ and $\zeta > 0$. The minimal values will be reached for cube-like objects as can be proved with the same proving method as for shape factor f in $\mathbb{Z}^2$.

References

1. Beutelspacher, A. (1982). Einführung in die endliche Geometrie, Band I, Wissenschaftsverlag, Mannheim.
2. Bieri, H., Nef, W. (1984). Algorithms for the Euler characteristics and related additive functionals of digital objects, Computer Vision, Graphics, and Image Processing 28, pp. 166–175.
3. Freeman, H. (1961). On the encoding of arbitrary geometry configurations, IRE Trans. EC-10 pp. 260–268.
4. Gray, S.B. (1971). Local properties of binary images in two and three dimensions, IEEE Trans. C-20, pp. 551–561.
5. Klette, R. (1983). M-dimensional Cellular Spaces, Techn. Rep. CAR-TR-6, Univ. of Maryland, Center for Automation Research, College Park MD.
6. Klette, R. (1983). The m-dimensional Grid Point Space, Techn. Rep. TR-1256, Univ. of Maryland, Comp.Sc.Center, College Park MD.
7. Lee, C.N., Rosenfeld, A. (1986). Computing the Euler number of a 3D Image, Techn. Rep. CAR-TR-205, Center for Automation Research, University of Maryland, College Park MD.
8. Park, C.M., Rosenfeld, A. (1971). Connectivity and genus in three dimensions, Techn.Rep. TR-153, Computer Science Center, University of Maryland, College Park MD.

9. Rosenfeld, R.A., Jaglom, I.M. (1971). Nichteuklidische Geometrie. In: Alexandroff, P.S., Markuschewitsch, A.I., Chintschin, A.J. (eds.): Enzyklopödie der Elementarmathematik, Bd.5, Deutscher Verlag der Wissenschaften, Berlin, pp. 458–464.
10. Rosenfeld, A., Kong, T.Y. (1989). Digital surfaces, Techn. Rep. CAR-TR-467, University of Maryland, Center for Automation Research, College Park MD.
11. Santolo, L.A. (1976). Integral Geometry and Geometrical Probability, Addison-Wesley, London.
12. Voss, K. (1988). Theoretische Grundlagen der digitalen Bildverarbeitung, Akademie-Verlag, Berlin.
13. Voss, K. (1991). Images, objects, and surfaces in $\mathbb{Z}^n$. Int. J. Pattern Recognition and Artif. Intell. 5, pp. 797–808.
14. Voss, K. (1993). Discrete Images, Objects, and Functions in $\mathbb{Z}^n$, Springer-Verlag, Berlin.

On Boundaries and Boundary Crack-Codes of Multidimensional Digital Images *

T. Yung Kong

Department of Computer Science, Queens College, Flushing, NY 11367, USA

Abstract. An asymmetric, anisotropic relation on the boundary elements of a binary image, due to Gordon and Udupa (Gordon, D., Udupa, J.K. (1989). Fast surface tracking in three-dimensional binary images, Computer Vision, Graphics and Image Processing 45, pp. 196–214.), is used to generalize the 2-D concept of a difference crack-code to crack-codes that represent boundaries in higher-dimensional binary images. For an n-dimensional binary image where $n \geq 3$, it is shown how each connected component of the boundary (with respect to Gordon and Udupa's relation) can be represented by a "crack code" consisting of a single pair of sequences. It is also shown that the amount of memory required to store such a crack code for each component of the boundary does not exceed $(4+\lceil\log(n-1)\rceil)(1-1/n)$ bits per boundary element. In particular, the memory requirement is no more than $3\frac{1}{3}$ bits per boundary element for 3-D images, and no more than $4\frac{1}{2}$ bits per boundary element for 4-D images.

Keywords: digital topology, digital geometry, binary images, 3-dimensional images, multidimensional images, adjacency, crack-code.

1 Introduction

The difference crack-code representation of boundaries in 2-D binary images ([8, p. 199]) is well known. This paper will explain how an asymmetric, anisotropic adjacency relation on boundary elements introduced by Gordon and Udupa [2] (and further studied by Kong and Udupa [5]) can be used to give a space-efficient generalization of difference crack-codes to higher-dimensional binary images.

* A part of the work reported in this paper was done while the author held a visiting appointment at the Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA. The author has enjoyed many useful and stimulating discussions with Dr G. T. Herman and Dr J. K. Udupa on the subject of boundaries in multidimensional digital images.

2 The 2-Dimensional Case

2.1 The Simple 8-Boundaries of a Set of Pixels

In this paper the term *pixel* means a unit square in the Euclidean plane whose corners have all integer coordinates. If C is any set of pixels then $\overline{C}$ denotes the set of all pixels that are not in C. The reader is assumed to be familiar with the standard concepts of κ-adjacency, κ-connectedness and κ-components for $\kappa = 4$ or 8. (For definitions of these concepts, see [4].)

A *boundary element* of a set of pixels C is a pixel edge $p \cap q$ where $p \in C, q \in \overline{C}$ and p is 4-adjacent to q. The set of all boundary elements of C is called the *boundary* of C, and written ∂C. Note that $\partial C = \partial \overline{C}$.

Let X be a 4-connected finite set of pixels. A finite 8-component of $\overline{X}$ is called an *8-hole* of X. If X has just m 8-holes (where $m \geq 0$) then the boundary of X can be decomposed into $m+1$ parts, where one part separates X from the outside (more precisely, from the unbounded 8-component of $\overline{X}$) and each of the other parts separates X from one of the 8-holes. Each of these parts of ∂X will be called a *simple 8-boundary* of X. More precisely, a simple 8-boundary of X is a set ∂Y in which Y is an 8-component of $\overline{X}$. Thus ∂X can be partitioned into $m+1$ simple 8-boundaries, where m is the number of 8-holes of X.

2.2 Strong Neighbours and Strong Successors

Let C be any set of pixels. If $p \cap u$ and $q \cap v$ are boundary elements of C, where $p, q \in C$ and $u, v \in \overline{C}$, then we will say $p \cap u$ is a *strong neighbour of* $q \cap v$ *relative to* C if one of the following holds:

1. $p = q$, and u is 8-adjacent to v.
2. p is 4-adjacent to q, and u is 4-adjacent to v.
3. $u = v$, p is 8-adjacent to q, and the other common 4-neighbour of p and q is in C.

The different cases of this definition are illustrated in Fig. 1, where the spotted pixels are in C and the white pixels are in $\overline{C}$.

Let $e = p \cap q$ be a boundary element of C, where $p \in C$ and $q \in \overline{C}$. Then e^*_C will denote the directed edge obtained by orienting e in such a way that p lies on the right and q on the left.

Every element e of ∂C has exactly two strong neighbours relative to C: the initial point of e^*_C is an endpoint of one strong neighbour, and the final point of e^*_C is an endpoint of the other strong neighbour. The former strong neighbour will be called the *strong predecessor of* e *relative to* C; the latter strong neighbour will be called the *strong successor of* e *relative to* C. These definitions are also illustrated in Fig. 1.

For $e \in \partial C$, let $\sigma_C(e)$ denote the strong successor of e relative to C. Then σ_C is a bijection of ∂C onto ∂C and, for every $e \in \partial C$, $\sigma_C^{-1}(e)$ is the strong predecessor of e relative to C.

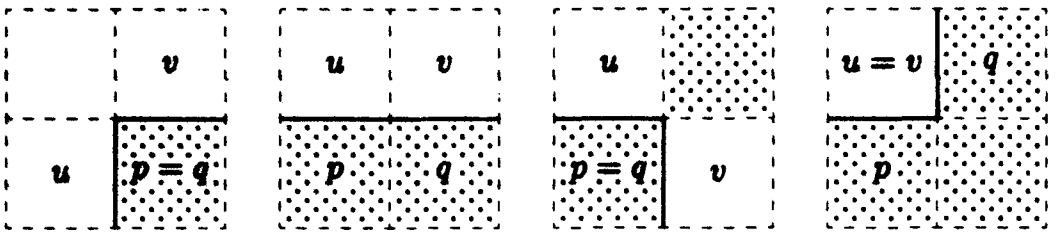


Fig. 1. If the spotted pixels are in C and the white pixels are in $\overline{C}$, then in each case: (i) the boundary elements $p \cap u$ and $q \cap v$ are strong neighbours of each other relative to C; (ii) $q \cap v$ is the strong successor of $p \cap u$ relative to C, and $p \cap u$ is the strong predecessor of $q \cap v$ relative to C.

Now let X be an arbitrary finite 4-connected set of pixels. Note that a strong neighbour of $b \in \partial X$ relative to X must lie on the same simple 8-boundary of X as b itself. If the simple 8-boundary of X that contains b has just λ boundary elements, then $\sigma_X^\lambda(b) = b$; moreover, the sequence $(b, \sigma_X(b), \sigma_X^2(b), \ldots, \sigma_X^{\lambda-1}(b))$ contains all elements of that simple 8-boundary.

2.3 A Difference Crack-Code for Simple 8-Boundaries

Again, let X be an arbitrary finite 4-connected set of pixels, let b be any boundary element of X and let λ be the number of boundary elements in the simple 8-boundary of X that contains b. For every non-negative integer j let $c_j(b) \in \{-1, 0, 1\}$ be $2/\pi$ times the anticlockwise angle change from the direction of $(\sigma_X^j(b))_X^*$ to the direction of $(\sigma_X^{j+1}(b))_X^*$. The sequence $c(b) = (c_0(b), c_1(b) \ldots c_{\lambda-2}(b))$ is called a *difference crack-code* of the simple 8-boundary of X that contains b.

It is evidently possible to reconstruct that simple 8-boundary from the difference crack-code $c(b)$ provided the position and orientation of b_X^* are known. Thus if X has exactly m 8-holes and each of $b_0, b_1, \ldots, b_m$ is a boundary element on a different simple 8-boundary of X, then it is possible to reconstruct X itself from the pairs $\{(c(b_i), b_i{}_X^*) \mid 0 \le i \le m\}$. Whenever the number of boundary elements of X is much smaller than the number of pixels in X, difference crack-codes provide a means of representing X compactly.

In addition, many properties of X can be computed directly from the pairs $\{(c(b_i), b_i{}_X^*) \mid 0 \le i \le m\}$ in no more than $O(\ell)$ time, where ℓ is the sum of the lengths of the crack-codes $c(b_i)$ (so ℓ is the total number of boundary elements of X). In fact, if f is any function for which $\sum\{f(x_0, y) \mid y_1 \le y \le y_2\}$ or $\sum\{f(x, y_0) \mid x_1 \le x \le x_2\}$ can be evaluated in $O(1)$ time, then the sum $\sum\{f(x, y) \mid (x, y)$ are the coordinates of a pixel in $X\}$ is such a property of X. Thus the area of X ($f(x, y) \equiv 1$) and all moments of X ($f(x, y) \equiv x^s y^t$) are such properties of X, and hence so are the coordinates of the centroid of X and the central moments of X (i.e., the moments of X when the origin is shifted to the centroid of X).

3 Generalization to n Dimensions

A generalization of the strong successor relation to n dimensions will now be defined which yields concise difference crack-codes for boundaries in n-dimensional binary images. This relation was introduced by Gordon and Udupa in [2] for boundaries of 3-D images. For this reason it will be called the *GU-successor* relation.

In the rest of this paper n denotes an arbitrary but fixed integer greater than or equal to 2, except that in Sect. 3.6 we will assume $n \geq 3$.

3.1 n-Dimensional Digital Topology: Digital k-Cells, α_i-Adjacency, Boundary Elements

Let k be an arbitrary integer such that $0 \leq k \leq n$. A *digital k-cell* is a cartesian product $I_1 \times I_2 \times \ldots I_n$ in which just k of the I_j are closed unit intervals of form $[i_j, i_j + 1]$ and the remaining $n - k$ of the I_j are singleton sets $\{i_j\}$, where each i_j is an integer. (Note that in the case $n = 2$ a digital n-cell is a pixel, and in the case $n = 3$ a digital n-cell is a voxel.) If C is a set of digital n-cells then $\overline{C}$ denotes the set of all digital n-cells that are not in C.

For $0 \leq r \leq k$, a *$(k-r)$-face* of a digital k-cell $I_1 \times I_2 \times \ldots I_n$ is a digital $(k-r)$-cell $I'_1 \times I'_2 \times \ldots I'_n$ where there are exactly r values of j for which I_j is a unit interval and I'_j consists of one endpoint of I_j, and where $I'_j = I_j$ for all the other values of j.

For any positive integer $i \leq n$, *α_i-adjacency* is the symmetric binary relation on digital n-cells such that one digital n-cell is α_i-adjacent to another if and only if they are distinct n-cells that share at least an $(n-i)$-face. In this paper α_i-adjacency will only be used in the cases where $i = 1, 2$, or n.

When $n = 2$, α_1 and α_2 are the standard 4-adjacency and 8-adjacency relations. When $n = 3$, α_1, α_2 and α_3 are respectively the standard 6-adjacency, 18-adjacency and 26-adjacency relations.

Every digital $(n-1)$-cell f is an $(n-1)$-face of exactly two digital n-cells. Those two digital n-cells are α_1-adjacent to each other and their intersection is just f.

A *boundary element* of a set C of digital n-cells is a digital $(n-1)$-cell $p \cap q$ in which $p \in C, q \in \overline{C}$ and p is α_1-adjacent to q. The *boundary of C*, written ∂C, is the set of all boundary elements of C. As in the 2-D case, $\partial C = \partial \overline{C}$.

In the next section the 2-D concept of a simple 8-boundary will be generalized to n-dimensional images.

3.2 Simple ρ-Boundaries

Let X be an α_1-connected finite set of digital n-cells. In the 2-D case the 8-adjacency relation was used on $\overline{X}$ to define the holes of X. When $n > 2$ it is not immediately clear what adjacency relation should be used on $\overline{X}$ to define the "cavities" of X. Any symmetric relation ρ satisfying $\alpha_1 \subseteq \rho \subseteq \alpha_n$ would be a reasonable candidate for use as the adjacency relation on $\overline{X}$. (Indeed, if in an

empirical digital image two such adjacency relations yield substantially different sets of connected components of $\overline{X}$, then one might infer that the resolution of that image is too low.) Note that ρ need not be one of the relations α_i.

For any symmetric relation ρ satisfying $\alpha_1 \subseteq \rho \subseteq \alpha_n$, a finite ρ-component of $\overline{X}$ will be called a *ρ-cavity* of X. (The concept of ρ-component is derived from ρ in the standard way. A set of digital n-cells is *ρ-connected* if it cannot be partitioned into two non-empty subsets which are such that no member of one subset is ρ-adjacent to a member of the other. A *ρ-component* of a non-empty set C of digital n-cells is a maximal ρ-connected subset of C.)

If X has just m ρ-cavities (where $m \geq 0$) then ∂X can be decomposed into $m+1$ parts, where one part separates X from the outside (more precisely, from the unbounded ρ-component of $\overline{X}$), and each of the other parts separates X from one of the ρ-cavities. We shall call each of these parts of ∂X a *simple ρ-boundary* of X. More precisely, a simple ρ-boundary of X is a set ∂Y in which Y is a ρ-component of $\overline{X}$. Thus ∂X can be partitioned into $m+1$ simple ρ-boundaries, where m is the number of ρ-cavities of X.

Note that in the case $n = 2$ a simple 8-boundary may also be called a *simple α_2-boundary*.

3.3 Review of Three Properties of the Strong Successor Relation

Before defining the GU-successor relation, we observe that in 2-D images the relation "is the strong successor relative to X of" has the following important properties for any finite 4-connected set of pixels X:

1. The relation is "locally determined". (To find the strong successor of a boundary element e relative to X it is only necessary to inspect the four pixels incident on the final point of the directed line segment e^*_X.)
2. The relation's transitive closure is an equivalence relation whose equivalence classes are the simple 8-boundaries of X.
3. In the digraph of the relation each node has just one successor and just one predecessor.

These are the properties which make it possible to represent each simple 8-boundary of X by a difference crack-code. Because of property 3, no boundary element is represented more than once in a difference crack-code.

3.4 The GU-Successor Relation

As before, let X be an α_1-connected finite set of digital n-cells. The GU-successor relation relative to X, which is defined below in Definition 1, is a relation on ∂X that satisfies property 1.

Moreover, for a certain symmetric relation P which satisfies $\alpha_1 \subset P \subseteq \alpha_2$ (see Definition 2 below) the GU-successor relation relative to X satisfies property 2 when "simple 8-boundaries" is replaced by "simple P-boundaries". This fundamental relationship between the GU-successor and P-adjacency relations

will be stated below as Theorem 3, which will also be referred to as the Main Theorem.

By themselves, these facts about the GU-successor relation are not particularly impressive because they are also true of a simpler relation on ∂X. This simpler relation may be called the *Artzy-Frieder-Herman successor relation* because it was introduced by Artzy, Frieder and Herman in [1] for 2-D and 3-D images. A definition of the Artzy-Frieder-Herman successor relation can be obtained by omitting the word "horizontal" in the definition of the GU-successor relation given below.

But for purposes of this study the GU-successor relation is a substantial improvement on the Artzy-Frieder-Herman successor relation when $n \geq 3$. This is because it also satisfies property 3 at a majority of the boundary elements of most α_1-connected sets of digital n-cells X, and because it relates each boundary element to only $2 - 2/n$ other elements, on average, whereas the Artzy-Frieder-Herman successor relation relates each boundary element to $n - 1$ others.

The definition of the GU-successor relation given below uses some additional notation and terminology, which will now be defined. Let e be a digital $(n-2)$-cell $I_1 \times I_2 \ldots I_n$ in which I_j and I_ℓ are singleton sets and $j < \ell$. Then π_e will denote the 2-D plane normal to e that passes through the centroid of e, with the cartesian coordinate system in which the first coordinate is the jth coordinate of n-space and the second coordinate is the ℓth coordinate of n-space. In the case $n = 2$, π_e is the whole Euclidean plane.

Note that if v is any digital n-cell that meets π_e then the centroid of v lies on π_e and $v \cap \pi_e$ is a pixel in the plane π_e. If Z is a set of digital $(n-i)$-cells, where $i \in \{0, 1, 2\}$, then $\pi_e(Z)$ will denote the set $\{z \cap \pi_e \mid z \in Z\} - \{\emptyset\}$. (Thus $\pi_e(Z)$ is the "cross-section" of Z determined by π_e.) According as $i = 0, 1$, or 2 $\pi_e(Z)$ is regarded as a set of pixels, pixel edges, or pixel corners in the plane π_e. If C is a set of digital n-cells and $f \in \partial C$ meets π_e, then the pixel edge in $\pi_e(\{f\})$ is a boundary element of $\pi_e(C)$, and so it has a strong successor and a strong predecessor relative to $\pi_e(C)$.

Say that a digital k-cell $I_1 \times I_2 \times \ldots I_n$ is *horizontal* if I_n is a singleton set, *vertical* if I_n is a closed unit interval.

Definition 1. Suppose C is a set of digital n-cells and $f, f' \in \partial C$. Then f' is a *GU-successor (GU-predecessor) of f relative to C* if $f \cap f'$ is a horizontal digital $(n-2)$-cell, e say, and the pixel edge in $\pi_e(\{f'\})$ is the strong successor (strong predecessor) of the pixel edge in $\pi_e(\{f\})$ relative to $\pi_e(C)$.

The GU-successor and GU-predecessor relations are illustrated in Fig. 2. In the case $n = 2$ the GU-successor and GU-predecessor relations are just the same as the strong successor and strong predecessor relations, since every digital 0-cell is horizontal according to the above definition. But when $n \geq 3$ the restriction that the shared digital $(n-2)$-cell e be horizontal substantially reduces the average number of GU-successors that a boundary element will have. It is because of this restriction that the GU-successor relation satisfies property 3 at most boundary elements.

Indeed, each vertical element of δC has just 1 pair of horizontal $(n-2)$-faces and therefore has just 1 GU-successor and just 1 GU-predecessor relative to C. Thus property 3 holds for all vertical boundary elements.

Each horizontal element of δC has just $n-1$ pairs of horizontal $(n-2)$-faces and therefore has just $n-1$ GU-successors and just $n-1$ GU-predecessors relative to C. So when $n \geq 3$ property 3 fails for horizontal boundary elements. But "on average" only $1/n$ of the elements of δC will be horizontal while $(n-1)/n$ of the elements will be vertical. Thus one might say that the expected number of GU-successors or GU-predecessors of a boundary element is $(1/n)(n-1)+(n-1)/n = 2-2/n < 2$.

3.5 The P-Adjacency Relation and the Main Theorem

A fundamental non-trivial result concerning the GU-successor relation, already mentioned in the previous section, is that there exists an adjacency relation P satisfying $\alpha_1 \subset P \subseteq \alpha_2$ such that the GU-successor relation satisfies property 2 when "simple 8-boundaries" is replaced by "simple P-boundaries". It is time to define the adjacency relation P and state the result precisely:

Definition 2. One digital n-cell is *P-adjacent* to another if they are α_1-adjacent to each other, or if they are α_2-adjacent to each other and they meet in a horizontal digital $(n-2)$-cell.

In the case $n = 2$ the P-adjacency relation is the same as the 8-adjacency relation, since every digital 0-cell is horizontal according to our definition. But when $n \geq 3$ the P-adjacency relation is strictly stronger than the α_2-adjacency relation. (See Fig. 2.)

Notice that if X is a finite α_1-connected set of digital n-cells and f is a boundary element of X then a GU-successor of f relative to X must lie in the same simple P-boundary of X as f itself. But the following theorem says much more:

Theorem 3 Main Theorem. *Let X be a finite α_1-connected set of digital n-cells. Then the transitive closure of the relation "is a GU-successor relative to X of" is an equivalence relation whose equivalence classes are the simple P-boundaries of X.*

This theorem shows that the relationship between the GU-successor relation and simple P-boundaries is a generalization to n dimensions of the relationship between the strong successor relation and simple 8-boundaries.

The Main Theorem was conjectured in the case $n = 3$ by Gordon and Udupa in [2], and a proof for the case $n = 3$ was given by Kong and Udupa in [5]. A proof that the theorem holds for all values of n is outlined in [3]. The author and Udupa plan to include a complete proof of this result in a more comprehensive paper on the GU-successor relation and its applications.

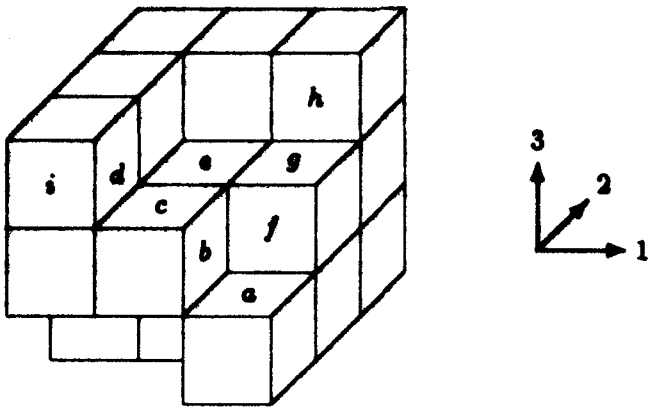


Fig. 2. In this 3-D example, the voxel with b and c as faces is P-adjacent to the voxel with d and i as faces, but is not P-adjacent to the voxel with f and g as faces. b and e are the GU-successors of c; c is the GU-successor of d; g is the GU-successor of f and is also a GU-successor of e; f is a GU-successor of a; h is a GU-successor of g; b is neither a GU-successor nor a GU-predecessor of a; f is neither a GU-successor nor a GU-predecessor of b; d is neither a GU-successor nor a GU-predecessor of i.

3.6 Difference Crack-Codes for Simple P-Boundaries

Suppose $n \geq 3$ and X is a finite α_1-connected set of digital n-cells. Whether a boundary element of X is horizontal or not depends on the way the coordinates are ordered — specifically, it depends on which coordinate is regarded as the n^{th} coordinate. Now cyclically permute the coordinates in such a way as to minimize the number of horizontal boundary elements of X. This ensures that X has at most $|\partial X|/n$ horizontal boundary elements.

Let S be any simple P-boundary of X. A "difference crack-code" for S will now be described.

Let $\mathcal{G}$ denote the digraph of the GU-successor relation relative to X. Every weak component of $\mathcal{G}$ is also a strong component, so we may refer to the *components* of $\mathcal{G}$ without ambiguity. By the Main Theorem, S corresponds to the set of nodes in a component $\mathcal{G}_S$ of $\mathcal{G}$. Since each element of ∂X has just as many GU-predecessors relative to X as GU-successors (1 or $n-1$ of each, according as the boundary element is vertical or horizontal) the indegree of each node of $\mathcal{G}$ is equal to its outdegree. Consequently $\mathcal{G}_S$ has directed Euler circuits.

Let $(b_0, b_1, \ldots, b_r)$ be a sequence of elements of ∂X that corresponds to the sequence of nodes in some directed Euler circuit γ of $\mathcal{G}_S$, where $b_r = b_0$ is a vertical boundary element and the n^{th} coordinate of the centroid of b_1 is greater than the n^{th} coordinate of the centroid of b_0. For $0 \leq i \leq r-2$ let $e_i = b_i \cap b_{i+1}$. Then e_i is a horizontal digital $(n-2)$-cell $I_1 \times I_2 \times \ldots I_n$ where I_n and exactly one other I_j are singleton sets and the other $n-2$ I_js are closed unit intervals. Let ℓ_i denote the unique positive integer $\ell \leq n-1$ for which I_ℓ is a singleton set.

For $0 \leq i \leq r-2$ let $c_i \in \{-1, 0, 1\}$ be $2/\pi$ times the anticlockwise angle change from the pixel edge in $\pi_{e_i}(\{b_i\})$ (oriented towards the centroid of e_i) to the pixel edge in $\pi_{e_i}(\{b_{i+1}\})$ (oriented away from the centroid of e_i). For $1 \leq i \leq r-2$ let $d_i = \ell_i - \ell_{i-1} \pmod{n-1}$, so $d_i \in \{0, 1, \ldots, n-2\}$. Notice that d_i can only be non-zero when b_i is a horizontal boundary element. This is because when $b_i = I_1 \times I_2 \times \ldots I_n$ is vertical $\ell_i = \ell_{i-1} = \ell$ where ℓ is the unique positive integer for which I_ℓ is a singleton set.

Call the pair of sequences $((c_0, c_1, c_2, \ldots, c_{r-2}), (d_1, d_2, \ldots, d_{r-2}))$ the *preliminary code for* S induced by the directed Euler circuit γ. If the cyclic permutation applied to the coordinates at the outset is specified, then it is possible to reconstruct S from this pair of sequences and the coordinates of the centroid of b_0.

The *difference crack-code* for S induced by γ is obtained by taking the preliminary code for S induced by γ and omitting from its sequence of d_is every d_i for which b_i is a vertical boundary element. No information is lost when all such d_is are omitted, for the following reasons. Firstly, we already know that these d_is are all equal to 0. Secondly, it is easy to determine whether b_i is vertical from the subsequence $(c_0, c_1, \ldots, c_{i-1})$. Indeed, since b_0 is vertical it follows that b_i is vertical if and only if $c_0 + c_1 + \ldots c_{i-1}$ is even. A very simple example of the crack-code is given in Fig. 3.

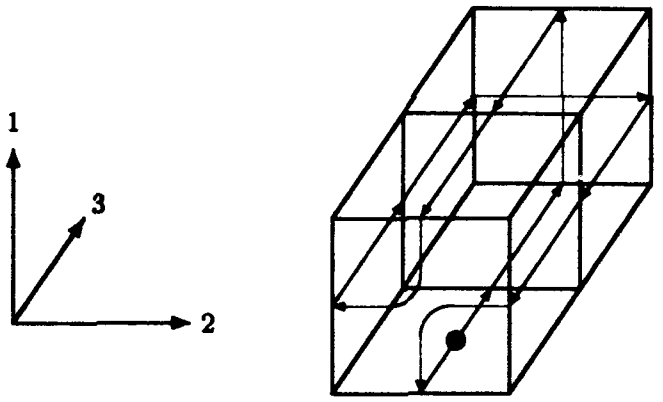


Fig. 3. Suppose $n = 3$, X is this set of two voxels, $S = \partial X$, b_0 is the dotted voxel face, and the directed Euler circuit γ is the circuit given by the arrows. Then the difference crack-code for S induced by γ is $((0, -1, -1, 0, -1, -1, 0, -1, -1, 0, -1), (0, 1, 0))$.

Let $h(S)$ denote the number of horizontal elements in S, and $v(S)$ the number of vertical elements in S. The difference crack code contains $(r-1)$ c_is and $(r-2-v(S))$ d_is. Each c_i can be stored in 2 bits and each d_i in $\lceil \log(n-1) \rceil$ bits. Hence the code can be stored in no more than $2r + \lceil \log(n-1) \rceil (r - v(S))$ bits. Horizontal elements of S occur just $n-1$ times in the sequence $(b_0, b_1, \ldots, b_{r-1})$ whereas vertical elements occur just once. So $r = (n-1)h(S) + v(S)$, and the code can be stored in no more than $2((n-1)h(S)+v(S)) + \lceil \log(n-1) \rceil (n-1)h(S)$

bits.

Let $S_0, S_1, \ldots, S_m$ be the simple P-boundaries of X, and let H denote the number of horizontal boundary elements of X. Then $\sum_{i=0}^{m} h(S_i) = H$ and $\sum_{i=0}^{m} v(S_i) = |\partial X| - H$. So if such a difference crack-code is constructed for each S_i then the number of bits needed to store all the codes is no more than $2((n-1)H + (|\partial X| - H)) + \lceil \log(n-1) \rceil (n-1)H$. Since $H \leq |\partial X|/n$, the number of bits required is no more than $(4 + \lceil \log(n-1) \rceil)(1 - 1/n)|\partial X|$.

In particular, on putting $n = 3$ or 4 it can be seen that the memory requirement of the codes is no more than $3\frac{1}{3}$ bits per boundary element in the 3-D case, and no more than $4\frac{1}{2}$ bits per boundary element in the 4-D case.

This difference crack-code is reminiscent of an Euler circuit based representation of digital surfaces suggested by Rosenfeld, Kong and Wu at the end of Section 4 in [9]. However, Theorem 3 had not yet been proved when [9] was written, and the possibility of constructing a difference crack-code based on Gordon and Udupa's GU-successor relation was not considered in that paper.

Kovalevsky [6] describes a related method of representing boundaries of 3-D binary images. Kovalevsky's representation is also based on the GU-successor relation. However, whereas the difference crack-code presented above uses a single pair of sequences to encode an entire simple P-boundary, Kovalevsky's representation does not. Instead, it regards the boundary of a 3-D binary image as a union of boundary element "hoops", each of which is represented by a separate sequence.

References

1. Artzy, E., Frieder G., Herman, G.T. (1981). The theory, design, implementation, and evaluation of a three-dimensional surface detection algorithm, Computer Graphics and Image Processing 15, pp. 1–24,
2. Gordon, D., Udupa, J.K. (1989). Fast surface tracking in three-dimensional binary images, Computer Vision, Graphics and Image Processing 45, pp. 196–214.
3. Kong, T.Y. (1993). Justification of a type of fast anisotropic boundary tracker for multidimensional binary images, to appear in: Proc. on Vision Geometry (November 15–16, 1992, Boston, Mass., USA), SPIE Volume 1832.
4. Kong, T.Y., Rosenfeld, A. (1989). Digital topology: introduction and survey, Computer Vision, Graphics and Image Processing 48, pp. 357–393.
5. Kong, T.Y., Udupa, J.K. (1992). A justification of a fast surface tracking algorithm, CVGIP: Graphical Models and Image Processing 54, pp. 162–170.
6. Kovalevsky, V.A. (1993). Topological foundations of shape analysis, this volume, pp. 21–36.
7. Newman, M.H.A. (1951). Elements of the Topology of Plane Sets of Points, 2nd Edition, Cambridge University Press, Cambridge, UK.
8. Rosenfeld, A., Kak, A.C. (1982). Digital Picture Processing, 2nd Edition, Vol. 2, Academic Press, New York.
9. Rosenfeld, A., Kong, T.Y., Wu, A.Y. (1991). Digital surfaces, CVGIP: Graphical Models and Image Processing 53, pp. 305–312.

Studying Shape Through Size Functions *

Claudio Uras[1] *and Alessandro Verri*[1,2]

[1] Dipartimento di Fisica dell'Università di Genova, Genova, Italy
[2] International Computer Science Institute, Berkeley CA, USA

Abstract. According to a recent mathematical theory the intuitive concept of shape can be formalized through functions, named *size functions*, which convey information on both the topological and metric properties of the viewed shape. In this paper the main concepts and results of the theory are first reviewed in a somewhat intuitive fashion. Then, an algorithm for the computation of discrete size functions is presented. Finally, by introducing a suitable distance function, it is shown that size functions can be useful for both shape description and recognition from real images.

Keywords: Shape description, size function.

1 Introduction

Shape description and recognition are important stages of vision. From the computational perspective, many problems stem from the well-known difficulty of dealing with qualitative and quantitative changes in shape within the same scheme.

The study of shape through integer-valued functions, called *size functions* [1, 2, 3], has recently been proposed. The key idea underlying the concept of a size function is that of setting *metric* bounds to the classical notion of *homotopy*, i.e., of continuous deformation. Size functions are very good candidates for shape representation because they (*i*) convey information about both the qualitative and quantitative structure of the viewed shape, (*ii*) can be tailored to suit the invariant properties of the shapes to be studied, and (*iii*) are inherently "stable" against small changes in shape.

The aim of this paper is to assess the potential of the theory of size functions for computer vision. Therefore, after a brief summary of the main concepts of the theory, an algorithm for the computation of size functions in the discrete case is described. Then a simple way to measure distances between size functions is

* C. Uras is supported by a fellowship from ELSAG-Bailey S.p.A. We thank Massimo Ferri and Patrizio Frosini for many helpful discussions. Patrizio Frosini and Steve Omohundro made valuable comments on the paper. Clive Prest checked the English.

proposed and tested on real images. Finally, the main conclusions which can be drawn from our research are summarised.

2 A Simple Example

First, let us introduce the notion of a size function through a simple example. The aim of this section is to generate a description of the curve in Fig. 1a which is useful for shape recognition.

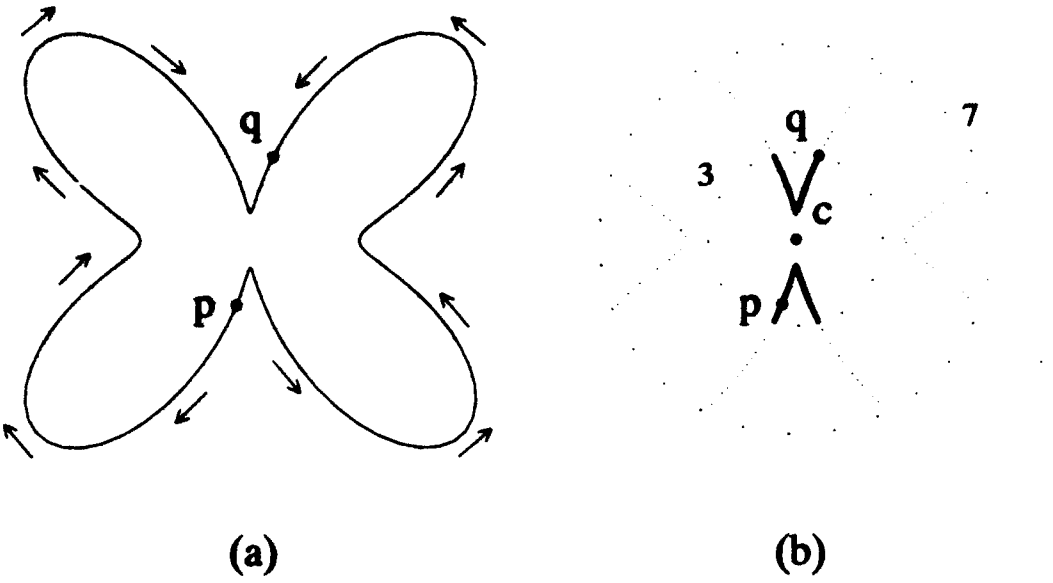


Fig. 1. Topological and metric obstructions. (a) Since the curve α has no topological obstruction, the point p can be brought into coincidence with q without leaving α by following either one of the two trajectories indicated by the arrows. (b) If D_c, the distance from the centre of mass c, along either trajectory cannot be larger than 7 (the metric obstruction), p cannot be brought into coincidence with q any longer. The size function $l_{D_c}(\alpha)$ at the point $(3,7)$ equals 2 because two of the four connected components of the set of points within the larger circle (the points with $D_c \leq 7$) contain at least a point within the smaller circle (the points with $D_c \leq 3$).

As a preliminary step, let us define a transformation H which brings a point of α onto some other point of α without leaving the curve. The arrows in Fig. 1a, for example, help visualize two possible "trajectories" along which H brings the point p onto the point q. The transformation H induces an equivalence relation on the points of α, where two points u and v are said to be *H-equivalent* if there exists a continuous trajectory on α which brings u onto v. Since, independent of the shape of α, all the points fall into the same single equivalence class, the purely topological concept of H-equivalence is clearly not sufficient to characterize the shape of α. Intuitively, this is due to the absence of "topological obstructions" between points of α.

Let us now change the definition of H-equivalence slightly by introducing "metric obstructions" along the trajectories of H on α. For example, let c be the centre of mass of α and $D_c(s)$ denote the distance between c and a point s of α. In Fig. 1b the continuous lines identify the points with $D_c \leq 3$, the dashed lines the points with $3 < D_c \leq 7$, while the points with $D_c > 7$ have not been drawn. The gaps in Fig. 1b make it clear that D_c, which is called a *measuring function*, eventually exceeds 7 (the "metric obstruction") along any trajectory from p to q. This suggests that H should be redefined so that two points u and v are said to be $H(D_c \leq y)$-*equivalent* if a trajectory exists on α from u to v along which D_c never exceeds y. It is evident that not all the points of α are $H(D_c \leq y)$-equivalent for some value of y and that the number of equivalence classes depends on the shape of α. In Fig. 1b, for example, p is not $H(D_c \leq 7)$-equivalent to q.

The notion of $H(D_c \leq y)$-equivalence is essential for the definition of size function. For each pair of real numbers (x, y), the *size function* $l_{D_c}(\alpha; x, y)$ induced by the measuring function D_c counts the number of equivalence classes in which the equivalence relation $H(D_c \leq y)$ divides the set of points of α with $D_c \leq x$ (for $y \geq x$). In practice, the size function $l_{D_c}(\alpha; x, y)$ can be computed by counting the number of connected components of the set of points of α with $D_c \leq y$ that contains at least one point with $D_c \leq x$. In the particular example of Fig. 1b we have $l_{D_c}(\alpha; 3, 7) = 2$, because the set of points with $D_c \leq 7$ has 4 connected components, two of which contain at least one point with $D_c \leq 3$ (i.e., a point on a continuous line).

The diagram of $l_{D_c}(\alpha; x, y)$ is shown in Fig. 2a (the horizontal and vertical axes are the x- and y-axes respectively). The size function $l_{D_c}(x, y)$ is a piecewise constant function which, within the triangular region $T = \{(x, y) : 0 \leq x < 10.56..., x \leq y < 10.56...\}$, equals 0 for $0 \leq x < 1$, 2 for $1 \leq x < 4$, and 4 for $4 \leq x < 10.56....$ The numbers 1, 4, and 10.56... are the critical values of the measuring function D_c (see Fig. 2b). The value of l_{D_c} elsewhere is independent of α. In essence, $l_\varphi = 0$ on the left of the vertical axis (there are no points to start with), $l_\varphi = 1$ for $y \geq 10.56...$ and $0 \geq x \geq y$ (all the metric obstructions are removed), and $l_\varphi = \infty$ for $x > 0$ and $y < x$ (each point belongs to a different equivalence class).

3 Main Definitions

Let us now define and comment on the notion of size function in more general terms. We first establish some basic notation.

In this section a shape is an n-dimensional, compact, boundaryless, piecewise C^∞ submanifold $\mathcal{M}$ of the Euclidean space E^m $(n < m)$ [1]. The set of k-tuples p of points p_i of $\mathcal{M}$, $i = 1, ..., k$, is denoted by $\mathcal{M}^k$ (in the example of Fig. 1, it was simply $k = 1$ and thus $\mathcal{M}^1 = \alpha^1 = \alpha$). If p and q are in $\mathcal{M}^k$, let $d_k(p, q) = \max_{0 \leq i \leq k}\{d(p_i, q_i)\}$ be the distance between p and q, where $d(p_i, q_i)$ is the usual Euclidean distance between p_i and q_i. The important concept of measuring function can now be defined.

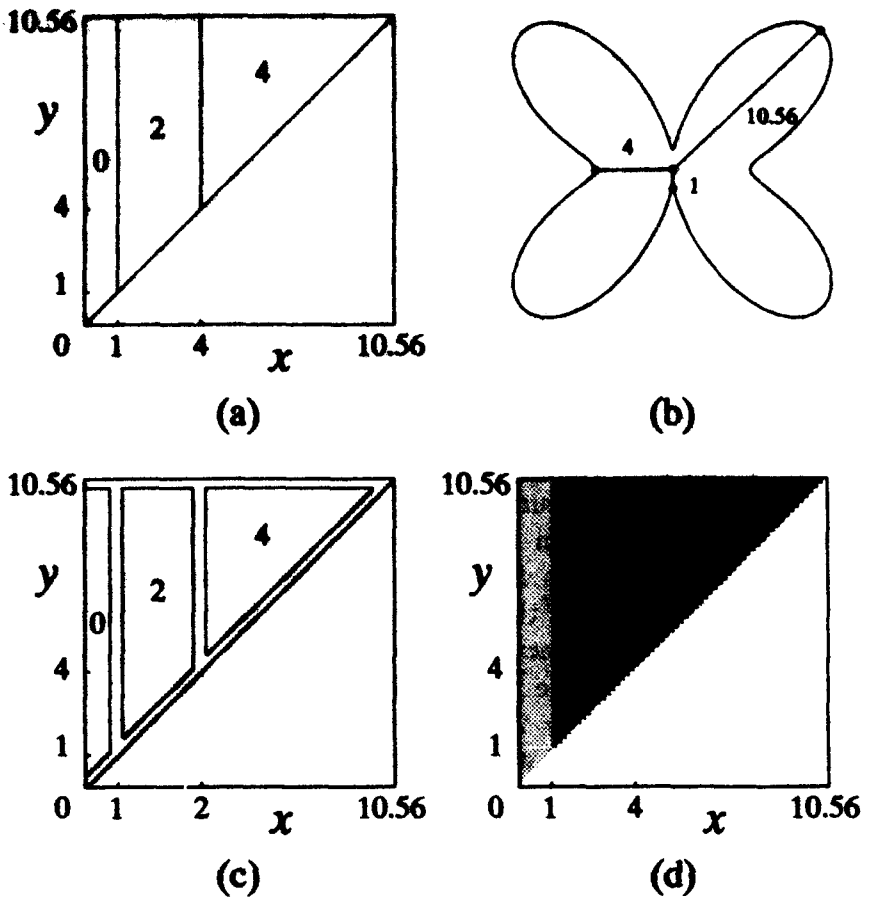


Fig. 2. Representing size functions. (a) The size function $l_{D_c}(x, y)$ of the curve α is a piecewise constant function which equals 0, 2, and 4 for $0 \leq x < 1$, $1 \leq x < 4$, and $4 \leq x < 10.56...$ respectively, within the triangular region $T = \{(x, y) : 0 \leq x < 10.56..., x \leq y < 10.56...\}$. (b) The numbers 1, 4, and 10.56... are the local minimum and maximum values of the measuring function D_c. (c) The size function $\bar{l}_{D_c}(x, y)$ (output of the algorithm described in the text) consists of thin regions, called "blind stripes", which go all around the regions where $\bar{l}_\varphi$ is known exactly and in which the "true" value of $\bar{l}_\varphi$ is bounded by the monotonicity constraints but otherwise uncertain. (d) However, a grey-value coded representation of $\bar{l}_{D_c}$ shows that the estimated value of $\bar{l}_{D_c}$ within the blind stripes is mostly correct.

Definition 1. A *measuring function* [1] is any continuous function

$$\varphi : \mathcal{M}^k \to \mathbf{R}.$$

The notion of measuring functions leads to the key concept of metric homotopy [1].

Definition 2. A *metric $H(\varphi \leq y)$-homotopy* between p and q in $\mathcal{M}^k$ is a continuous function $H : [0, 1] \to \mathcal{M}^k$ such that

- $H(0) = p$, $H(1) = q$;
- $\varphi[H(\tau)] \leq y \ \ \forall \tau \in [0,1]$.

We write $p \simeq_{\varphi \leq y} q$, if such a metric $H(\varphi \leq y)$-homotopy exists. Now let $\mathcal{M}^k(\varphi \leq x)$ be the set of points p in $\mathcal{M}^k$ with $\varphi(p) \leq x$ (e.g. the set of points on the continuous lines of Fig. 1b). We have the following:

Definition 3. The size function $l_\varphi(\mathcal{M}) : \mathbf{R}^2 \to N \cup \{+\infty\}$ [1] can be defined as

$$(x,y) \hookrightarrow \begin{cases} \#\{\mathcal{M}^{k+1}(\varphi \leq x)/\ \simeq_{\varphi \leq y}\} \text{ if finite,} \\ \\ +\infty \text{ otherwise.} \end{cases}$$

A fundamental theorem of the theory [1] ensures that the value of the size function inside the triangular region $T_\varphi(\mathcal{M}) = \{(x,y) : \varphi^{\min} \leq y \leq \varphi^{\max}, \varphi^{\min} \leq x \leq y\}$, where $\varphi^{\min}$ and $\varphi^{\max}$ are the minimum and maximum value of φ on α respectively, is finite. In what follows, for the sake of simplicity, let us assume that the value of the size function on the boundary of $T_\varphi(\mathcal{M})$ is also finite.

The definition of size function in the previous section has been extended in two important directions. First, a size function can be defined on piecewise smooth surfaces of arbitrary dimension. Second, the measuring function does not need to be the distance from the centre of mass and can be defined on k-tuples of the shape. In the case of a curve, for example, the curvature, the distance between pairs of points, and the area of the triangle whose vertices lie on the curve, could equally have been used as measuring functions with $k = 1, 2$, and 3 respectively.

A size function $l_\varphi(\mathcal{M}; x, y)$ has a number of general properties. First, by definition, l_φ is non-decreasing in x and non-increasing in y (see Fig. 2a). Second, l_φ inherits the invariant properties of the measuring function φ (thus, in the previous section $l_{D_c}(\alpha)$ is invariant for translation and rotation of α on the plane of Fig. 1a). Third, although l_φ can be defined over the entire plane, the relevant information is contained within the triangular region $T_\varphi(\mathcal{M})$.

The relevance of the notion of size function to shape analysis is due to the fact that the main properties of size functions can be extended to the discrete case [2] with little change. In the next section an algorithm for the discrete computation of size functions will be described and the main properties of the obtained representation discussed. A rigorous account of the mathematical foundations of the theory of size functions in both the continuum and discrete case can be found in [1, 2].

4 Computing Size Functions

This section describes the implementation of an algorithm for the discrete computation of the size function of a planar curve α and discusses the main properties of size functions in both the continuum and discrete case. For the sake of simplicity, the discussion is restricted to the case in which the measuring function

φ is defined on single points of α (that is, $k = 1$) with $\varphi \geq 0$ (a more general description can be found in [4]). In addition, let $B(p)_\delta$ be the open circle of centre p and radius δ, and l_φ and $\bar{l}_\varphi$ the size function in the continuum and discrete case respectively. The algorithm consists of four steps.

1. Sample (or approximate) the curve α at a finite number N of points p^i, $i = 1, ..., N$, so that (*i*) $\alpha \subset \cup_{i=1}^{N} B(p_i)_\delta$ and (*ii*) the set $B(p_i)_\delta \cap \alpha$ is non-empty and connected for $i = 1, ..., N$. Compute $\varphi(p_i)$, $i = 1, ..., N$.
2. Define the graph G whose vertices are the points p^i and whose edges link vertices which correspond to adjacent points on α.
3. Compute the maximum $\varphi^{\max}$ of $\varphi(p^i)$, $i = 1, ..., N$ and set $\Delta \geq \epsilon_\varphi(\delta)$, where $\epsilon_\varphi(\delta)$ is the modulus of continuity of φ at δ.
4. For $y = 0$ to $y \leq \varphi^{\max}$
   (a) Define the subgraph $G_{\varphi \leq y}$ of G induced by the set of vertices of G for which $\varphi \leq y$.
   (b) For $x = 0$ to $x \leq y$
      i. Let $\bar{l}_\varphi(\alpha; x, y)$ be the number of connected components of $G_{\varphi \leq y}$ which contain at least a vertex p^i such that $\varphi(p^i) \leq x$.
      ii. $x \rightarrow x + \Delta$.
   (c) $y \rightarrow y + \Delta$.

The conditions (*i*) and (*ii*) of the first step ensure that the curve α is covered in such a way that each open circle contains exactly one connected arc of α. The graph G, in the second step, is a discrete representation of α such that a path on G between the vertices p_i and p_j is the discrete counterpart of a trajectory between points of the two arcs $B(p_i)_\delta \cap \alpha$ and $B(p_j)_\delta \cap \alpha$. The third step determines the minimal resolution at which $\bar{l}_{D_c}$ can be computed. In the final step $\bar{l}_{D_c}$ is computed over a grid of equally spaced points within the triangular region $T_\varphi(\alpha) = \{(x, y) : 0 \leq y \leq \varphi^{\max}, 0 \leq x \leq y\}$.

The diagram of $\bar{l}_{D_c}$, that is, the output of the algorithm when α is sampled at 100 points, is shown in Fig. 2c. Somewhat unexpectedly, the diagram of Fig. 2c does not consist of a set of discrete estimates. This is made possible by a fundamental theorem of the theory [2] which ensures that if $\Delta x, \Delta y \geq \epsilon_\varphi(\delta)$ (see the third step of the algorithm) and $\bar{l}_\varphi = n$ at two different points, then, at "every" point in-between, $\bar{l}_\varphi = n$ and, most importantly, $l_\varphi = n$. Consequently, there are three areas in the diagram of Fig. 2c where $\bar{l}_{D_c}$ is known to be equal to l_{D_c} and to 0, 2, and 4 respectively with "no margin of error". The differences between $\bar{l}_\varphi$ and l_φ are located near the points where $\bar{l}_\varphi$ is not constant. In fact, it can be shown [2] that if $\bar{l}_\varphi$ takes on different values at two adjacent points along either axis, neither of the two estimates is *a priori* equal to l_φ. Thus, there are regions, named "blind stripes", which go all around the locations where $\bar{l}_\varphi$ is known exactly and in which the "true" value of l_φ is bounded by the monotonicity constraints along the coordinate axes but is otherwise uncertain (see Fig. 2d, however). Intuitively, the width of the blind stripes reflects the coarseness of the sampling stage and the ambiguity of the finite covering of the first step of the algorithm. A finer sampling would narrow down the width of the blind stripes,

thereby reducing the uncertainty in the location and value of the discontinuities of l_φ.

5 Experimental Results

Let us present some experimental results on the computation and use of size functions for object recognition from real images. First, a distance between size functions needs to be defined.

5.1 A Distance between Size Functions

Let φ be the measuring function with $\varphi \geq 0$, α_1 and α_2 two planar curves, and $\varphi^{\max}(\alpha_i)$ the maximum of φ on α_i, for $i = 1, 2$. Let us scale φ by defining $\hat{\varphi} = \varphi/\varphi^{\max}(\alpha_i)$ on α_i, for $i = 1, 2$. As a result, $\hat{\varphi}(\alpha_1) = \hat{\varphi}(\alpha_2) = 1$ and a scale-invariant distance D between the size functions $l_\varphi(\alpha_1)$ and $l_\varphi(\alpha_2)$ can be defined simply as

$$D(l_\varphi(\alpha_1), l_\varphi(\alpha_2)) = 2 \int_0^1 dy \int_0^y dx |l_\varphi(\alpha_1; x, y) - l_\varphi(\alpha_2; x, y)|.$$

Similarly, in the discrete case, $\bar{l}_\varphi(\alpha_1)$ and $\bar{l}_\varphi(\alpha_2)$ can be computed at the same fixed resolution R and regarded as triangular matrices $\bar{l}_\varphi(\alpha_1)_{i,j}$ and $\bar{l}_\varphi(\alpha_2)_{i,j}$ with $i = 1, \ldots, R-1$ and $j = 1, \ldots, R-i$. The distance D can then be redefined as

$$D\left(\bar{l}_\varphi(\alpha_1), \bar{l}_\varphi(\alpha_2)\right) = \frac{2}{R(R-1)} \sum_{i=1}^{R-1} \sum_{j=1}^{R-i} | \bar{l}_\varphi(\alpha_1)_{i,j} - \bar{l}_\varphi(\alpha_2)_{i,j} |, \qquad (1)$$

where the normalization factor is chosen so that $D = 1$ if, on average, the triangular matrices $\bar{l}_\varphi(\alpha_1)$ and $\bar{l}_\varphi(\alpha_2)$ differ by 1 at each entry. The entries on the diagonal of the triangular matrices are not considered because they may be affected by large quantization errors.

In order to test the algorithm for the computation of size functions of the previous section and then assess the usefulness of the concept of size functions for object recognition, some experiments on sets of "real" objects were performed. Let us now describe one of these experiments in some detail.

5.2 Leaf Recognition

Figure 3 shows the images of six leaves from six different plant species (from upper left to lower right: ivy, lemon, oleander, pittosporum, oak, and olive). Each leaf was picked from a set of eight leaves of the same species yielding a total of 48 leaves and one image of each leaf against a dark background was taken. Standard edge detection techniques were applied to extract the silhouette of each leaf [5] and the size function l_{D_c} of each leaf was then computed over a grid of fixed resolution. The distance from the centre of mass was always

normalized between 0 and 1. In order to test the invariance of l_{D_c} for translation and rotation, the position and orientation of each leaf on a plane nearly parallel to the image plane was varied from image to image.

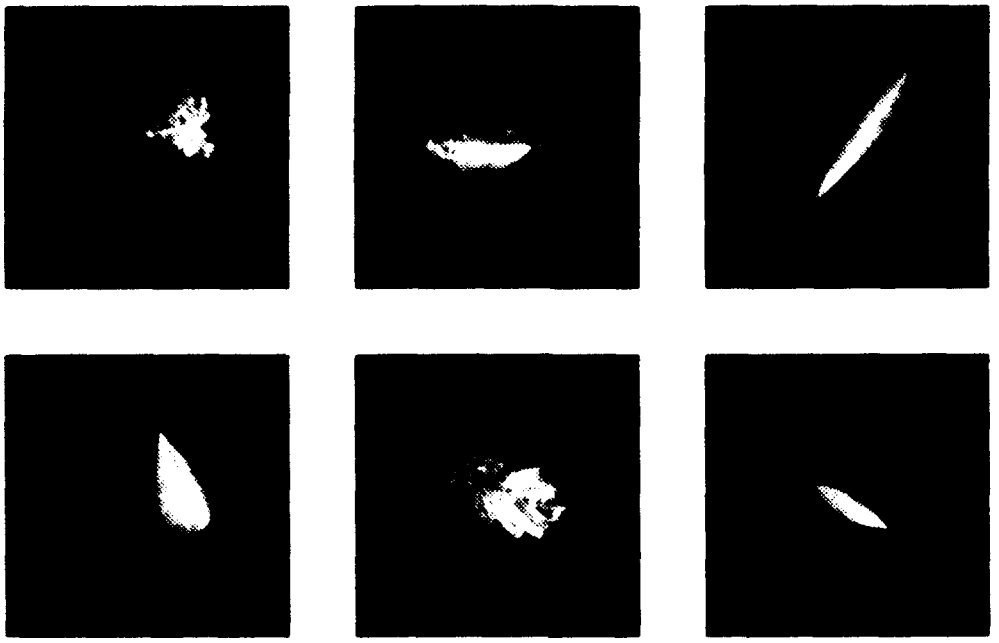


Fig. 3. Six images of six different species of leaves. From upper left to lower right: ivy, lemon, oleander, pittosporum, oak, and olive.

Figure 4 shows the grey-value coded "size functions" which have been obtained by averaging the size functions obtained from leaves of the same species (the value at each point of these "size functions" is the average of the value of the size functions at that point). A number of qualitative conclusions can already be drawn by a simple inspection of Fig. 4.

First, the size functions of the ivy leaves appear to be consistently different from all the other size functions. Second, the size functions of oleander and olive leaves are qualitatively similar but quantitatively different (the location of the main discontinuity along the horizontal axis of the oleander size function is further to the left). Third, the difference between the size functions of oak and lemon leaves is localized near the diagonal (which is where the shape "details" can be detected). Fourth, it might not be easy to distinguish between lemon, olive, and pittosporum leaves. Lastly, the size functions of lemon, pittosporum, and oak leaves present a higher degree of variability (larger regions over which the average takes on non-integer values). It is interesting to note that similar conclusions could have been drawn by looking at the original set of leaves.

From the quantitative point of view, the average "size functions" of Fig. 4 have been used to classify each leaf according to a simple recognition scheme. Each "size function" of Fig. 4 has been considered as a "model" for each of the six species, and the distance D between each leaf and each model has been computed by using (1). Finally, each leaf was classified depending on the minimum distance.

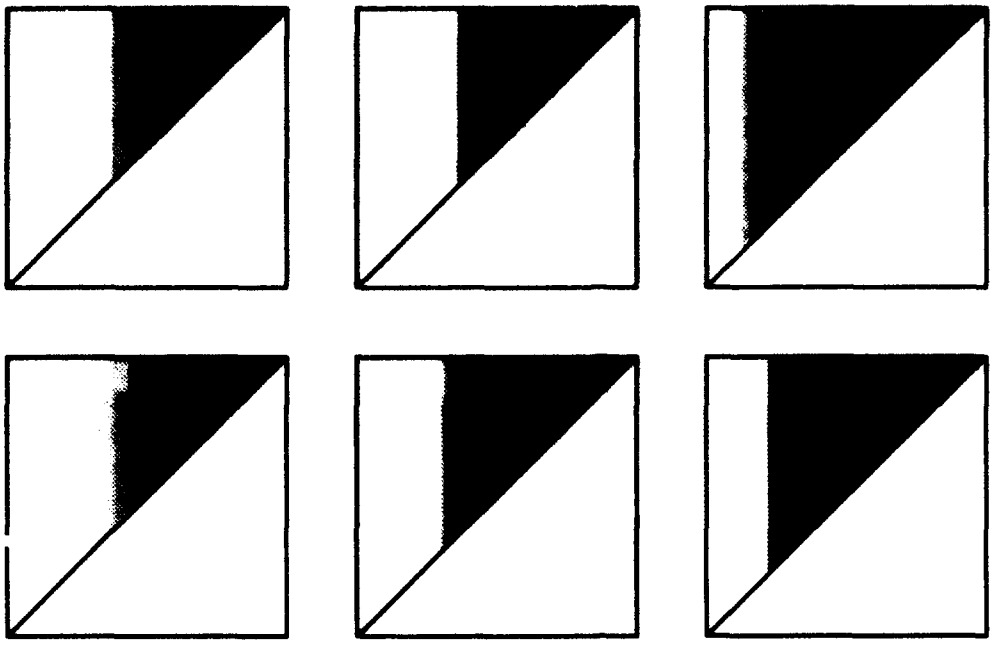


Fig. 4. Grey-value coded "size functions" which have been obtained by averaging the size functions obtained from leaves of the same species. From upper left to lower right: ivy, lemon, oleander, pittosporum, oak, and olive.

The method was able to classify all the leaves correctly with the exception of a lemon leaf which was mistaken for a pittosporum leaf. A more robust technique which combines descriptions from different size functions and correctly classifies each leaf will be described in a forthcoming paper [6].

Similar results (with no classification errors) have been obtained by looking at images of tools such as pliers, screwdrivers, scissors, wrenches, and hammers of different sizes and quantitatively different shapes.

6 Conclusion

The aim of this paper was to assess the relevance of a recent mathematical theory in relation to the problems of shape description and recognition. The theory is based on the concept of size function which combines topological and metric properties of shape.

Our analysis has shown that size functions in the discrete (*i*) can be computed reliably from real images, (*ii*) preserve the invariant properties of the chosen measuring function, (*iii*) can easily be made independent of change-of-scale, and (*iv*) are inherently robust against small qualitative and quantitative changes of shape. In conclusion, experiments on real images indicate that the shape representation which can be obtained through size functions is likely to be suitable for object recognition.

References

1. Frosini, P. (1992). Metric homotopies, to appear.
2. Frosini, P. (1992). Discrete computation of size functions. Journal of Combinatorics, Information and System Sciences, to appear.
3. Frosini, P. (1991). Measuring shapes by size functions. Proc. of SPIE on Intelligent Robotic Systems, Boston MA.
4. Uras, C. (1992). Riconoscimento di forme con strategie metrico-topologiche. Tesi di Laurea in Fisica, Università di Genova.
5. De Micheli, E., Caprile, B., Ottonello, P., and Torre, V. (1989). Localization and noise in edge detection. IEEE Trans. Patt. Anal. Mach. Intell. 11, pp. 1106–1117.
6. Uras, C. and Verri, A. (1993). Describing the shape of non-rigid objects with size functions, IEEE Workshop on Qualitative Vision, New York, June 1993.

Introduction to Categorical Shape Theory, with Applications in Mathematical Morphology

Mirek Hušek

Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 18600 Prague, Czech Republic

Abstract. An introduction to category theory for workers in shape theory with some elementary examples, including some from mathematical morphology.

Keywords: categorical shape theory, mathematical morphology.

Introduction

Category theory was founded in 1945 by Eilenberg and Mac Lane [4] in order to define natural transformations (see Definition 4). After the basic paper by Mac Lane [11] and the usage of categories in algebraic topology and homological algebra (started by Eilenberg and Steenrod [5]) category theory developed rapidly, motivated mostly from algebra. In the 1960s its rich usage in continuous structures like topological spaces or topological groups started, and also relations with set theory were investigated. Recently, useful applications in computer science have been found. Generally, one can say that category theory is an abstract language convenient for describing general situations in mathematics; it helps in finding similarities in theories of various areas of mathematics.

This paper is written for workers in shape theory who are not aware of possible usage of category theory in their work. As further reading I would recommend the recently published book by Adámek, Herrlich and Strecker [1], and the classic book by Mac Lane [12].

1 Basics of Category Theory

Perhaps the most illustrative example of a category is that of sets. In the next definition, the reader should bear in mind that $\mathrm{ob}\mathcal{C}$ is the class of sets, and $\mathcal{C}(A,B)$ is the set of mappings from the set A into the set B. As can be seen even from that example, we must distinguish between sets and proper classes. Readers who are not too familiar with such set-theoretical notions and who work only with small sets (e.g., with finite sets, or with at most countable sets, etc.) need not be concerned.

Definition 1. A *category* C is composed of a class $\mathrm{ob}C$, of disjoint sets $C(A,B)$ for $A, B \in \mathrm{ob}C$, and of an associative composition $\circ$ of members of $\bigcup_{A,B} C(A,B)$ satisfying the properties:
(a) $f \circ g$ is defined for $g \in C(A,B), f \in C(C,D)$ iff $B = C$;
(b) for each $A \in \mathrm{ob}C$, there exists $1_A \in C(A,A)$ such that $1_A \circ f = f$ and $g \circ 1_A = g$ whenever the composition is defined (i.e., if $f \in C(B,A), g \in C(A,B)$ for some object B).

The members of $\mathrm{ob}C$ are called *objects* of C, the members of $C(A,B)$ *morphisms* (or *arrows*) from A to B. The fact that $f \in C(A,B)$ is often denoted by $f : A \to B$ or by $A \xrightarrow{f} B$; then A is called the *domain* of f, B is called the *range* or *codomain* of f. The composition $f \circ g$ from (a) of Definition 1 can be expressed by

$$A \xrightarrow{g} B \xrightarrow{f} D.$$

Remarks. (1) In Definition 1 two assumptions are made about $C(A,B)$ that need not be regarded as important in the usual practice. First of all the fact that $C(A,B)$ is requested to be a set, not a proper class. The other assumption is disjointness of all the sets $C(A,B)$. It is an auxiliary or technical assumption. Usually one does not worry about it in practice; in case the sets $C(A,B)$ are not disjoint, it is easy to change their definition to make them disjoint (e.g., instead of f, one takes triples (f,A,B)).

(2) If in Definition 1, condition (a), one replaces the equality $B = C$ by the equality $A = D$, we have again a category with the same objects and morphisms but with the "converse" composition. This category is called *dual* to the original one. One may get the dual category by converting arrows: if $f : A \to B$ in the original category, then $f : B \to A$ in the dual category. Such a "duality" helps a lot with definitions and proofs. The situation is similar to the concept of conjugate matrix. If one defines a notion for categories, it is defined also in the dual category and can be carried back to the original category by converting arrows to get its dual notion. See Definition 6 for an example.

(3) Careful readers might notice that all the objects A are in one-to-one correspondence with the unit morphisms 1_A, and so categories can be defined without mentioning objects at all, as a partial semigroup (being a class in general, not a set) having enough units.

Before we come to examples, two kinds of special morphisms will be defined, namely isomorphisms and retracts, and then subcategories. Other important special morphisms like monomorphisms (corresponding to one-to-one maps) or epimorphisms (corresponding to surjections, i.e., onto-maps) will not be defined. Note that only those concepts which are defined usually by means of points can (and must) be defined in categories without using points.

The definition of subcategories is quite natural and corresponds to subsemigroups if one adopts the approach explained in the last remark (item (3)).

Definition 2. (a) A morphism $f : A \to B$ in a category $\mathcal{C}$ is called *isomorphism* if there is a morphism $g : B \to A$ in $\mathcal{C}$ with $f \circ g = 1_B, g \circ f = 1_A$ (g is often denoted by f^{-1}).
(b) A morphism $f : A \to B$ in a category $\mathcal{C}$ is called *retraction* if there is a morphism $g : B \to A$ in $\mathcal{C}$ with $f \circ g = 1_B$.
(c) A category $\mathcal{A}$ is a *subcategory* of a category $\mathcal{C}$ if
$\mathrm{ob}\,\mathcal{A} \subset \mathrm{ob}\,\mathcal{C}$,
$\mathcal{A}(A, B) \subset \mathcal{C}(A, B)$ for all objects A, B of $\mathcal{A}$,
every $\mathcal{A}(A, A)$ contains $1_A \in \mathcal{C}(A, A)$,
the composition of $\mathcal{A}$ is the restriction of that of $\mathcal{C}$ to $\mathcal{A}$.
If $\mathcal{A}(A, B) = \mathcal{C}(A, B)$ for all objects A, B of $\mathcal{A}$ then $\mathcal{A}$ is said to be *full* in $\mathcal{C}$.

Remarks. (1) Notice that if $\mathcal{A}$ is a full subcategory of $\mathcal{C}$, then it is completely described inside $\mathcal{C}$ by specifying its objects. Usually one says that $\mathcal{A}$ is a full subcategory of $\mathcal{C}$ *generated* by the class of objects $\mathrm{ob}\,\mathcal{A}$.

(2) In any category, isomorphic objects cannot be distinguished by means of the inner procedures of the category. That means, that uniqueness of objects having a given property is always up to isomorphism (an object of a category $\mathcal{A}$ has a property defined inside $\mathcal{A}$ iff every object isomorphic to it has that property, too).

We shall start with examples of structures on sets. In that case, the morphisms are mappings between the underlying sets (usually, they preserve the structure in some sense) and the composition is just the usual composition of mappings. The unit morphisms 1_A are the identity mappings. So it suffices to describe the category by describing the objects and morphisms. One must check that the identity mappings are morphisms, and that the composite of two morphisms is again a morphism.

Examples. (1) The category Set of sets: the class of objects is the class of sets and the set of morphisms $\mathrm{Set}(A, B)$ is the set of all mappings from A to B. The isomorphisms in Set are exactly the bijections (one-to-one and onto-maps); retractions are exactly onto-maps.

(2) The category Gr of groups: the class of objects is the class of groups and the set of morphisms $\mathrm{Gr}(A, B)$ is the set of all group homomorphisms from A to B. The isomorphisms in Gr are exactly group isomorphisms.

The category AbGr is the full subcategory of Gr generated by the Abelian groups, that is, between Abelian groups the homomorphisms are the same as the group homomorphisms.

(3) The category Top of topological spaces: the class of objects is the class of topological spaces and the set of morphisms, $\mathrm{Top}(A, B)$ is the set of all continuous mappings from A to B. The isomorphisms in Top are exactly the homeomorphisms onto.

One can define many full subcategories of Top like Comp generated by all compact spaces, Haus generated by all Hausdorff spaces, etc.

(4) The category Met_u has for objects the metric spaces and for morphisms the uniformly continuous mappings. The category Met_c has for objects the metric

spaces and for morphisms the continuous mappings (note that $\mathbf{Met}_c$ is not a subcategory of Top because different metrics may induce the same topology). The category $\mathbf{Met}_l$ has for its objects the metric spaces and for morphisms the non-expansive maps (i.e., the Lipschitz maps f with the Lipschitz constant 1: $d(fx, fy) \leq d(x, y)$).

Clearly, $\mathbf{Met}_l$ is a (non-full) subcategory of $\mathbf{Met}_u$ which is a (non-full) subcategory of $\mathbf{Met}_c$. The isomorphisms in $\mathbf{Met}_l$ are the distance-preserving onto-maps.

(5) The category Poset of partially ordered sets: the class of objects is the class of partially ordered sets, and the set of morphisms, $\mathbf{Poset}(A, B)$ is the set of all order-preserving mappings from A to B. The isomorphisms in Poset are the order-isomorphisms.

The category Latt is the (non-full) subcategory of Poset with the class of objects being lattices, and with $\mathbf{Latt}(A, B)$ being the mappings from A to B preserving infima. One can define here various modifications taking into account other kind of mappings (preserving suprema or both infima and suprema).

The next example belongs to those categories in which objects are points of some structure and morphisms are relations between the points:

Example. If X is a partially ordered set, then the category $\mathcal{X}$ determined by X has for its class of objects the set X, and $\mathcal{X}(x, y)$ is at most a one-element set: it is non-empty iff $x \leq y$. If one wants to specify the morphisms, it is possible to define $\mathcal{X}(x, y) = (x, y)$ iff $x \leq y$, i.e., the class (in fact, the set) of morphisms of $\mathcal{X}$ coincides with order regarded as a subset of $X \times X$. The composition here is in fact the transitivity of the order. The isomorphisms are identities.

If $Y \subset X$, then Y determines the full subcategory $\mathcal{Y}$ of $\mathcal{X}$ generated by the points of Y.

It is said that topological spaces were defined in order to define continuous mappings. One may understand in general that relations between objects are sometimes more important than the objects themselves. If categories are objects of interest, there should be some relations between them; we shall define them as functors. Of course, one can continue: if functors are objects of interest, there should be some relations between them, and these can be defined (by natural transformations). We shall not go into higher levels. Since categories may be regarded as partial semigroups, functors should be mappings between them, preserving units and composition.

Definition 3. (a) A *functor* F from a category $\mathcal{A}$ to a category $\mathcal{B}$ is a mapping which maps objects of $\mathcal{A}$ into objects of $\mathcal{B}$ and sets $\mathcal{A}(A, B)$ into $\mathcal{B}(FA, FB)$ such that it preserves the units and the composition. Notation: $F : \mathcal{A} \to \mathcal{B}$.

(b) Two categories $\mathcal{A}, \mathcal{B}$ are said to be *isomorphic* if there exist functors $F : \mathcal{A} \to \mathcal{B}, G : \mathcal{B} \to \mathcal{A}$ with $F \circ G = 1_{\mathcal{B}}$ and $G \circ F = 1_{\mathcal{A}}$. In that case the functors F, G are called *isofunctors*.

Remarks. (1) The last conditions of Definition 3 (a) mean that $1_{FA} = F(1_A)$ for each object of $\mathcal{A}$ and that $F(f \circ g) = Ff \circ Fg$ whenever $f \circ g$ makes sense, i.e.,

$$F(A \xrightarrow{g} B \xrightarrow{f} C) = FA \xrightarrow{Fg} FB \xrightarrow{Ff} FC\,.$$

(2) The notion of functor we have just defined has a more precise term, namely *covariant functor*. That is because there is also a *contravariant functor*, which is, in fact, a (covariant) functor into the dual category of the original range category. For the purpose of this study, the first kind of a functor is sufficient; when one needs the second kind, the dual category can be used.

Examples. (1) Except for the identity functor $1_{\mathcal{A}} : \mathcal{A} \to \mathcal{A}$, the simplest examples are the so-called *forgetful functors* from categories of structures into Set or into categories having some weaker structures. The forgetful functors Gr $\to$ Set, Top $\to$ Set "forget" the structure; they assign to a group or to a topological space its underlying set (i.e., the set X to a topological space $(X, \mathcal{G})$, and the map $f : X \to Y$ is assigned to a continuous map $f : (X, \mathcal{G}) \to (Y, \mathcal{H})$). The forgetful functor Ring $\to$ AbGr forgets the multiplication structure of rings but not the additive structure.

(2) If $\mathcal{A}$ is a subcategory of $\mathcal{B}$ then the inclusion of the class of objects of $\mathcal{A}$ into that of $\mathcal{B}$, together with, for each pair of objects A, B of $\mathcal{A}$, the inclusion of $\mathcal{A}(A, B)$ into $\mathcal{B}(A, B)$ form a functor.

(3) If X, Y are partially ordered sets, then functors from $\mathcal{X}$ into $\mathcal{Y}$ (the categories determined by X, Y, resp.) are exactly the order-preserving mappings $X \to Y$.

(4) If S, T are semigroups (or groups), then functors from $\mathcal{S}$ into $\mathcal{T}$ (the categories determined by S, T, resp.) are just the homomorphisms $S \to T$.

Definition 4. (a) A *natural transformation* $\eta : F \to G$ between functors $F, G : \mathcal{A} \to \mathcal{B}$ is a class of morphisms $\{\eta_A : FA \to GA\}$, indexed by objects of $\mathcal{A}$ and having the property that for every $f \in \mathcal{A}(A, A')$ one has $Gf \circ \eta_A = \eta_{A'} \circ Ff$, i.e., the following diagram commutes.

$$\begin{array}{ccc} FA & \xrightarrow{\eta_A} & GA \\ {\scriptstyle Ff}\downarrow & & \downarrow{\scriptstyle Gf} \\ FA' & \xrightarrow[\eta_{A'}]{} & GA' \end{array}$$

(b) Two functors F, G are said to be *naturally isomorphic* if there exist natural transformations η, ϵ such that $\eta \circ \epsilon = 1_F, \epsilon \circ \eta = 1_G$.

Examples. (1) Let Ban be the category of Banach spaces and continuous linear mappings, F be the functor Ban $\to$ Ban assigning to each Banach space B its second dual B^{**}, and to a continuous linear mapping $f : B_1 \to B_2$ its second dual $f^{**} : B_1^{**} \to B_2^{**}$. The canonical embeddings $B \to B^{**}$ generate a natural transformation $1_{\mathsf{Ban}} \to F$.

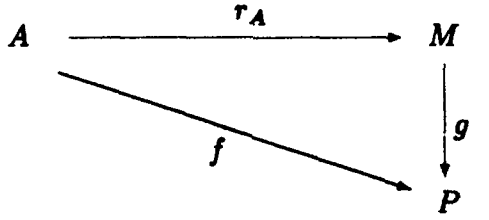
(2) The classical example showing the difference between natural transformations and "usual" transformations is the category $\mathcal{A}$ of all finite-dimensional vector spaces and of linear mappings. It is well-known that the first dual X^* is isomorphic to X for every finite-dimensional vector space X, but there is no

natural transformation expressing that isomorphism because it depends on the choice of a base in X; however the embedding $X \to X^{**}$ does not depend on the choice of a base.

2 Shapes, Closings and Reflections

In this section possible applications of category theory in shape theory will be described. We shall start with a simple case where we have a class of objects with a given subclass (members of which we may call models or prototypes), and our task is to assign to each object a model which is as close to the object as possible. In other words, we are looking for the shape of an object, expressed by means of properties of models. One may take for the class of objects all letters of a set of typewriters and for the class of models the letters of one specified typewriter; our task is assign to a letter, say "J", of a typewriter the closest letter of the specified typewriter. We must know what "close" means and so, we shall suppose that the objects together with relations between them form a category $\mathcal{C}$; the models form a subcategory $\mathcal{M}$. To describe precisely the category and its subcategory may not be easy in practice (e.g., in the example with letters of typewriters it will not be easy to describe the relations, i.e. the morphisms) and so we shall use for an illustrative example the following slightly artificial one. Let $\mathcal{C}$ be the category determined by the partially ordered set of all subsets of the plane and $\mathcal{M}$ be its full subcategory generated by closed convex sets. To say that a model M is "close" to an object A could mean that there is a (specified) morphism $f \in \mathcal{C}(A, M)$ (in our example it means $A \subset M$). The model M which is closest to an object A is then described in the following way:

There exists a morphism $r_A : A \to M$ such that for any other morphism $f : A \to P$ with P being a model (an object of $\mathcal{M}$) there exists a morphism $g : M \to P$ in $\mathcal{M}$ such that $f = g \circ r_A$, i.e., the following diagram is commutative.



In our example the closest model to a set A is its closed convex hull, of course. In that connection, a reader may ask why the relation was defined by means of single morphisms and not by the relation $\mathcal{C}(A, M) \neq \emptyset$. That is also possible, but it is equivalent to the investigation of a partial order on the class of objects. The more relations (morphisms) between objects we have at our disposal, the more information we get and the closer model we can find (but also, the more complicated the investigation we will have).

In our example, the closest model is unique and corresponds to our intuition of what it should be. Is it so in general? Consider the following example: Take the category Set as $\mathcal{A}$ and its full subcategory generated by infinite sets as $\mathcal{M}$.

The closest model in our sense described above, say to the singleton, exists and it is an arbitrary infinite set.

To avoid such an unwanted situation, one must add uniqueness somewhere; the best way is to require it for the morphism g in the diagram above. Using that modified version of the above notion of the closest model, we come to the following definition.

Definition 5. Let $\mathcal{C}$ be a category and $\mathcal{M}$ a subcategory. An object M of $\mathcal{M}$ is said to be a *reflection* of an object A of $\mathcal{C}$ if there exists a morphism $r_A : A \to M$ such that for each morphism $f : A \to P$ with $P \in \mathrm{ob}\,\mathcal{M}$ there exists a unique morphism $g : M \to P$ in $\mathcal{M}$ with $f = g \circ r_A$.

If every object of $\mathcal{C}$ has a reflection in $\mathcal{M}$, then $\mathcal{M}$ is said to be *reflective* in the category $\mathcal{C}$.

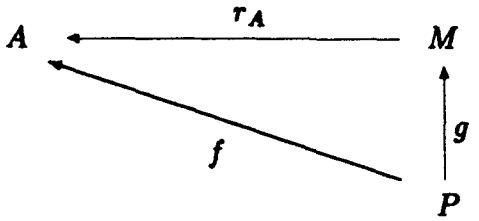
In the case where the uniqueness of the shape is not required, it is possible to use the notion described before Definition 5, that is, not to require the uniqueness of the morphism g. Such situations have been studied in categories and they are called *weak reflections* (see e.g., the survey paper [9]).

Using the last definition, we may illustrate the duality mentioned in remarks after the definition of categories. We shall reverse all the arrows in the above definition and get the notion of coreflectivity.

Definition 6. Let $\mathcal{C}$ be a category and $\mathcal{M}$ its subcategory. An object M of $\mathcal{M}$ is said to be a *coreflection* of an object A of $\mathcal{C}$ if there exists a morphism $r_A : M \to A$ such that for each morphism $f : P \to A$ with $P \in \mathrm{ob}\,\mathcal{M}$ there exists a unique morphism $g : P \to M$ in $\mathcal{M}$ with $f = r_A \circ g$.

If every object of $\mathcal{C}$ has a coreflection in $\mathcal{M}$, then $\mathcal{M}$ is said to be *coreflective* in the category $\mathcal{C}$.

The commutative diagram for coreflections looks as follows:



It should be mentioned that we could proceed dually from the beginning of this section, and define the closest model for an object as the coreflection.

Examples. (1) The subcategory AbGr is reflective in Gr. Indeed, the reflection of a group G is the quotient group G/C, where C is the commutator of G, i.e., the smallest subgroup of G containing all the elements $aba^{-1}b^{-1}$ for $a, b \in G$.

(2) The full subcategory of torsion abelian groups is coreflective in AbGr. The embedding of the torsion subgroup T of G into G is the coreflection.

(3) The following full subcategories of Top are coreflective in Top: the ones generated by all discrete spaces (or by locally connected spaces, or by sequential spaces). One gets the coreflection by modifying the given topology on the same

underlying set: take the smallest topology having the given property which is larger than the given topology. All coreflections in Top can be described in that way.

(4) Except in trivial cases like indiscrete spaces, to prove that a subcategory of Top is reflective in Top, one must know more about properties of reflections. So, here we shall just say that, for example, Haus, Comp are reflective in Top.

(5) The nonfull subcategory of Poset composed of partially ordered sets having suprema of nonvoid subsets and of mappings preserving suprema, is reflective in Poset.

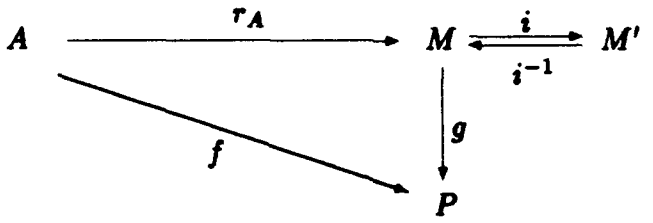
(6) If a partially ordered set A is regarded as a category $\mathcal{A}$ and $B \subset A$ is regarded as a full subcategory $\mathcal{B}$ of $\mathcal{A}$, then $a \in A$ has a reflection in $\mathcal{B}$ iff the infimum of all elements of B greater than a exists in B; then this infimum is the reflection. Dually for coreflections.

If B is reflective in $(A, \leq)$ and we assign to every $a \in A$ its reflection Ra, then the map $R : A \to B$ is idempotent ($R \circ R = R$), extensive ($a \leq Ra$) and order preserving. Conversely, every idempotent, extensive and order preserving map $R : A \to A$ gives rise to a reflective subcategory $B \subset (A, \leq)$, namely $B = \{a : Ra = a\}$. For coreflections one must use compressive maps ($Ra \leq a$) instead of the extensive ones.

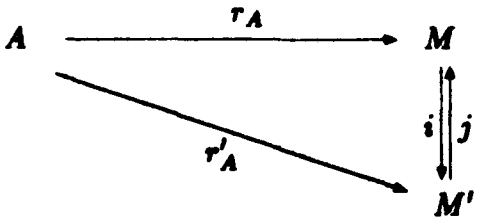
The uniqueness of the shape is expressed in the following result.

Theorem 7. *Let $\mathcal{M}$ be a subcategory of a category $\mathcal{A}$. A reflection in $\mathcal{M}$ of an object A is determined uniquely up to isomorphism in $\mathcal{M}$.*

Proof. Let $\mathcal{C}$ be a category and $\mathcal{M}$ its subcategory. If M is a reflection of A in $\mathcal{M}$ and $i : M \to M'$ is an isomorphism in $\mathcal{M}$, then M' is also a reflection of A in $\mathcal{M}$:



Conversely, if $r_A : A \to M, r'_A : A \to M'$ are two reflections of A in $\mathcal{M}$, then there are (by Definition 5) $i \in \mathcal{M}(M, M'), j \in \mathcal{M}(M', M)$ such that $i \circ r_A = r'_A, j \circ r'_A = r_A$, as in the following diagram:



Then $i \circ j \circ r'_A = r'_A = 1_{M'} \circ r'_A$; because of the uniqueness of the factorisation via r'_A, one has $i \circ j = 1_{M'}$ and similarly $j \circ i = 1_M$. Consequently, M, M' are isomorphic in $\mathcal{M}$. □

Remark. Let $\mathcal{M}$ be reflective in $\mathcal{A}$. It is an exercise to show that assigning to $A \in \mathcal{A}$ its reflection RA is idempotent (in the sense that $R(RA)$ is isomorphic to RA in $\mathcal{M}$) iff $\mathcal{M}$ is full in $\mathcal{A}$. In shape theory, it is natural to assume that the shape of a model (prototype) M is M itself; to achieve that situation, one must have $\mathcal{M}$ full in $\mathcal{A}$. Perhaps, it may be convenient in some situations to accept also non-idempotency of the shape.

Example. Now, we shall look at connections between mathematical morphology and reflections. To do so, we shall recall several basic concepts (for more details see other papers concerning mathematical morphology in this volume, e.g., the contributions of Heijmans [6] or of Roerdink [20]). Mathematical morphology is a part of image analysis. When investigating objects, it transforms them conveniently. If one looks carefully at such transformations, one can see very general features which are not difficult to express in categorical language. The transformations were investigated from a topological point of view [15], later also from a lattice–theoretical point of view [23, 7, 21] and applied to lattices of functions (grey-level morphology) [18, 10]. The following transformations are of particular interest. If $X, B \subset \mathbb{R}^n$ then the

dilation of X by B is the set $\{x : X \cap (x - B) \neq \emptyset\}$,
erosion of X by B is the set $\{x : B + x \subset X\}$,
closing of X by B is the set $\{x : y - x \in B \Rightarrow X \cap (y - B) \neq \emptyset\}$,
opening of X by B is the set $\bigcup\{B + x : B + x \subset X\}$.

Other important examples are similar transformations of functions used in the investigation of grey–level images: If f is a real–valued function on $\mathbb{R}^n$ and B is again a subset of $\mathbb{R}^n$, then

dilation of f by B is the function $\sup\{f(x - b) : b \in B\}$,
erosion of f by B is the function $\inf\{f(x + b) : b \in B\}$,
closing of f by B is the function $\inf\{\sup\{f(x + b - b') : b' \in B\} : b \in B\}$,
opening of f by B is the function $\sup\{\inf\{f(x - b + b') : b' \in B\} : b \in B\}$.

In a sense, dilations form B–hulls and erosions B–interiors of the investigated object and they may be rather far from it. Closings and openings are closer to the object and can describe it in a better way; that is, closings fill small holes and gulfs, openings delete small promontories, and both smooth some edges.

As defined here, the closing is the composition of dilation followed by erosion; the opening is the composition of erosion followed by dilation. There are other approaches to the above notions, for example, not by using the set B but by testing the set X or the function f by an area or by a function. Moreover, the above definitions of erosion, dilation, opening, and closing can be used without any change in Abelian groups instead of in Euclidean spaces (one did not use any geometric or topological properties of the given space). If one takes for B a ball with a given radius, then it is easy to change the definitions to work in

metric spaces (and in uniform spaces), or in another kind of distance space: If (X, d) is a metric space and $\varepsilon > 0$, one can define an ε-dilation (resp. ε-erosion) of a set $A \subset X$ as the set

$$U_\varepsilon(A) = \{x \in X : d(x, A) < \varepsilon\} \text{ (resp. } V_\varepsilon(A) = \{x \in X : U_\varepsilon(x) \subset A\}).$$

Then the closing $V_\varepsilon(U_\varepsilon(A))$ contains the metric closure of A but is usually bigger (the metric closure of A is the intersection of the above closings for all positive ε s).A similar statement holds for openings and metric interiors.

Sometimes, openings and closings can be defined without any reference to dilations and erosions; if we are testing a set of pixels and cannot decide whether certain pixels belong to the set or not, the extreme solutions are to add them (closing) or to delete them (opening). The natural question arises: can one assign to such primarily defined closings and openings some corresponding erosions and dilations? To answer that question, one must have a categorical description of the properties of all the concepts under consideration. We shall do it now for closings and openings. For erosions and dilations, it will be done in the next section; there one can find the answer to the above question.

Take for X a Euclidean space or an Abelian group or a metric space and let $\mathcal{A}$ be the category determined by the partially ordered set of all subsets of X, for $\mathcal{B}$ its full subcategory generated by all closings with respect to some B. Then $\mathcal{B}$ is reflective in $\mathcal{A}$, reflection of $A \subset X$ is its closing. Similarly, openings form a coreflective subcategory of $\mathcal{A}$.

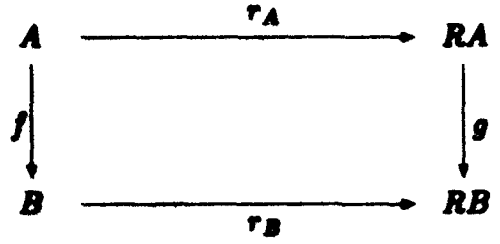
In fact, we may say that two sets have the same shape with respect to a set B if they have the same closing with respect to B (or if they have the same opening, or if they have the same closing *and* opening).

In summary one may say that looking at closings as reflections and at openings as coreflections (in partially ordered sets, or in lattices, or in any other situation) should be a unifying point of view for investigating their general properties and relations.

Coming back to general theory, we could see in the example of reflections in partially ordered sets that reflections generate some mappings. More generally, every morphism between objects determines uniquely a morphism between the corresponding reflections and this assignment generates a functor. Thus every relation between objects under consideration determines a relation between the corresponding models of the objects (or in other words, every relation between objects determines a relation between their shapes - but not conversely).

Theorem 8. *If $\mathcal{M}$ is reflective in $\mathcal{C}$, then there is a functor $R : \mathcal{C} \to \mathcal{M}$ assigning to every object of $\mathcal{C}$ its reflection in $\mathcal{M}$. The class $\{r_A : A \to RA\}$ forms a natural transformation $1_\mathcal{C} \to IR$, where $I : \mathcal{M} \to \mathcal{C}$ is the embedding.*

Proof. If $f \in \mathcal{C}(A, B)$ then there is a unique morphism $g \in \mathcal{M}(RA, RB)$ such that the diagram



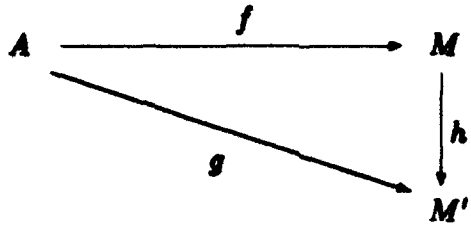
commutes. Putting $Rf = g$ and using the uniqueness in the definition of reflections, one gets that R is a functor. Indeed, if $f = 1_A$ then $g = 1_{RA}$ makes the above diagram commutative and there is no other possibility, so $R(1_A) = 1_{RA}$. If, in addition to the above $f \in C(A, B)$, one has $h \in C(B, C)$ then $Rh \circ Rf \circ r_A = r_C \circ h \circ f = R(h \circ f) \circ r_A$; again by the uniqueness, $Rh \circ Rf = R(h \circ f)$.

The fact that $\{r_A\}$ forms the requested natural transformation follows directly from the above diagram. □

Careful readers will notice from the proof above that the functor R having the properties of Theorem 8 is determined uniquely up to natural isomorphism. That does not mean that there is a unique functor R from C into $\mathcal{M}$; there are usually many such functors, but only one is such that there is a natural transformation $\{r_A\} : 1_C \to R$ having the property that r_{FA} is an isomorphism for each object A. Therefore, not every "enlargement" of a set in a Euclidean space can be a closing.

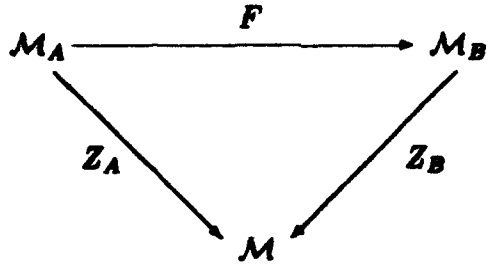
Let us return to the shape motivation of reflections. If the class of models is reflective in the whole class of all the objects, then we can assign to each object its closest model, that is, we can say what shape objects have, and also whether two objects have the same shape or not (i.e. whether they have isomorphic reflections or not). That is the ideal situation. Very often such a situation does not occur. If we return to our original example of closed convex plane sets as models and all plane sets as objects, and change the class of models to be all convex closed polyhedra in the plane, then only exceptionally does an object have a reflection. In that case, we cannot assign closest models to all objects. Can we decide whether two objects have the same shape? In our example the answer is very easy and it is in the affirmative. It suffices to realize that the closed convex hull of a set A is the intersection of all closed convex polyhedra containing A; so, two sets have the same shape if they have the same collection of closed convex polyhedra containing them.

We can use such an approach in general. We must find out how to express "the same collection of closed convex polyhedra containing them" in categorical language. From the above procedure we can deduce what corresponds to "collection of closed convex polyhedra containing them". For a given object A of a category $\mathcal{A}$ we denote by $\mathcal{M}_A$ the class $\bigcup\{\mathcal{A}(A, M) : M$ is an object of $\mathcal{M}\}$ (hence, it is the class of arrows starting at A and ending at objects of $\mathcal{M}$) We shall regard $\mathcal{M}_A$ as a category with the morphisms from an arrow f into an arrow g being a morphism h of $\mathcal{M}$ such that $h \circ f = g$, i.e., the following diagram is commutative:



The expression "the same" remains to be explained. In this situation, when we compare two categories $\mathcal{M}_A, \mathcal{M}_B$, the natural meaning is "isomorphic". But such an approach is too general. If we look at our example of planar sets and polyhedra, all bounded sets would have the same shape. Indeed, for bounded sets A, B, the families $\{K : K$ is a closed convex polyhedron containing $A\}$, $\{L : L$ is a closed convex polyhedron containing $B\}$ are isomorphic as partially ordered sets; of course, this isomorphism cannot preserve the polyhedra K, that is, in general it maps K into another polyhedron. That is the main difference to the case described above and so, to maintain what we meant by "the same", we must put some restriction onto the isomorphism, namely it must preserve the ranges of objects of our categories (i.e., the image of an $f : A \to M$ from $\mathcal{M}_A$ in $\mathcal{M}_B$ must be a morphism $g : B \to M$ with the same M). It is possible to express this in categorical language: denote by Z_A the functor $\mathcal{M}_A \to \mathcal{M}$ assigning the range M to the object $f : A \to M$ of $\mathcal{M}_A$, and $h : M \to P$ to the morphism h of $\mathcal{M}_A$ (from $f : A \to M$ to $g : A \to P$). It is a kind of a forgetful functor, it forgets the morphism $f : A \to M$ and remembers only the range M. The situation when no restrictions are required from an isofunctor $\mathcal{M}_A \to \mathcal{M}_B$ may be useful in some cases; one might then say that the objects have the same $\mathcal{M}$-shape in a weak sense.

Definition 9. Let $\mathcal{M}$ be a subcategory of a category $\mathcal{A}$. Two objects A, B have the *same $\mathcal{M}$-shape* provided there is an isofunctor $F : \mathcal{M}_A \to \mathcal{M}_B$ such that $Z_B \circ F = Z_A$, i.e., the following diagram is commutative:



We could proceed dually and take for $\mathcal{M}_A$ not the morphisms starting at A and ranging in $\mathcal{M}$ but those starting in $\mathcal{M}$ and ranging in A. One could denote such a property *to have the same $\mathcal{M}$-coshape*. Of course, it is also possible to require that two objects have the same $\mathcal{M}$-shape as well as the same $\mathcal{M}$-coshape. It is not difficult to modify our procedure to investigate also those concepts.

We show now, that the property "having the same shape" really generalizes the property "having isomorphic reflections".

Theorem 10. *Let $\mathcal{M}$ be a subcategory of a category $\mathcal{A}$ and let two objects A, B have isomorphic reflections in $\mathcal{M}$. Then they have the same $\mathcal{M}$-shape.*

Proof. Every $f : A \to M, M \in \mathrm{ob}\mathcal{M}$ can be written uniquely as $g \circ r_A$ for some $g \in \mathcal{M}(RA, M)$, where $r_A : A \to RA$ is the reflection of A in $\mathcal{M}$, and similarly for B. Let $i : RA \to RB$ be an isomorphism in $\mathcal{M}$. Then one can define, for $f \in \mathcal{M}_A$, $Ff = g \circ i^{-1} \circ r_B$. It is easy to show that F is an isofunctor $\mathcal{M}_A \to \mathcal{M}_B$ commuting with the forgetful functors Z. □

If A, B have the same $\mathcal{M}$-shape and A has a reflection in $\mathcal{M}$, then B has the same reflection in $\mathcal{M}$. If A, B would have the same $\mathcal{M}$-shape in the weak sense only, and if A has a reflection in $\mathcal{M}$ then B has also a reflection in $\mathcal{M}$ but one cannot say in general that their reflections are isomorphic; it may happen that there are no morphisms between these reflections.

In some cases it is easy to modify the situation above to show that A, B have the same shape iff they have isomorphic reflections in a modified category. Such a modification is artificial and makes sense theoretically, but not in practice. We shall not go into details here.

Several easy examples showing various possibilities of shapes will now be given. The main task of the examples is to be illustrative and that is why they are "mathematical" and not "practical". The second and the third examples show how the comparison "to have the same $\mathcal{M}$-shape" depends on relations between objects and models.

Examples. (1) Let $\mathcal{A}$ be a category where all the morphisms are isomorphisms (e.g., take the partially ordered set of planar sets as objects, and translations as morphisms). Let $\mathcal{M}$ be a full subcategory of $\mathcal{A}$ such that every object of $\mathcal{A}$ is isomorphic to an object of $\mathcal{M}$. Then A, B have the same $\mathcal{M}$-shape iff A, B are isomorphic. This example may be found in the paper by Roerdink in this volume [20].

(2) Take for $\mathcal{A}$ the bounded planar sets as objects, and for $\mathcal{M}$ the full subcategory of the planar closed convex polyhedra. We consider each of the following cases in which the morphisms of $\mathcal{A}$ are defined by:

(a) $\mathcal{A}(A, B)$ consists of inclusions. Then A, B have the same $\mathcal{M}$-shape iff the closed convex hulls of A, B coincide. This example was our illustrative example for the basic definition above.

(b) $\mathcal{A}(A, B)$ consists of restrictions of the translations (or the rotations, or the scalings) of the plane mapping A into B. Then A, B have the same $\mathcal{M}$-shape iff the closed convex hull of A is a translation (or a rotation, or a scaling) of the closed convex hull of B.

(c) $\mathcal{A}(A, B)$ consists of affine mappings (i.e., compositions of translations, rotations, and scalings, or of translations and linear maps) of the plane mapping A into B. Then A, B have the same $\mathcal{M}$-shape iff the closed convex hulls of A, B are affine isomorphic.

(d) $\mathcal{A}(A, B)$ consists of continuous maps (or uniformly continuous maps) of the plane mapping A into B. Then A, B have the same $\mathcal{M}$-shape iff they are homeomorphic (or uniformly homeomorphic).

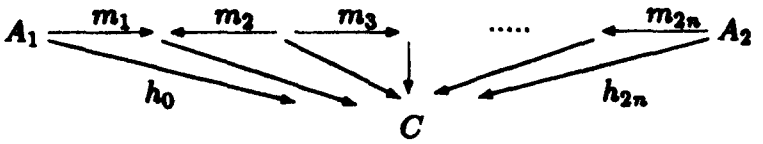
(e) $\mathcal{A}(A,B)$ consists of non-expansive maps (see the example 4 after the definition of subcategories) of the plane mapping A into B. Then A, B have the same $\mathcal{M}$-shape iff they are isometric.

(3) We adapt the previous example in the sense that fullness is no longer a prerequisite.
(a) $\mathcal{A}(A,B)$ consists of continuous maps of the plane mapping A into B; $\mathcal{M}(P,Q)$ consists of inclusions. Then A, B have the same $\mathcal{M}$-shape iff their closed convex hulls are homeomorphic.
(b) $\mathcal{A}(A,B)$ consists of continuous maps of the plane mapping A into B; $\mathcal{M}(P,Q)$ consists of identities. Then A, B always have the same $\mathcal{M}$-shape.
(c) $\mathcal{A}(A,B)$ consists of continuous maps of the plane mapping A into B; $\mathcal{M}(P,Q)$ consists of translations. Then A, B always have the same $\mathcal{M}$-shape.

Usually, the categories $\mathcal{M}_A$ are rather big. Instead of $\mathcal{M}_A$ one may take smaller subcategories $\mathcal{N}$, which are in some sense initial in the whole category. An object A of a category $\mathcal{C}$ is said to be initial, if $\mathcal{C}(A,C)$ consists of one morphism for every object C of $\mathcal{C}$. If one wants to transfer this definition to a bigger class than just a single object, one must require for every object C some morphism from an object of the initial subcategory, and also some kind of uniqueness. The following definition is taken from [3].

Definition 11. A subcategory $\mathcal{A}$ is said to be *initial* in $\mathcal{C}$ if
(i) for every object C of $\mathcal{C}$ there exists a morph $g \in \mathcal{C}(A,C)$ for some object A of $\mathcal{A}$;
(ii) for every two such morphisms $g_1 \in \mathcal{C}(A_1,C), g_2 \in \mathcal{C}(A_2,C)$, there exists a finite sequence $\{h_i \in \mathcal{C}(A_i,C)\}_{i\le 2n}$ with $h_0 = g_1, h_{2n} = g_2$ and morphisms $\{m_j : 1 \le j \le 2n\}$ such that $h_{2i+1} \circ m_{2i+1} = h_{2i} = h_{2i-1} \circ m_{2i}$.

The last condition of the Definition 11 means that the following diagram is commutative:



Of course, the best situation occurs if the initial subcategory is the simplest possible one, namely if it consists of a single morphism. The following result shows that it is just the case when the reflection exists:

Theorem 12. *An object A of $\mathcal{A}$ has a reflection in $\mathcal{M}$ iff $\mathcal{M}_A$ has an initial subcategory consisting of a single morphism (and a single object).*

Proof. If $r_A : A \to RA$ is a reflection of A in $\mathcal{M}$, then the subcategory of $\mathcal{M}_A$, consisting of the unique object r_A and its identity morphism is, clearly, initial. If, conversely, a subcategory of $\mathcal{M}_A$ consisting of a single object $f : A \to M$ and its identity morphism is initial, then to prove that $f : A \to M$ is a reflection of A in $\mathcal{M}$, one must show that for any $g : A \to P$ with P in $\mathcal{M}$ there is a morphism

$h : M \to P$ in $\mathcal{M}$ with $h \circ f = g$ (that follows from the first condition of initiality in Definition 11) and that such a morphism h is unique (that follows from the second condition of Definition 11). □

A very important example of the preceding theory is Borsuk's shape theory transferred into categorical language by Mardešić and Segal [14]. Details can be found in Segal's contribution to this volume [22]. It is briefly described in the next example.

Example. In the category $\mathcal{C}$ =CompHom of compact metric spaces and homotopy classes of continuous mappings (i.e., two continuous mappings are regarded the same if they are homotopic) we take the full subcategory $\mathcal{M}$ of polyhedra. Then one can show that for each compact metric space A there is an inverse sequence in $\mathcal{M}$ having A as its limit and generating an initial subcategory of $\mathcal{M}_A$. In other words, there is a sequence $\{A_n\}$ of objects of $\mathcal{M}$, a sequence of morphisms $\{p_n : A_{n+1} \to A_n\}$ in $\mathcal{M}$ and a sequence of morphisms $\{f_n : A \to A_n\}$ such that $f_{n+1} \circ p_n = f_n$ for each n; moreover, the sequence $\{f_n\}$ is in some sense universal with respect to the sequence $\{p_n\}$ and their properties.

When dealing with non-metric spaces, then one must admit more complicated systems than just sequences (e.g., for compact Hausdorff spaces, one deals with an inverse system of finite polyhedra, i.e., with families $\{A_\alpha, p_{\alpha,\beta}\}_\Omega$, where Ω is a directed partially ordered set, A_α's are finite polyhedra, $p_{\alpha,\beta} : A_\beta \to A_\alpha$ for $\alpha \leq \beta$). See, e.g., [14, 16]. In this case, the "zigzag" sequence $\{m_i\}$ from Definition 11 consists of m_2, m_3 only (the other m_i s are identities).

There is no room here to explain in detail how to work with initial subcategories of $\mathcal{M}_A$. As the previous example shows, the initial subcategories for different A, B may have no objects of $\mathcal{M}$ in common, and so one cannot use Definition 9 without change. The possibility of using much smaller categories than $\mathcal{M}_A$ has an unpleasant consequence that one must use more complicated formulas for functors between the initial subcategories. They are described for a special case in the paper by Segal in this volume [22]. If $\mathcal{N}_A, \mathcal{N}_B$ are initial in $\mathcal{M}_A, \mathcal{M}_B$, resp., the admissible functors $\mathcal{N}_A \to \mathcal{N}_B$ consist of a functor $\Phi : \mathcal{N}_A \to \mathcal{N}_B$ and a mapping Ψ from $\mathcal{N}_A$ into morphisms of $\mathcal{M}$ such that for each $p : f \to g$ in $\mathcal{N}_A$ one has $p \circ \Psi(f) \circ \Phi(f) = \Psi(g) \circ \Phi(g)$. Such (and only such) functors can be extended uniquely to functors $\mathcal{M}_A \to \mathcal{M}_B$ commuting with the forgetful functors Z.

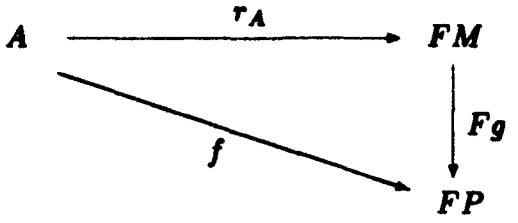
3 Shapes, Dilations, and Adjunctions

In the previous section, we investigated the situations when the class of models is a subclass of all the objects under consideration. It may happen that this is not the case. The models form a category $\mathcal{M}$, objects under consideration form another category $\mathcal{A}$, and the relation between objects and models is given by a functor $F : \mathcal{M} \to \mathcal{A}$. This means that models and/or relations between them cannot be described by means of properties of $\mathcal{A}$, and one must go outside $\mathcal{A}$

to describe them. A simple example is a classification of some physical experiments where some features are not taken into account explicitly; if one "forgets" temperature, there are various models with different temperature having "the same" other features considered.

The relation for a model M to be close to an object A now means to specify a morphism $A \to FM$, that is, we still work in the category $\mathcal{A}$ as before, but the relations between models are taken in $\mathcal{M}$, i.e., outside of $\mathcal{A}$. The model M which is closest to an object A is then described in the following way:

There exists a morphism $r_A : A \to FM$ such that for any other morphism $f : A \to FP$ with P being a model (an object of $\mathcal{M}$) there exists a morphism $g : M \to P$ in $\mathcal{M}$ such that $f = Fg \circ r_A$, i.e., the following diagram is commutative:



Again, as in the previous section, to assure uniqueness of the closest model, one must require g to be unique (see Definition 5).

Definition 13. Let $F : \mathcal{M} \to \mathcal{A}$ be a functor. An object M of $\mathcal{M}$ is said to be a *F-universal* for an object A of $\mathcal{A}$ if there exists a morphism $r_A : A \to FM$ such that for each morphism $f : A \to FP$ with $P \in \operatorname{ob}\mathcal{M}$ there exists a unique morphism $g : M \to P$ in $\mathcal{M}$ with $f = Fg \circ r_A$.

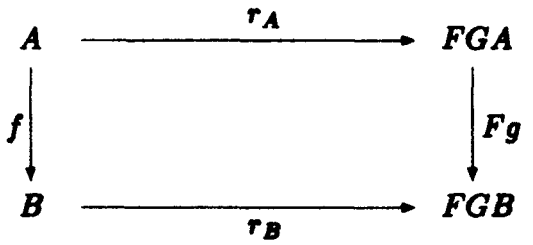
If every object of $\mathcal{A}$ has an F-universal object in $\mathcal{M}$, then $\mathcal{M}$ is said to be *F-universal* in $\mathcal{A}$ and F is called *adjoint* functor.

The dual concepts, obtained much as coreflection was from reflection in Definition 6, are *F-couniversal* and *coadjoint*.

The counterpart of Theorem 8 is as follows:

Theorem 14. *If $\mathcal{M}$ is F-universal in $\mathcal{A}$ then there exists a unique functor $G : \mathcal{A} \to \mathcal{M}$ such that every GA is the F-universal object for A.*

Proof. For every object A of $\mathcal{A}$, denote by GA its F-universal object in $\mathcal{M}$. For $f : A \to B$ in $\mathcal{A}$ there exists a unique morphism $g : GA \to GB$ in $\mathcal{M}$ such that the following diagram commutes:



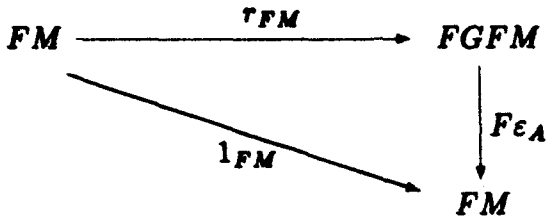
Define $Gf = g$; it is routine (an exercise) to show that G is a functor. □

Example. **In some sense, adjoint functors generalise Galois connections. If $F : M \to A, G : A \to M$ are order-preserving mappings between partially ordered sets A, M, then they form a Galois connection if, by definition, $Ga \leq m \Leftrightarrow a \leq Fm$. If $\mathcal{A}, \mathcal{M}$ are the categories determined by A, M, then the mappings F, G form a Galois connection iff F is an adjoint functor and G is the corresponding functor from Theorem 13.**

Remarks. **(1) When $F : \mathcal{M} \to \mathcal{A}$ is an inclusion, then F is adjoint iff $\mathcal{M}$ is reflective in $\mathcal{A}$.**

(2) If F is adjoint, then F preserves limits (the notion of limit is not defined in this paper). In some cases, the converse is true, e.g., in complete lattices: *a mapping F between complete lattices is adjoint iff F preserves infima.*

Using the notation of Theorem 14, the closest model from Definition 13 to an object A is now of the form $FG(A)$. The morphisms $r_A : A \to FG(A)$ form a natural transformation, $r : 1_{\mathcal{A}} \to F \circ G$, as the commutative diagram from the previous proof shows. It is interesting that there also exists a natural transformation $\varepsilon : G \circ F \to 1_{\mathcal{M}}$. Its existence follows from the F-universality of $r_{FM} : GM \to FGFM$ used for the identity morphism 1_{FM}; there exists a unique morphism $\varepsilon_M : GFM \to M$ making the following diagram commutative:



One can prove easily that ε is really a natural transformation and in addition to the property from the previous diagram, i.e.,

$$FM \xrightarrow{r_{FM}} FGFM \xrightarrow{F\varepsilon_M} FM = 1_{FM},$$

for every object M one has also

$$GA \xrightarrow{Gr_A} GFGA \xrightarrow{\varepsilon_{GA}} GA = 1_{GA}$$

for every object A. The situation described above (that is the existence of $G : \mathcal{A} \to \mathcal{M}, r : 1_{\mathcal{A}} \to GF, \varepsilon : FG \to 1_{\mathcal{M}}$, with the latter two properties) characterizes adjointness of the functor F, and is called an *adjoint situation*. In the case when F is an embedding of a full reflective subcategory, ε is the identity.

Another characterization of adjointness of the functor F requires also a functor $G : \mathcal{A} \to \mathcal{M}$ and a natural isomorphism $\mathcal{A}(A, FM) \to \mathcal{M}(GA, M)$ (compare the last example above).

In the previous example, it is stated that every Galois connection between partially ordered sets forms an adjoint functor. One can proceed to higher levels of Galois connections and get other examples of adjoint functors. Details can

be found in [8] by Herrlich and Hušek. But adjoint situations are too general to replace Galois connections everywhere; in the case that F is an adjoint functor $\mathcal{M} \to \mathcal{A}$ such that r_{FM} is an isomorphism for every object M, then F together with its corresponding functor G from Theorem 14 may be called a *generalized Galois connection* (it is called Galois connection of the fourth kind, or Galois adjunction in [8]). Such a situation preserves many basic properties of Galois connections; one will be stated in the next result.

Theorem 15. *Let F be an adjoint functor $\mathcal{M} \to \mathcal{A}$ and G be its corresponding functor $\mathcal{A} \to \mathcal{M}$. Then (F, G) is a generalized Galois connection iff the full subcategory $\mathcal{B}$ of $\mathcal{A}$ generated by all the images FGA is reflective in $\mathcal{A}$, the full subcategory $\mathcal{C}$ of $\mathcal{M}$ generated by all the images GFM is coreflective in $\mathcal{M}$, and the restrictions of F, G to $\mathcal{C}, \mathcal{B}$ respectively, are isofunctors.*

Thus, each such adjoint situation can be decomposed as a coreflection and a reflection. In the case that an adjoint functor does not form a generalized Galois connection, that is, some r_{FM} is not an isomorphism, then no combinations of F, G, r applied to this M can be an isomorphism (they are retractions or coretractions). In the case of small categories (e.g., those determined by partially ordered sets) such a situation cannot occur.

In the next example we come back to erosions, dilations, closings, and openings defined in the previous section in partially ordered sets of all subsets of some structures, and will apply our preceeding theorem to them.

Example. The assignment of the erosion to a set, regarded as a functor, gives rise to an adjoint functor F; the corresponding functor G assigns to a set its dilation. Of course, the corresponding natural transformations r, ε are formed of inclusions in this situation. It follows from Theorem 15 that the composition FG generates a reflection and the composition GF generates a coreflection. In our case they are just the closings and openings. The lattice of closings is isomorphic to the lattice of openings.

Now, we may repeat the procedure from the previous section for the case that the F-universal objects do not exist. Again we can form special categories of "approximations" of a given object A in order to be able to speak about objects having the same shape. For the present situation, for every object A of $\mathcal{A}$, we define the category F_A as follows:

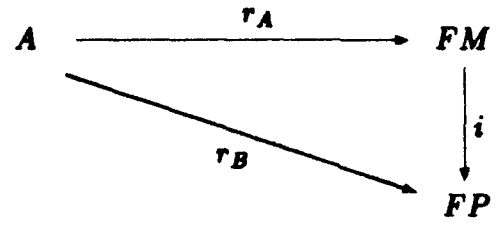
the objects of F_A are morphisms $f : A \to FM$ for objects M of $\mathcal{M}$;

the morphisms of $F_A(f, g)$, where $f : A \to FM, g : A \to FP$, are morphisms $m : M \to P$ of $\mathcal{M}$ with $Fm \circ f = g$.

Again we define the forgetful functor $Z_A : F_A \to \mathcal{M}$ by $Z_A(f : A \to M) = M, Z_A(h) = h$. The fact that $H : F_A \to F_B$ commutes with the forgetful functors (i.e., $Z_B \circ H = Z_A$) means that $H(f : A \to FM) \in \mathcal{A}(B, FM)$.

Definition 16. Let $F : \mathcal{M} \to \mathcal{A}$ be a functor. Two objects A, B of $\mathcal{A}$ have the same F-shape provided the categories F_A, F_B are isomorphic by a functor commuting with the forgetful functors Z.

The category $\mathcal{M}_A$ is now identical with the category I_A where I is the embedding functor $\mathcal{M} \to \mathcal{A}$. What follows is then quite analogous to that from the preceding section. We shall state modifications of the two Theorems 10, 12 without proofs. The first result says that the expression "A, B have the same F-shape" really generalizes the expression "A, B have isomorphic F-universal objects", where the F-universal objects $r_A : A \to FM, r_B : B \to FP$ of A, B, resp., are isomorphic if there is an isomorphism $i : M \to P$ in $\mathcal{M}$ such that the following diagram commutes:



Theorem 17. *Let $F : \mathcal{M} \to \mathcal{A}$ be a functor and let two objects A, B have isomorphic F-universal objects. Then they have the same F-shape.*

Theorem 18. *Let $F : \mathcal{M} \to \mathcal{A}$ be a functor. An object A of $\mathcal{A}$ has an F-universal object iff F_A has an initial subcategory consisting of a single morphism (and a single object).*

Also the following analogue of a result mentioned in the previous section holds: if A, B have the same F-shape and A has an F-universal object, then B has also an F-universal object coinciding with that of A.

4 Conclusions

It is not a task of category theory to prove deep results concerning shape theory. Category theory should help to find basic properties of various concepts and deduce from them general results valid for all those concepts. Using categorical theorems may help in finding at least basic results or good questions. In shape theory, one may ask what constructions preserve the shape; among the constructions one can include products, sums, various hulls, and transforms. For instance, the fact that some closings are invariant with respect to certain transformations, has a categorical background and follows from more general results. Knowing such general results may help in finding other invariants or showing that not only closings of sets, but also closings of functions in grey-level theory are invariant under certain transformations. To bring some example of another kind, in the recent paper [13] by Mardešić it is proved that a Tychonoff space X has the same shape (or strong shape) as its Čech-Stone compactification iff X is pseudocompact.

References

1. Adámek, J., Herrlich, H., Strecker, G. (1990). Abstract and Concrete Categories, John Wiley, New York.
2. Borsuk, K. (1975) Theory of Shape, PWN, Warsaw.
3. Cordier, J.M., Porter, T. (1988). Pattern recognition and categorical shape theory, Pattern Recognition Letters 7, pp. 73–76.
4. Eilenberg, S., Mac Lane, S. (1945). General theory of natural equivalences, Trans. Am. Math. Soc. 58, pp. 231–294.
5. Eilenberg, S., Steenrod, N. (1952). Foundations of Algebraic Topology, Princeton Univ. Press, Princeton.
6. Heijmans, H.J.A.M. (1993). Mathematical morphology as a tool for shape description, this volume, pp. 147–176.
7. Heijmans, H.J.A.M., Ronse, C. (1990). The algebraic basis of mathematical morphology. Part I: dilations and erosions, Comp. Vision, Graphics and Image Processing 50, pp. 245–295.
8. Herrlich, H., Hušek, M. (1990). Galois connections categorically, J. Pure and Appl. Algebra 68, pp. 165–180.
9. Herrlich, H., Hušek, M. (1993). Categorical topology. In: Hušek, M., van Mill, J., Recent Progress in General Topology, Elsevier, Amsterdam, in press.
10. Hušek, M. (1989). Categories and mathematical morphology, Lect. Notes in Comp. Sci. 393, Springer-Verlag, Berlin, pp. 294–301.
11. Mac Lane, S. (1950). Duality for groups, Bull. Am. Math. Soc. 56, pp. 485–516.
12. Mac Lane, S. (1971). Categories for the Working Mathematician, Springer-Verlag, Berlin.
13. Mardešić, S. (1992). Strong shape of the Čech-Stone compactification, Comment. Math. Univ. Carolinae 33, pp. 533–539.
14. Mardešić, S., Segal, J. (1982). Shape Theory, North-Holland, Amsterdam.
15. Matheron, G. (1975). Random Sets and Integral Geometry, John Wiley, New York.
16. Morita, K. (1975). On shapes of topological spaces, Fund. Math. 86, pp.251–259.
17. Pavel, P. (1983). "Shape theory" and pattern recognition, Pattern Recogn. 16, pp. 349–355.
18. Roerdink, J.B.T.M. (1992). Mathematical morphology with non-commutative symmetry groups. In: Dougherty, E.R. (ed.), Mathematical Morphology in Image Processing, Ch.7, pp. 205–204, Marcel Dekker, New York.
19. Roerdink, J.B.T.M. (1993). On the construction of translation and rotation invariant morphological operators. In: Haralick, R.M. (ed.), Mathematical Morphology: Theory and Hardware, Oxford Univ. Press, in press.
20. Roerdink, J.B.T.M. (1993). Manifold shape: from differential geometry to mathematical morphology, this volume, pp. 209–223.
21. Ronse, C., Heijmans, H.J.A.M. (1991). The algebraic basis of mathematical morphology. Part II: openings and closings, Comp. Vision, Graphics and Image Processing 54, pp. 74–97.
22. Segal, J. (1993). Shape theory: an ANR-sequence approach, this volume, pp. 111–125.
23. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.

Shape Theory: An ANR-Sequence Approach

Jack Segal

Department of Mathematics, University of Washington, Seattle, WA 98195, USA

Abstract. This is an expository article about the relationship of shape theory and some geometric notions. The ANR-sequence approach to shape theory is described. Then the shape classification of subcontinua of the plane is given. This is used to show that the dyadic solenoid is not the shape of any planar continuum. The shape classification of (m-sphere)-like continua is also obtained.

Keywords: shape, inverse sequence, inverse limit, ε-mapping, $\mathcal{P}$-like, ANR, pro-group, solenoid.

1 Introduction

Although shape theory applies to general topological spaces, in this expository article only compact metric spaces will be considered. Moreover, we will concentrate on a few examples and techniques in an attempt to convey some of the basic ideas and yet keep the paper reasonably self-contained. To reduce the abstractness of the subject only the relationship of shape theory and some geometric notions will be dealt with.

Shape theory is like homotopy theory in that it studies the global properties of topological spaces. However, the approach used in homotopy theory is of such a nature that it yields interesting results only for spaces which behave well locally (like ANRs or polyhedra). On the other hand, the tools of shape theory are so designed that they yield interesting results in the case of complicated local behavior (like that which occurs in metric compacta). Moreover, shape theory does not modify homotopy theory on ANRs, that is, it agrees with homotopy theory on such spaces.

It should be mentioned that one cannot ignore spaces with complicated local properties since they arise in nice settings, for example, they show up as fibres of maps between spaces with good local properties. In an attempt to overcome such difficulties, Borsuk [3] undertook the development of shape theory in 1968. One would expect shape theory to yield a classification of metric compacta, weaker than homotopy type but coinciding with it when applied to ANRs.

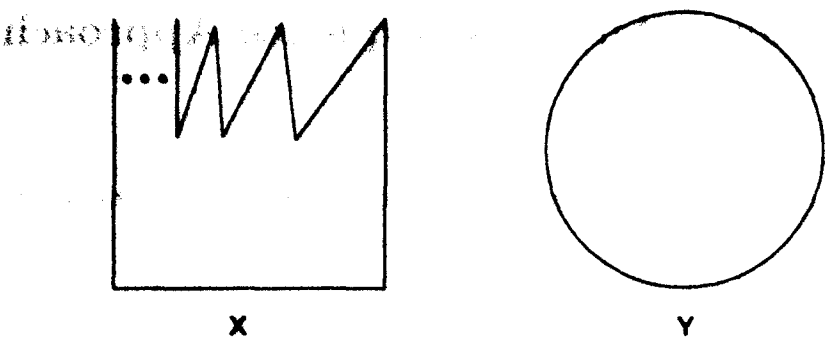


Fig. 1. The Warsaw circle and unit circle

Example 1. Let X denote the Warsaw circle $W \subset \mathbb{R}^2$ (see Fig. 1), and Y denote the unit circle $S^1 \subset \mathbb{R}^2$. Then there are maps $f : X \to Y$ which are essential (i.e., not homotopic to a constant map), but there is no essential map of Y into X since the image of Y in X would have to be a locally connected continuum and the only locally connected subcontinua of X are arcs or points. Since all maps $g : Y \to X$ are homotopic to a constant map, $fg \simeq 0$, and so they are not homotopic to the identity map on Y. Therefore, $Y \not\simeq X$ (i.e., they are of different homotopy type). Informally, the reason X and Y are not of the same homotopy type is that there are not enough maps of Y into X due to local difficulties.

Borsuk's idea to overcome this difficulty was to introduce a notion more general than a mapping. He was able to generalize mappings and yet maintain a great deal of the geometry inherent in the original notion. For example, one expects the Warsaw circle W to be in the same shape class as S^1 because of their global similarities (e.g., they both divide the plane into two components).

In 1970 Mardešić and Segal [12] (or [13]) gave a more categorical description of shape theory using ANR-systems. In this paper a brief description of this ANR-sequence approach to shape theory is given for compact metric spaces. In Sect. 2 a brief description of inverse sequences and their limits is given. In Sect. 3 ANRs and ANR-sequences are described. In Sect. 4 various shape invariants such as the Čech homology groups, homology pro-groups and movability are considered. In Sect. 5 the shape classification of planar continua are considered and used on some embedding problems. In Sect. 6 ε-mappings and $\mathcal{P}$-like compacta are considered and their connection with limits of $\mathcal{P}$-sequences is investigated. In Sect. 7 the shape classification of S^m-like compacta is derived. Finally, in Sect. 8 the shape classification of 0-dimensional compacta is obtained.

2 Inverse Sequences and Limits

The notion of the limit of an inverse sequence appeared, in a slightly different form, in a 1929 paper by Alexandroff [1]; the present definition was first stated by Lefschetz in [8]. Mappings of inverse sequences and induced limit mappings were first studied in Freudenthal's paper [6]. In full generality inverse systems

were defined by Lefschetz in [9]. An extensive treatment of inverse systems was presented by Eilenberg and Steenrod in their 1952 work [5].

An *inverse sequence* of compact metric spaces and bonding maps is a family

$$\boldsymbol{X} = \{X_n, p_{n,n'}\} \ ,$$

where

$$p_{n,n'} : X_{n'} \to X_n$$

is a map of $X_{n'}$ into X_n for each pair n, n' with $n \leq n'$. Moreover for $n \leq n' \leq n''$ the following conditions must hold:

$$p_{n,n'} p_{n',n''} = p_{n,n''} \ ,$$

and

$$p_{n,n} = \mathrm{id}_{X_n} \ ,$$

for all n.

Given an inverse sequence $\boldsymbol{X} = \{X_n, p_{n,n'}\}$ consider the Cartesian product $\prod X_n$; the element $\{x_n\}$ of this product is called a thread of the inverse sequence if for each pair n, n' of positive integers such that $n \leq n'$ we have

$$p_{n,n'}(x_{n'}) = x_n \ .$$

The subspace of the product space $\prod X_n$ consisting of all threads of the inverse sequence $\boldsymbol{X}$ is called the limit of the inverse sequence $\boldsymbol{X}$ and is denoted by $X = \lim \boldsymbol{X}$ *or* $X = \lim\{X_n, p_{n,n'}\}$.

The limit space X is a closed subspace of $\prod X_n$ and since the latter is a compact metric space so is the former. Moreover, if each X_n is a non-empty compact metric space, then the limit space X is non-empty. The projection of the product space $\prod X_n$ into X_n is denoted by the map $\varphi_n : \prod X_n \to X_n$ defined by $\varphi_n(\{x_n\}) = x_n$ for each n. The *natural projection* of the limit space X into X_n is the restriction of φ_n to X, i.e., $p_n = \varphi_n | X$. For $n \leq n'$ one has

$$p_n = p_{n,n'} p_{n'} \ .$$

Let $\boldsymbol{X} = \{X_n, p_{n,n'}\}$ and $\boldsymbol{Y} = \{Y_n, q_{n,n'}\}$ be inverse sequences. Then a *map* of inverse sequences

$$\boldsymbol{F} : \boldsymbol{X} \to \boldsymbol{Y}$$

consists of an increasing function $f : N \to N$ of the positive integers N and for each $n \in N$ a map

$$f_n : X_{f(n)} \to Y_n \ ,$$

such that if $n' \geq n$, then commutativity holds in the following diagram

$$\begin{array}{ccc} X_{f(n)} & \xleftarrow{p_{f(n),f(n')}} & X_{f(n')} \\ f_n \downarrow & = & \downarrow f_{n'} \\ Y_n & \xleftarrow[q_{n,n'}]{} & Y_{n'} \end{array}$$

In other words we have

$$f_n p_{f(n),f(n')} = q_{n,n'} f_{n'} \ .$$

Suppose $F : X \to Y$ is a map of the inverse sequence X into Y. Then the *inverse limit* of F is a map f of the limit of X into the limit of Y, $f : X \to Y$. The map f is defined as follows: if $x = (x_n) \in X$ then $f(x) = (f_n(x_{f(n)}))$. It follows from the commutativity condition that f takes X into Y. Moreover, we have commutativity holds in the diagram

$$\begin{array}{ccc} X_{f(n)} & \xleftarrow{p_n} & X \\ f_n \downarrow & = & \downarrow f \\ Y_n & \xleftarrow[q_n]{} & Y \end{array}$$

One of the most useful results on inverse limits is the following.

Theorem 1. *Any compact metric space X is the limit of an inverse sequence $X = \{X_n, p_{n,n+1}\}$ of compact polyhedra.*

3 ANR-sequences

Let A be a subset of a space X and $f : A \to Y$ a mapping of A into a space Y. A mapping $F : X \to Y$ of X into Y satisfying

$$F(x) = f(x) \text{ for } x \text{ in } A$$

is called an extension of f over X (with respect to Y).

Tietze's Extension Theorem. *Suppose C is a closed subset of a space X and f is a continuous real-valued function defined over C and bounded by a constant k:*

$$|f(x)| \leq k \ .$$

Then there exists a real-valued extension F of f over X such that

$$|F(x)| \leq k \ .$$

As a corollary to the Tietze Extension Theorem we have:

Corollary 2. *Let C be a closed subset of a space X and f a mapping of C into the n-sphere S^n. Then there is an open set in X containing C over which f can be extended (with respect to S^n).*

If a space has the property stated here for S^n then it is called an *absolute neighbourhood retract* or ANR. Every polyhedron is an ANR. More generally, the class of ANRs contains the class of polyhedra as a proper subset but the two are closely related since every compact metric ANR has the homotopy type of some compact polyhedron. However, while polyhedra are constructive in nature (belonging to elementary geometry), the ANRs are usually described axiomatically and are purely topological objects.

Consider ANR-sequences, that is, $\boldsymbol{X} = \{X_n, p_{n,n+1}\}$, where X_n is a compact ANR for compact metric spaces and $p_{n,n+1} : X_{n+1} \to X_n$ is a continuous map for all $n \in N$, the positive integers. These ANR-sequences will be organized into equivalence classes so that one can use any representative to denote the class. In this paper an ANR-sequence $\boldsymbol{X}$ is said to be associated with a space X if $X = \lim \boldsymbol{X}$, i.e., X is the inverse limit of $\boldsymbol{X}$. This allows one to use any such sequence associated with X. Either X is described this way to begin with as in the case of the solenoids or one obtains such an ANR-sequence associated with X through some construction. Such an ANR-sequence associated with X always exists. A map of ANR-sequences $\boldsymbol{f} : \boldsymbol{X} \to \boldsymbol{Y}$ consists of an increasing function $f : N \to N$ and a collection of maps $\{f_n\}, f_n : X_{f(n)} \to Y_n$ such that the following diagram

$$\begin{array}{ccc} X_{f(n)} & \xleftarrow{p} & X_{f(n+1)} \\ f_n \downarrow & \simeq & \downarrow f_{n+1} \\ Y_n & \xleftarrow[q]{} & Y_{n+1} \end{array} \tag{1}$$

commutes up to homotopy (where we delete subscripts from bonding maps in the diagram) i.e., $f_n p_{f(n),f(n+1)} \simeq q_{n,n+1} f_{n+1}$.

The identity map $\boldsymbol{1} : \boldsymbol{X} \to \boldsymbol{X}$ is given by $1(n) = n$, $1_n = \mathrm{id}$. The composition of maps of sequences $\boldsymbol{f} : \boldsymbol{X} \to \boldsymbol{Y}$, $\boldsymbol{g} : \boldsymbol{Y} \to \boldsymbol{Z} = \{Z_n, r_{n,n+1}\}$ is the map $\boldsymbol{h} = \boldsymbol{gf} : \boldsymbol{X} \to \boldsymbol{Z}$ defined by $h = fg : N \to N$ and for $h_n : X_{h(n)} \to Z_n$ we take $g_n f_{g(n)}$. So the increasing function $h = f(g) : N \to N$ and the collection $\{h_n\}$ form a map of ANR-sequences $\boldsymbol{h} : \boldsymbol{X} \to \boldsymbol{Z}$.

Now we define homotopy for maps of sequences. The maps $\boldsymbol{f}, \boldsymbol{g} : \boldsymbol{X} \to \boldsymbol{Y}$ are homotopic (written $\boldsymbol{f} \simeq \boldsymbol{g}$), if for every $n \in N$, there is an $n' \in N$ such that the following diagram commutes up to homotopy.

$$\begin{array}{ccccc} & & X_{n'} & & \\ & p \swarrow & & \searrow p & \\ X_{f(n)} & & \simeq & & X_{g(n)} \\ & f_n \searrow & & \swarrow g_n & \\ & & Y_n & & \end{array} \tag{2}$$

This homotopy relation for maps of ANR-sequences is an equivalence relation and classifies all maps of ANR-sequences associated with X to those associated with Y. These classes are called the *shape maps* from X to Y, written $\boldsymbol{f} : X \to Y$. A continuous map $f : X \to Y$ always determines a shape map $\boldsymbol{f} : X \to Y$. The

converse is not true in general. For example, consider a shape map $f : X \to Y$, that is,

$$\begin{array}{ccccccc} X_{f(1)} & \xleftarrow{p} & X_{f(2)} & \xleftarrow{p} & X_{f(3)} & \leftarrow \cdots & X \\ f_1 \downarrow & \simeq & \downarrow f_2 & \simeq & \downarrow f_3 & & \\ Y_1 & \xleftarrow[q]{} & Y_2 & \xleftarrow[q]{} & Y_3 & \leftarrow \cdots & Y \end{array} \tag{3}$$

and note that the squares in the diagram only commute up to homotopy (not exactly), so one does not expect to get a continuous map from X to Y but only a shape map from X to Y. We do have a special case though when Y is an ANR; then any shape map into Y is induced by a continuous map into Y.

In analogy with homotopy theory we define two spaces X and Y to be of the *same shape* ($\operatorname{Sh} X = \operatorname{Sh} Y$) if and only if there exist shape maps $f : X \to Y$ and $g : Y \to X$ such that (a) $fg \simeq 1_Y$ and (b) $gf \simeq 1_X$. In cases (a) and (b) hold f is called a shape equivalence. Further we say that X is shape dominated by Y ($\operatorname{Sh} X \leq \operatorname{Sh} Y$) provided (b) holds. In this case f is called a shape domination. Note that in the special case that X and Y are ANRs if $\operatorname{Sh} X = \operatorname{Sh} Y$ then $X \simeq Y$.

Example 2. In general X and Y may have the same shape but be of different homotopy type. Let X denote the Warsaw circle W, and Y denote the circle S^1. Then we consider the ANR-sequence $\boldsymbol{X} = \{X_n, p_{n,n+1}\}$ associated with X given by $X_n = S^1$ for each n and $p_{n,n+1} : X_{n+1} \to X_n$ is a properly chosen degree-one map. We also use an ANR-sequence $\boldsymbol{Y} = \{Y_n, q_{n,n+1}\}$ associated with Y given by $Y_n = S^1$ for each n and $q_{n,n+1}$ is the identity map on S^1. For the shape maps $f : \boldsymbol{X} \to \boldsymbol{Y}$, $g : \boldsymbol{Y} \to \boldsymbol{X}$ we take $f_n = g_n =$ identity on S^1. Then $fg \simeq 1_Y$ and $gf \simeq 1_X$. To obtain the necessary commutativity up to homotopy in the diagrams below we recall certain well-known facts about the degree of maps from S^1 to itself. These are $\deg(f(g)) = \deg g \cdot \deg f$, if $\deg f = \deg g$, then $f \simeq g$, and $\deg(\mathrm{id}) = 1$.

$$\begin{array}{ccccc} & & Y_{g(f(n))} & & \\ & \mathrm{id} \swarrow & & \searrow q & \\ Y_{g(f(n))} & & \simeq & & Y_n \\ & f_n g_{f(n)} \searrow & & \swarrow \mathrm{id} & \\ & & Y_n & & \end{array} \tag{4}$$

$$\begin{array}{ccccc} & & X_{f(g(n))} & & \\ & \mathrm{id} \swarrow & & \searrow p & \\ X_{f(g(n))} & & \simeq & & X_n \\ & g_n f_{g(n)} \searrow & & \swarrow \mathrm{id} & \\ & & X_n & & \end{array}$$

Thus we have $\operatorname{Sh} X = \operatorname{Sh} Y$ but by Example 1 we know X and Y are of different homotopy type.

Example 3. Here we describe two circle-like continua of different shape. Let X denote the dyadic solenoid, that is, X is the inverse limit of an inverse sequence of circles of unit radius in the complex plane and bonding maps given by $p_{n,n+1}(z) = z^2$. Let Y denote S^1 as described in Example 2. Then any shape map $g : Y \to X$ must have the property that $\deg g_n \neq 1$ (due to the commutativity up to homotopy in the diagram below). Then relation (2) of the definition of the homotopy of two maps of ANR-sequences cannot hold, since if it did we would have $f_n(g_n) \simeq 1_Y$ so $\deg f_n \cdot \deg g_n = 1$ which is impossible. Therefore we have that X and Y are of different shape.

$$\begin{array}{ccc} X_n & \xleftarrow{p} & X_{n+1} \\ g_n \uparrow & \simeq & \uparrow g_{n+1} \\ Y_{g(n)} & \xleftarrow[q]{} & Y_{g(n+1)} \end{array} \tag{5}$$

4 Shape Invariants and Pro-groups

Various functors of algebraic topology such as Čech homology or cohomology are shape invariants. Furthermore, if one does not pass to the limit in this situation, one obtains the homology pro-groups which are an even more delicate shape invariant.

For every ANR-sequence $\boldsymbol{X} = \{X_n, p_{n,n+1}\}$ one can define the homology pro-groups. These are inverse sequences of groups $H_m(\boldsymbol{X}) = \{H_m(X_n), p_{n,n+1*}\}$, which are objects in the category of pro-groups. Here we are taking the integers as the coefficient groups but one could use any abelian group. If $\boldsymbol{X}$ and $\boldsymbol{X}'$ are two ANR-sequences associated with X, then $H_m(\boldsymbol{X})$ and $H_m(\boldsymbol{X}')$ are naturally isomorphic pro-groups, that is, they are isomorphic objects of pro-Group. Therefore, one can define the homology pro-groups of a compact metric space X as the homology pro-groups of an associated ANR-sequence $\boldsymbol{X}$, since they are determined up to isomorphism in pro-Group. Clearly, isomorphic pro-groups have isomorphic inverse limits but the converse is not always true. For example, consider the pro-group $\boldsymbol{G} = \{G_n, \varphi_{n,n+1}\}$ where each G_n is a copy of the integers Z and each $\varphi_{n,n+1}$ is the homomorphism determined by multiplication by 2. Then $\boldsymbol{G}$ is not isomorphic to the zero pro-group $\{0\}$ although they both have as their inverse limit the zero group.

The inverse limit of the homology pro-group $H_m(\boldsymbol{X})$ is the usual Čech homology group $\check{H}_m(X)$. Homology pro-groups are finer invariants than the Čech homology groups. For example, $\check{H}_1$ of the dyadic solenoid is zero but the corresponding pro-group is nontrivial (as shown in the previous paragraph).

Borsuk [4] also introduced a far-reaching generalization of ANRs, called movability. The name of this shape invariant comes from a geometric interpretation of Borsuk's original definition. After its restatement in the ANR-sequence approach it became apparent that this is a categorical notion. So although the definition which follows is for spaces and maps it applies more generally, for example, to inverse sequences of groups and homomorphisms.

A compactum X is *movable* if there exists an ANR-sequence $\boldsymbol{X} = \{X_n, p_{n,n+1}\}$ associated with X such that for every positive integer n, there exists $n' \geq n$ such that for all $n'' \geq n$ there is a map $r : X_{n'} \to X_{n''}$ satisfying $p_{n,n''}r \simeq p_{n,n'}$.

Example 4. Recall from Example 3 the description of the dyadic solenoid X as the inverse limit of a sequence of circles with bonding maps of degree 2. Applying the H_1 functor to this sequence, we get pro-$H_1(X)$ as the following inverse sequence

$$Z \leftarrow Z \leftarrow Z \leftarrow \dots \ .$$

with each bonding homomorphism given by multiplication by 2. Then pro-$H_1(X)$ is not movable as a pro-group (and so X is not movable as a space). To see this suppose otherwise and take $n = 1$ in the definition. Then for $n'' = n' + 1$ we would have a homomorphism $r_* : H_1(S^1) \to H_1(S^1)$. So we would have $p_{n,n''*}r_* = p_{n,n'*}$. Thus $2^{n''-1}r_*(1) = 2^{n'-1}$ which implies that $r_*(1)$ is not an integer, a contradiction.

5 The Shape Classification of Planar Continua

Here the shape classification of planar continua is described.

Theorem 3. *Every planar continuum has the shape of some wedge of circles P_n, $0 \leq n \leq \infty$ (where P_0 is a point, P_n for n finite and ≥ 1 is the wedge of n circles, and P_∞ is the Hawaiian earring, i.e., an infinite wedge of circles).*

Proof. There exists a sequence of compact connected polyhedra $X_n \subseteq \mathbb{R}^2$ such that $X_{n+1} \subseteq \operatorname{Int} X_n$ and $\cap X_n = X$. Since a regular neighbourhood of X_n in $\mathbb{R}^2$ is a compact connected 2-manifold with boundary, one can assume from the beginning that each X_n is such a manifold. By the classification theorem for 2-manifolds, each X_n is a perforated disk, i.e., is of the form

$$X_n = D_n \setminus \bigcup_i \operatorname{Int} D_{ni} \ ,$$

where $\{D_{ni}\}$, $i = 1, \dots, r_n$, is a finite collection of disjoint disks contained in the interior of a disk $D_n \subseteq \mathbb{R}^2$. Observe that $\partial D_{n+1} \subseteq X_{n+1} \subseteq \operatorname{Int} X_n \subseteq \operatorname{Int} D_n$ so that $D_{n+1} \subseteq \operatorname{Int} D_n$. One can assume that no D_{ni} belongs to the unbounded component of $\mathbb{R}^2 \setminus X$ (otherwise, one modifies D_n so as to exclude D_{ni}). Similarly, one can assume that for $i \neq j$ the disks D_{ni}, D_{nj} belong to different components of $\mathbb{R}^2 \setminus X$. It is now clear that each disk D_{ni} is contained in D_{n+1} for otherwise it would belong to the unbounded component of $\mathbb{R}^2 \setminus X$. □

Notice that

$$\begin{aligned} D_{ni} \subseteq \bigcup_i D_{ni} &= \operatorname{Int} D_n \setminus \operatorname{Int} X_n \subseteq \operatorname{Int} D_n \setminus X_{n+1} \\ &= (\operatorname{Int} D_n \setminus D_{n+1}) \cup (D_{n+1} \setminus X_{n+1}) \\ &= (\operatorname{Int} D_n \setminus D_{n+1}) \cup (\bigcup_j \operatorname{Int} D_{n+1,j}) \ . \end{aligned}$$

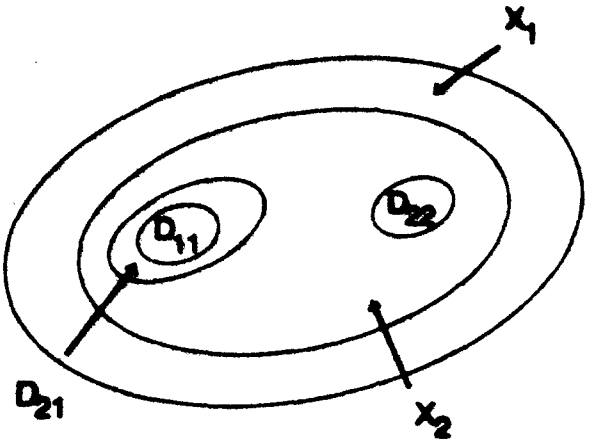


Fig. 2. Perforated disks

Since $D_{ni} \subseteq D_{n+1}$, we have $D_{ni} \subseteq \bigcup_j \operatorname{Int} D_{n+1,j}$. In fact each D_{ni} is contained in $\operatorname{Int} D_{n+1,j}$ for some $j = 1, \ldots, r_{n+1}$, because these are disjoint open sets. Moreover for a given j, one can have at most one i such that $D_{ni} \subseteq \operatorname{Int} D_{n+1,j}$ so that $D_{n+1,j} \setminus \operatorname{Int} D_{ni}$ is an annulus contained in X_n. This is because otherwise we would have two D_{ni} in the same component of $\mathbb{R}^2 \setminus X$. Consider the inclusion sequence $(\boldsymbol{X}, *) = ((X_n, *), i_{n,n+1})$.

It is not difficult to define an inverse sequence $(\boldsymbol{Y}, *) = ((Y_n, *), q_{n,n+1})$ and a shape map $(f_n) : (\boldsymbol{X}, *) \to (\boldsymbol{Y}, *)$ with the following properties. Each $(Y_n, *)$ is a finite wedge of circles, each $(Y_{n+1}, *)$ is of the form $(Y_n, *) \vee (Z_n, *)$, $q_{n,n+1}|Y_n = 1_{Y_n}$, $q_{n,n+1}|Z_n = *$, (see Fig. 3) and each f_n is a pointed homotopy equivalence. Consequently, (f_n) induces a shape equivalence $(X, *) \to (Y, *) = \lim(\boldsymbol{Y}, *)$. Observe that $(Y, *)$ is either a finite wedge of circles or the Hawaiian earring. We have thus obtained a complete shape classification of planar continua.

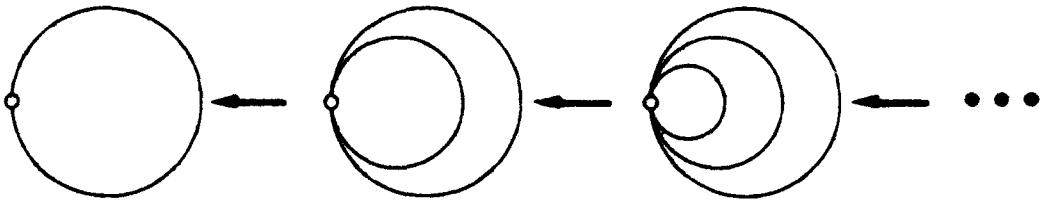


Fig. 3. Inverse sequence construction of the Hawaiian earring

Next consider an application of this classification to the question of embedding continua in the plane. In 1930 Kuratowski characterized 1-dimensional polyhedra which are embeddable in the 2-sphere S^2 as those which do not contain either of the two primitive skew curves K_1 or K_2. The polyhedron K_1 is the 1-skeleton of a tetrahedron with midpoints of a pair of non-adjacent edges joined

by a segment, and K_2 is the complete graph on five vertices. In 1966 Mardešić and Segal [11] showed that the connected polyhedra which are embeddable in S^2 are characterized as those which do not contain any one of the three polyhedra K_1, K_2 or T, where T is the "spiked disk" consisting of a disk and an arc which have only one point in common and this point is an interior point of the disk and an endpoint of the arc. They also conjectured that such a result was also true for ANRs. This was shown to be the case by Patkowska [14] in 1969.

But the question of embedding an arbitrary continuum in the plane is much more difficult. However, it is now shown how the above shape classification and shape invariants may be used to settle the question for the dyadic solenoid X which was described in Example 3. Suppose X embeds in $\mathbb{R}^2$ then by the above classification it would have to have the shape of some P_n. Then since the Čech homology functor is a shape invariant we would have

$$\check{H}_1(X) = \check{H}_1(P_n) \ ,$$

but

$$\check{H}_1(X) = 0 \text{ and } \check{H}_1(P_n) \neq 0 \text{ for } n \neq 0 \ ,$$

so the only possibility is $n = 0$. However, pro-$H_1(X)$ is also a shape invariant and we have

$$\text{pro-}H_1(X) \neq 0 \text{ and pro-}H_1(P_0) = 0 \ .$$

So the dyadic solenoid does not embed in the plane.

6 $\mathcal{P}$-like Compacta

If $\varepsilon > 0$, we say a mapping $f : X \to Y$ between compacta is an *ε-mapping* provided $\operatorname{diam}[f^{-1}(y)] < \varepsilon$ for each y in Y. Let $\mathcal{P}$ be a class of (compact) polyhedra. We say a compactum X is *$\mathcal{P}$-like* provided for each $\varepsilon > 0$ there is a polyhedron P in $\mathcal{P}$ and an ε-mapping $f : X \to P$ of X onto P. If X is $\mathcal{P}$-like and $\mathcal{P}$ consists of a single polyhedron P, then we say X is P-like. For example, if X is $\mathcal{P}$-like and $\mathcal{P} = \{S^1\}$, then X is called circle-like.

Example 5. Let $X = \operatorname{Cl}\left\{(x, y) \in \mathbb{R}^2 \mid y = \sin\left(\frac{1}{x}\right), 0 < x \leq 2\pi\right\}$, be the continuum in the plane which is the closure of the $\sin\left(\frac{1}{x}\right)$-curve. Then X is $\mathcal{P}$-like where $\mathcal{P} = \{I\}$, $I = [-1, 1]$ or in other words X is arc-like. To see this take for any $\varepsilon > 0$ an ε-mapping which is a retraction of X into a subarc X' which we call $g : X \to X' \subset X$. Note that X' is an arc, and therefore there is a homeomorphism $h : X' \to I$ of X' onto I. Then $f = hg : X \to I$ is the desired ε-map of X onto I, since $f^{-1}(y)$ is either a point or a small set of points determined by the retraction g.

The notion of an ε-mapping is a generalization of that of homeomorphism. However, ε-mappings form a much broader class than homeomorphisms. Notice in the above example one can use ε-mappings to squeeze out some local difficulties (non-local connectedness) in the domain. On the other hand, ε-mappings can do things like raise dimension and create homology. This is illustrated by the next example.

Example 6. Let $X = I = [-1, 1]$ and $\mathcal{P} = \{\mathcal{S}^2\}$. Then X is $\mathcal{P}$-like. Let $\varepsilon > 0$ and divide X into subarcs $I_k = [i_{k-1}, i_k]$ of diameter $\operatorname{diam}(I_k) < \frac{\varepsilon}{2}$ so that

$$X = \bigcup_{k=1}^{n} I_k \quad \text{and} \quad I_j \cap I_k = \begin{cases} \emptyset, & j \neq k-1, k, k+1 \\ \{i_{k-1}\}, & j = k-1 \\ \{i_k\}, & j = k+1 \end{cases} .$$

Now decompose $\mathcal{S}^2$ the unit sphere in $\mathbb{R}^3$ as follows: $\mathcal{S}^2 = \bigcup_{k=1}^{n} S_k$, where S_1 and S_n are disks bounded by simple closed curves C_1 and C_{n-1} around the north and south poles respectively, and S_k, $1 < k < n$, is an annulus bounded by simple closed curves C_{k-1}, C_k so that

$$S_j \cap S_k = \begin{cases} \emptyset, & j \neq k, k, k+1 \\ \{C_{k-1}\}, & j = k-1 \\ \{C_k\}, & j = k+1 \end{cases} .$$

Since all the S_k are locally connected continua by the Hahn-Mazurkiewicz Theorem one can map each I_k onto S_k with a mapping $f_k : I_k \to S_k$ so that $f_k(i_k) = f_{k+1}(i_k)$. Then $f = \bigcup_{k=1}^{n} f_k : X \to \mathcal{S}^2$ is an ε-mapping of X onto $\mathcal{S}^2$ since

$$f^{-1}(y) \subset I_k \cup I_{k+1} (\text{ or } I_{k-1} \cup I_k), \qquad y \in S_k$$

and so

$$\operatorname{diam}(f^{-1}(y)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon .$$

Example 6 can be generalized to show that if C is any locally connected continuum which contains an open n-dimensional disk $n \geq 2$ as an open subset, then I can be ε-mapped onto C for all $\varepsilon > 0$. In particular, I is $\mathcal{S}^n$-like for $n \geq 2$.

On the other hand, I cannot be ε-mapped onto $\mathcal{S}^1$ for all $\varepsilon > 0$. This can be made more precise by the following result of Kuratowski [7]. If $f : I^n \to \mathcal{S}^n$ is an ε-mapping of the n-cell I^n onto the n-sphere $\mathcal{S}^n$, then there is a positive real number $l_n = \frac{2n+2-\sqrt{2n^2+2n}}{n+2}$ such that $\varepsilon \geq l_n$. ($l_1 = \frac{2}{3}, l_2 = \frac{3-\sqrt{3}}{2} = 0.63\ldots, l_\infty = 2 - \sqrt{2} = 0.586\ldots$) .

The following theorem summarizes the classical results on $\mathcal{P}$-like compacta.

Theorem 4. *If* $\mathcal{P} = \begin{cases} (1)\ \textit{all polyhedra} \\ (2)\ \textit{all connected polyhedra} \\ (3)\ \textit{all polyhedra of } \dim \leq n \end{cases}$, *then* $\mathcal{P}$*-like compacta consist of* $\begin{cases} (1)\ \textit{all compacta} \\ (2)\ \textit{all continua} \\ (3)\ \textit{all compacta of } \dim \leq n \end{cases}$.

(1) is proved by using geometric realizations of the nerves of open coverings and applying canonical mappings into nerves. (3) is Alexandroff's theorem characterizing dimension by polyhedral approximation.

If $\boldsymbol{X} = \{X_n, p_{n,n'}\}$ is an inverse sequence of polyhedra from a class $\mathcal{P}$ and if all the bonding maps are surjective, then $\boldsymbol{X}$ is called a $\mathcal{P}$-sequence. It follows that the inverse limit $X = \lim \boldsymbol{X}$ of a $\mathcal{P}$-sequence $\boldsymbol{X}$ is a $\mathcal{P}$-like compactum.

In 1963 Mardešić and Segal in [10] established the converse of this last statement in case X is a compactum and $\mathcal{P}$ is a class of connected polyhedra. In fact, such an X is the limit of an inverse sequence $\boldsymbol{X} = \{X_n, p_{n,n'}\}$ where all the bonding maps are surjective (i.e., $\boldsymbol{X}$ is a $\mathcal{P}$-sequence).

Theorem 5. *Let $\mathcal{P}$ be a class of connected polyhedra. Then the class of $\mathcal{P}$-like compacta coincides with the class of limits of $\mathcal{P}$-sequences.*

7 The Shape Classification of $\mathcal{S}^m$-like Compacta

The above results on $\mathcal{P}$-like compacta can be used to obtain shape classifications of $\mathcal{P}$-like compacta for various classes $\mathcal{P}$. As an example we will examine $\mathcal{S}^m$-like compacta.

We first consider the 1-dimensional case. Let S_P denote the P-adic solenoid where $P = (p_1, p_2, \ldots)$ is a sequence of primes and $S_P = \lim\{X_n, p_{n,n'}\}$ where $X_n = \{z \,|\, |z| = 1\}$ is the unit circle in the complex plane and the map $p_{n,n+1} : X_{n+1} \to X_n$ is given by $p_{n,n+1}(z) = z^{p_n}$ (a map of degree p_n). Two sequences of primes $P = (p_1, p_2, \ldots)$ and $Q = (q_1, q_2, \ldots)$ are said to be equivalent ($P \sim Q$) provided it is possible to delete a finite number of terms from each so that every prime occurs the same number of times in each of the deleted sequences.

Theorem 6. *Let S_P and S_Q be two solenoids. Then the following statements are equivalent:*

1. $\mathrm{Sh}\, S_P = \mathrm{Sh}\, S_Q$,
2. $P \sim Q$,
3. *S_P and S_Q are homeomorphic.*

Proof. (i) $\Rightarrow$ (ii) Since Čech cohomology is a shape invariant we have $\check{H}^1(S_P) \approx \check{H}^1(S_Q)$. From the continuity of Čech cohomology it follows that $\check{H}^1(S_P) \approx F_P$ the group of P-adic rationals of the form $\frac{m}{p_1 p_2 \cdots p_s}$, where $m \in \boldsymbol{Z}$. However, $F_P \approx F_Q$ implies $P \sim Q$.

(ii) $\Rightarrow$ (iii) As noted by Bing [2] the two solenoids S_P and S_Q are homeomorphic if $P \sim Q$.

(iii) $\Rightarrow$ (i) This is obvious since shape is a topological invariant.

For every sequence of primes $P = (p_1, p_2, \ldots)$ we now consider the inverse sequence $S_P^m = \{X_n, p_{n,n+1}\}$ where each X_n is the m-sphere S^m and $p_{n,n+1} : X_{n+1} \to X_n$ is a map of degree p_n. We denote the inverse limit $\lim S_P^m$ by S_P^m. The shape of S_P^m is completely determined since any two bonding maps of the same degree are homotopic and so the limit spaces will have the same shape. □

Theorem 7. *Two spaces S_P^m and S_Q^m are of the same shape if and only if $P \sim Q$.*

Proof. Consider the solenoid S_P and its ANR-sequence $\boldsymbol{S}_P$. Applying the $(m-1)$-fold suspension Σ^{m-1} we obtain the unique sequence

$$\Sigma^{m-1}(\boldsymbol{S}_P) = \{\Sigma^{m-1}(S^1), \Sigma^{m-1}(p_{n,n+1})\}$$

whose limit is $\Sigma^{m-1}(S_P)$. Let $f_n : X_n = S^m \to \Sigma^{m-1}(S^1) = S^m$ be any mapping of degree 1. Then the maps $\{f_n\}$ form a map of ANR-sequences $\boldsymbol{f} : \boldsymbol{S}_P^m \to \Sigma^{m-1}(\boldsymbol{S}_P)$ because of the Hopf Classification Theorem for maps of spheres. In fact, $\boldsymbol{f}$ is a homotopy equivalence and thus S_P^m is of the same shape as $\Sigma^{m-1}(S_P)$. In this way the problem reduces to show that $\Sigma^{m-1}(S_P)$ and $\Sigma^{m-1}(S_Q)$ are of the same shape if and only if $P \sim Q$.

If $\Sigma^{m-1}(S_P)$ and $\Sigma^{m-1}(S_Q)$ are of the same shape, then

$$\begin{array}{ccc} \check{H}^m(\Sigma^{m-1}(S_P)) & \xrightarrow{\approx} & \check{H}^m(\Sigma^{m-1}(S_Q)) \\ \approx\downarrow & & \downarrow\approx \\ \check{H}^1(S_P) & & \check{H}^1(S_Q) \end{array}$$

which implies $P \sim Q$. Conversely, if $P \sim Q$, then S_P and S_Q are homeomorphic and therefore so are $\Sigma^{m-1}(S_P)$ and $\Sigma^{m-1}(S_Q)$. □

Theorem 8. *Every S^m-like continuum X has the shape of a point, S^m or S_P^m.*

Proof. Since X is S^m-like it admits inverse sequence expansion $\boldsymbol{X} = \{X_n, p_{n,n+1}\}$ where each X_n is an m-sphere. Let $k_n = \deg(p_{n,n+1})$. We can assume that all $k_n \geq 0$ (this can be achieved by omitting a finite number of initial terms and by taking compositions of consecutive bonding maps with an even number of negative degrees). If there are infinitely many zeros among the degrees, then X is the shape of a point because the maps of degree 0 can be replaced by constant maps without affecting the shape. Thus, if X is not the shape of a point, we can assume $k_n \geq 1$. If there is an n_0 such that $k_n = 1$ for $n \geq n_0$, then by replacing the bonding maps $p_{n,n+1}$ by identity maps, we conclude that X is of the shape of S^m. Otherwise, we can assume that all $k_n \geq 2$ (this can be achieved by taking suitable compositions of consecutive bonding maps). We now decompose the bonding map $p_{n,n+1}$ into a product of maps from S^m into S^m each of prime degree. This yields a limit space S_P^m of the same shape as X and we are done. □

These last two theorems classify all S^m-like continua with respect to shape.

8 Shapes of 0-dimensional Compacta

Theorem 9. *Two 0-dimensional compact metric spaces are of the same shape if and only if they are homeomorphic.*

Proof. First notice that every 0-dimensional compact metric space X is the inverse limit of an ANR-sequence $\boldsymbol{X} = \{X_n, p_{n,n'}\}$, where all X_n are finite sets. Indeed, $\boldsymbol{X}$ can be obtained by considering the nerves X_n of finite coverings $\mathcal{U}$ of X formed by disjoint open sets and the projections $p_{n,n'}$ uniquely determined by inclusion.

Now assume that X and Y are compact metric spaces of the same shape and that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are associated ANR-sequences with X_n and Y_n finite sets. Then there exist maps of sequences $\boldsymbol{f} : \boldsymbol{X} \to \boldsymbol{Y}$ and $\boldsymbol{g} : \boldsymbol{Y} \to \boldsymbol{X}$ such that $\boldsymbol{gf} \simeq \boldsymbol{1}$ and $\boldsymbol{fg} \simeq \boldsymbol{1}$. Since the components of Y_n are single points, the homotopy

$$f_n p_{f(n),f(n')} \simeq q_{n,n'} f_{n'}, \quad n \leq n' \ ,$$

becomes an equality

$$f_n p_{f(n),f(n')} = q_{n,n'} f_{n'} \ .$$

Therefore, $\{f_n\}$ is actually a map of inverse sequences and so induces a map $f : X \to Y$ such that for every n

$$f_n p_{f(n)} = q_n f \ . \tag{6}$$

Similarly, we have a map $g : Y \to X$ such that for every n

$$g_n q_{g(n)} = p_n g \ . \tag{7}$$

Furthermore, the homotopy $gf \simeq 1$ implies that for every n there is an $n' \geq n, fg(n)$ such that

$$g_n f_{g(n)} p_{fg(n),n'} = p_{n,n'} \ . \tag{8}$$

Consequently, for every n,

$$g_n f_{g(n)} p_{fg(n)} = p_n \ . \tag{9}$$

By (6), (9) becomes

$$g_n q_{g(n)} f = p_n \ , \tag{10}$$

which, by (7), gives

$$p_n g f = p_n \ . \tag{11}$$

Since (11) holds for every n, we conclude that

$$gf = 1_X \ . \tag{12}$$

Similarly, we obtain

$$fg = 1_Y \ . \tag{13}$$

(12) and (13) show that $f : X \to Y$ is a homeomorphism which completes the proof. □

References

1. Alexandroff, P. (1929). Untersuchen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension, Ann. of Math. 30, pp. 101–187.
2. Bing, R. H. (1960). A simple closed curve is the only homogeneous bounded plane continuum that contains an arc, Canad. J. Math 12, pp. 209–230.
3. Borsuk, K. (1968). Concerning homotopy properties of compacta, Fund. Math. 62, pp. 223–254.
4. Borsuk, K. (1969/70). On movable compacta, Fund. Math. 66, pp. 137–146.
5. Eilenberg, S., Steenrod, N. (1952). Foundations of Algebraic Topology, Princeton University Press, Princeton.
6. Freudenthal, H. (1937). Entwicklungen von Räumen und ihren Gruppen, Compositio Math. 4, pp. 145–234.
7. Kuratowski, K., (1933). Sur les transformations des sphères en des surfaces sphèriques, Fund. Math. 20, pp. 206–213.
8. Lefschetz, S. (1931). On compact spaces, Ann. of Math. 32, pp. 521–538.
9. Lefschetz, S. (1942). Algebraic Topology. Am. Math. Soc., New York.
10. Mardešić, S., Segal, J. (1963). ε-mappings onto polyhedra, Trans. Am. Math. Soc. 109, pp. 146–164.
11. Mardešić, S., Segal, J. (1966). A note on polyhedra embeddable in the plane, Duke Math. J. 33, pp. 633–638.
12. Mardešić, S., Segal, J. (1971). Shapes of compacta and ANR-systems, Fund. Math. 72, pp. 41–59.
13. Mardešić, S., Segal, J. (1982). Shape Theory, North-Holland, Amsterdam.
14. Patkowska, H. (1969). Some theorems on the embeddability of ANR-spaces into Euclidean spaces, Fund. Math. 65, pp. 289–308.

Can Categorical Shape Theory Handle Grey-level Images?

*Timothy Porter**

School of Mathematics, University of Wales at Bangor, Dean Street, Bangor, Gwynedd, LL57 1UT, United Kingdom

Abstract. Categorical shape theory can be considered as a formal model of a recognition process, but can it handle grey-level images? In this paper, some of the available pure mathematics, mostly from topology and category theory, that may be useful in this context are considered and the feasibility of such a model is discussed both from the machine-implementation viewpoint and a biological one.

Keywords: shape, categorical shape theory, sheaf theory, hierarchical systems, neural networks, formal languages.

1 Introduction

Categorical shape theory is seen as providing a formal language for describing certain aspects of the pattern recognition process, and examining the theoretical limitations of pattern recognition, limitations that are there even if one assumes a theoretical possibility of potentially infinite processes. It emphasizes the role of archetypes or models in the comparison process that leads to recognition. It grew out of geometric shape theory which uses approximating systems of polyhedral spaces to obtain information on compact metric spaces (for instance (closed) subsets of a Euclidean space such as one imagines the real world to be). A straightforward extension of the ideas of geometric shape theory thus provides a geometric example of a categorical shape theory, although of course the historical development of shape theory was in the opposite direction. How then can one use insights from shape theory to handle grey-level images where one does not only have a space? What should be the mathematical models of the objects and of the archetypes? What mathematical machinery might provide a possible language for this? This paper reviews some mathematics that may be potentially useful for this problem and describes the first steps in an attempt to solve it.

* The author would like to thank Gavin Wraith for providing an explanation of Lawvere's ideas on metric spaces and Andrée Charles Ehresmann for discussions of her ideas on the modelling of brain functions.

2 Why Category Theory?

This mathematical context is richer than many of the usual methods of mathematics used in pattern recognition as it emphasises, at the same time, the algebraic, topological, *and* combinatorial aspects of the subject. In fact Pavel [44] has already argued for its use as a unifying language for various aspects of pattern recognition. Furthermore categorical methods are also increasingly being applied to problems in theoretical computer science (see the excellent discussion in Goguen [19]) and in the description of evolutive hierarchical systems (the work of Ehresmann and Vanbremeersch [see reference list]) in theoretical biology. In both of these latter situations, the applications use the power of categorical language to discuss the way in which the whole of a system is greater than the sum of its parts. This is evident, for instance, in the study of the denotational semantics of modular programming languages in Moggi [42], using indexed categories or in the description of a hierarchical system in Ehresmann and Vanbremeersch [again see reference list], where categorical (co)limits are the structure used at a very fundamental level. (As noticed by Ehresmann and Vanbremeersch [11], the processing of signals in the visual cortex would seem to correspond to such a colimiting process in a hierarchical system and this suggests that similar use of colimits may be of help here in describing mathematical models for images and archetypes.)

The structure suggested here to handle grey-levels or colour is based both on category theory and on *sheaf theory*. Sheaf theory is designed to handle the passage from the local to the global, to act as an integrator even when any integration in the usual analytical sense would seem inappropriate. It is thus well suited to describing the combination of local and global information needed for an adequate description of a visual object.

Sheaf theoretic models for the objects, models, and comparison maps will be described. This raises many questions as to the adequacy of such models. It also raises questions of the mechanism by which a machine might approximate such a complicated mathematical object as a space together with a grey-level. To examine this, a brief summary of some recent relevant results from the theory of neural networks is included.

The paper will attempt to interpret these pure mathematical concepts in such a way that their utility, or otherwise, for the theoretical problem under consideration may be better evaluated.

3 Categorical Background

A brief introduction to category theory including some of the definitions needed to describe categorical shape theory is given in the introductory article to this section of this volume [27]. The notions that will be assumed in addition to those of categories, functors, and natural transformations, include limits and colimits. More technical definitions will be given below. A general reference for category theory is Mac Lane [34], and a good introduction to its basic ideas and to how they are applied in theoretical computer science is to be found in Goguen [19].

The use of sets, or more usually of structured sets of some kind, is now commonplace in modelling situations. Such a theory is adequate as long as there is only one set of things being considered. When more than one such set is needed then structure-preserving functions or morphisms between the structured sets are usually considered. With the minimum of extra conditions (associativity of composition and existence of identity morphisms), the structured sets and the morphisms between them form a category. The motto is that *structure is only observable via comparison;* in other words, *if a structured object is not interacting with others, if there are no morphisms, then the structure is essentially a closed system and little can be said about it.* Hence to understand objects, one must also understand the morphisms between them.

In many situations the structure imposed on the sets includes that of an order $\leq$. Such an order structure, by itself, determines a category (see Mac Lane [34]). Another structure common in applications is that of a graph, or network, consisting of some vertices or nodes and some edges, which for simplicity we will assume are directed, that is each has a start vertex and an end vertex. Such a directed graph is often studied by examining paths in it. The paths in a graph G again form a category, which is sometimes denoted $Pa(G)$.

If $(X, \leq)$ is an ordered set, then one can consider it to be a category in which each element of X is thought of as being an object of the category and if x and y are elements of X, then there is a single arrow from x to y exactly if $x \leq y$. Composition is expressed precisely by transitivity of the order relation:

$$x \leq y \text{ and } y \leq z \text{ together imply that } x \leq z \ .$$

It is important to notice that in such a category the objects of the category are not usually sets, and the morphisms or arrows between them are not usually functions. The same comments apply to the category $Pa(G)$ of paths in a graph G. This category has the vertices of the graph as the objects of the category and the paths from a to b as the arrows from a to b. Composition is by concatenation of the sequences making up the paths, provided that the end of the first path is the start of the second. The identity path at a vertex a is the empty sequence of edges that start and end at a, so again the objects are not sets and the morphisms are not functions.

The point just made is worth repeating as, if misunderstood, it can lead to difficulties. The objects of a category are merely objects, the morphisms merely arrows. The only *structure* that an object has is by virtue of its interaction with other objects. It is interesting to note that a similar point is made in a recent article on theoretical computer science by Martí-Oliet and Meseguer [37]. Their examples are also useful for one of the general points of this article, namely how one may think of categorical shape theory as a formal language. In logical type theory, one thinks of formulae as giving rise to types and proofs to functions and this is formalized in the Curry-Howard correspondence by saying that concepts pair up

$$\textit{Formulae} \longleftrightarrow \textit{Types} \ ,$$

and

$$\textit{Proofs} \longleftrightarrow \textit{Functions} \ .$$

Lambek [28] and Lawvere [30] in categorical logic paired up

$$Formulae \longleftrightarrow Objects \ ,$$

and

$$Proofs \longleftrightarrow Morphisms \ ,$$

and while this looks naively as if it is just a reinterpretation of Curry-Howard, in fact it is much wider as it does not assume that the objects are sets or types and neither does it assume that morphisms are tied to being functions. Objects are objects, morphisms are morphisms, that is all. This has been a key point in the development by Girard [17, 18] of linear logic, whose connections with natural deductions, Petri nets and concurrency may be of general relevance here. Continuing these correspondences, this theory pairs up

$$Formulae \longleftrightarrow States \ ,$$

and

$$Proofs \longleftrightarrow Transitions \ ,$$

and to complete the triangle, Meseguer and Montanari [39], working with Petri nets, develop the correspondence

$$States \longleftrightarrow Objects \ ,$$

and

$$Transitions \longleftrightarrow Morphisms \ .$$

Thus a modern interpretation of a formal language may be in terms of states and transitions, or objects and morphisms, instead of being merely in terms of formulae and proofs, and in our attempt to interpret categorical shape theory as a formal language, it will be wise to keep in mind these correspondences and the imagery they generate.

In any category $\mathbf{C}$, a *diagram* in $\mathbf{C}$ is an interacting system of objects and morphisms in $\mathbf{C}$. More precisely, the diagram's organizational structure is given by a (small) category $\mathbf{D}$, sometimes called the *diagram scheme*, and then the diagram itself is a functor $\mathbf{F} : \mathbf{D} \to \mathbf{C}$. A morphism of diagrams from $\mathbf{F} : \mathbf{D} \to \mathbf{C}$ to $\mathbf{G} : \mathbf{D} \to \mathbf{C}$ is a natural transformation between the functors. A morphism of diagrams is thus given by a compatible family $\{\mathbf{f}(d) : \mathbf{F}(d) \to \mathbf{G}(d) \mid d \ in \ \mathbf{D}\}$ of $\mathbf{C}$-morphisms between the corresponding nodes of the two diagrams.

If, as Goguen [20] suggests, "systems are diagrams", how can a single object observe the behaviour of a system? Any object X determines a constant diagram, k_X whose node objects $k_X(d)$ are just copies of X itself and whose interconnecting morphisms are all the identity on that object. A diagram morphism from k_X to F allows X to observe the system F. A *limit* for F is a single object $LimF$ (together with a diagram morphism from k_{LimF} to F) such that there is a natural bijection

$$Diagram\ morphisms(k_X, F) \cong \mathbf{C}(X, LimF) \ .$$

This can intuitively be thought of as saying that when an object observes the system F it can only see the information available in $LimF$; that is, to quote Goguen [20] again, "Behaviour is limit".

If dually a system F interacts with X via morphisms from F to X, then a *colimit for* F is an object $ColimF$ (together with a morphism to k_{ColimF}) such that there is a natural bijection

$$Diagram\ morphisms(F, k_X) \cong \mathbf{C}(ColimF, X) .$$

For the introduction to sheaf theory in the next section and to categorical shape theory in Sect. 5, some intuition about limits and colimits, beyond their definitions, will be needed.

Example. In an ordered set, the limit of a diagram consisting of two elements is just their greatest lower bound or meet, whilst their colimit is their join or least upper bound.

Of course, in a given setting, these may not always exist. (See Ehresmann and Vanbremeersch [8] for a discussion of how one can add colimits to a category in a useful way. The case when the category is that associated with a neural network leads to the concept of categorical neurons, that is collections of neurons interacting in a coherent and concerted fashion, see Ehresmann and Vanbremeersch [11].) The intuition of the colimit of a diagram is that it is obtained by gluing together the objects in the diagram. Colimits thus are a bit like unions of sets. The dual construction of limits then corresponds loosely to intersections and such an intuition will probably be sufficient. For those readers with a computer science background the articles of Goguen, [19, 20], provide examples that may be of use.

4 Sheaf Theory

If X is a topological space, one often needs to study continuous or upper semi-continuous functions on X with, say, real values. For instance X may be a subset of the plane with an intensity function defined on it. The properties of such functions are locally defined, that is they are often defined using the *open* subsets of X. For example, to require that a continuous real-valued function f is non-zero is equivalent to specifying the open set $\{x \mid f(x) \neq 0\}$.

In some sense a continuous real-valued function can be thought of as a continuously varying real number. The notion of a sheaf corresponds to an idea of a continuously varying family of sets. A preliminary notion is that of a presheaf.

Definition 1 Presheaves and Sheaves. A *presheaf* F on a space X assigns to each open set U in X, a set $F(U)$. If V is a smaller open set $V \subset U$ there is a function $restF(U, V)$ given from $F(U)$ to $F(V)$. (This function is usually called the *restriction map* as in practice it usually is one.) Furthermore if $W \subset V$ then the composite $restF(V, W) \circ restF(U, V)$ is required to be the same as

$\mathrm{rest}F(U,W)$. Alternatively let $\mathsf{Open}(X)$ be the lattice of open sets of X, then a presheaf F is a functor from the dual of $\mathsf{Open}(X)$ to the category Sets.

Given two elements f_1 in $F(U_1)$ and f_2 in $F(U_2)$ such that the restriction of these elements to $F(U_1 \cap U_2)$ are equal, then the *sheaf condition* requires that there be *exactly one* f in $F(U_1 \cup U_2)$ which restricts to f_1 in $F(U_1)$ and to f_2 in $F(U_2)$. A presheaf that satisfies the sheaf condition is called a *sheaf*.

A colimiting process is used to complete a presheaf, converting it to a sheaf. Locally defined elements or sections (i.e. elements in the $F(U)$ for U open in X) are glued together to make *globally* defined ones, whilst any non-uniqueness that would result from this process is killed off by the formation of a quotient, again using a colimit.

Examples of sheaves abound in mathematics and are beginning to be noticed as a potentially useful tool in theoretical computer science, see Goguen [20] and Ehrich *et al* [15]). The most easily accessible example to the non-expert is perhaps the sheaf of continuous real-valued functions on a space, where $F(U)$ is the set of continuous functions on the open set U of X. It is easy to replace continuity by upper semi-continuity or if the space has extra structure (for instance if it is a differential or complex analytic manifold) then the functions used can be those appropriate to that structure. In general if $p : Y \to X$ is any continuous map, then one can define a sheaf of sections of p, often denoted $\Gamma(p)$, in which for an open set U of X, $\Gamma(p)(U) = \{s : U \to Y \mid ps(x) = x \text{ for all } x \in U\}$. In fact all sheaves arise in this way and given a sheaf on X, one can find a suitable $Y \to X$ giving the particular sheaf F as its sheaf of sections, or, more exactly, giving one that is isomorphic to it. The case of the sheaf of real-valued functions is given by taking $Y = \mathbb{R} \times X$ with p being the projection.

If F and G are two sheaves on X then they can be described efficiently as functors from the dual of the lattice $\mathsf{Open}(X)$ of open sets of X to the category Sets. As F and G are functors, the natural definition of a morphism from F to G is a natural transformation between the two functors. This translates as follows:

Definition 2. A *sheaf morphism* $\phi : F \to G$ between two sheaves F and G on X is a family $\{\phi(U) : F(U) \to G(U) \mid U \in \mathsf{Open}(X)\}$ of functions between the sets in the two families which are *compatible* with the restriction maps in the sense that if V is an open subset of U then

$$\phi(V)\mathrm{rest}F(U,V) = \mathrm{rest}G(U,V)\phi(U) \ .$$

If F is a sheaf on X, and $f : X \to Y$ is a continuous map, then f gives a morphism of ordered sets $f^{-1} : \mathsf{Open}(Y) \to \mathsf{Open}(X)$ that maps an open set U in Y to the open set $f^{-1}(U) = \{x \mid f(x) \in U\}$. Composing this with F gives a sheaf G on Y defined by $G(U) = F(f^{-1}(U))$. The notation often used for this induced sheaf is $f^{-1}F$ and it is called the *direct image* of F along f. Given a sheaf F on X and a sheaf G on Y, a morphism Φ from F to G consists of a pair (f, ϕ) where $f : X \to Y$ is a continuous map and $\phi : f^{-1}F \to G$ is a morphism of sheaves on Y.

The sheaves defined above should more precisely be referred to as sheaves of sets. A presheaf F of sets on X is thus just a functor from $\mathrm{Open}(X)^{op}$ to Sets and it is a sheaf if it satisfies the sheaf condition. There are many examples in which the sets $F(U)$ for U open in X have more structure. For instance, for F the sheaf of continuous real-valued functions on a space, each $F(U)$ is a ring or even a normed algebra and all the restriction maps are ring morphisms or continuous homomorphisms of normed algebras, depending on how much structure is being considered. In that case, F is a sheaf of rings or of normed algebras. It should be clear that any category might serve as a codomain category for presheaves but that for the sheaf condition to make sense some extra properties would be needed. The following singles out a class of categories having suitable properties.

Definition 3. A *concrete category* is a category $\mathbf{C}$ together with a faithful functor $\mathcal{U} : \mathbf{C} \to \mathrm{Sets}$, so for all $X, Y \in \mathbf{C}$, the natural mapping

$$\mathbf{C}(X, Y) \to \mathrm{Sets}(\mathcal{U}Y, \mathcal{U}X)$$

is one-to-one. The set $\mathcal{U}X$ is called the *underlying set* of the object X. The functor $\mathcal{U}$ is said to be the *forgetful functor*.

Typical examples of concrete categories include those of monoids, groups, rings, topological spaces, topological algebras, etc. Many examples are categories of single sorted algebras. For the future development of the ideas of this paper, it may to be necessary to replace Sets as the base category by a more complicated category, thus allowing many sorted algebras of various types, but for the purposes of this exposition the above will suffice.

Definition 4. Given such a concrete category $\mathbf{C}$ and a space X, a *presheaf F with values in* $\mathbf{C}$ or a *presheaf of* $\mathbf{C}$*-objects* is a functor from $\mathrm{Open}(X)^{op}$ to $\mathbf{C}$.

Note that if F is a presheaf with values in $\mathbf{C}$ then $\mathcal{U}F$ is a presheaf of sets.

Definition 5. A presheaf F of $\mathbf{C}$-objects is a *sheaf* if $\mathcal{U}F$ is a sheaf of sets.

Example (based on Goguen, [20]). Any object is known only by the observations made of it. These can be thought of as being functions from some space-time domain into some space of attributes $f : U \to A$. If more than one attribute is observed, then A may be a product $A_1 \times \ldots \times A_n$. The different observed attributes may not be independent and can be assumed to satisfy some functional or relational laws embodied in an expression $P(f)$, which is true if for each X in the domain of f, P is satisfied by the n-tuple $f(x) = (f_1(x), \ldots, f_n(x))$. In such a case, there is a presheaf given by

$$\mathcal{O}(U) = \{f : U \to A_1 \times \ldots \times A_n \mid P(f)\} \ .$$

The morphisms are the restriction maps. Note not all such presheaves need be sheaves, but in most situations such a presheaf is either a sheaf or can be completed to be a sheaf. The completion process may, however, warp the underlying relation P used to define the admissible n-tuples of attributes.

Goguen [20] gives many different specific examples from computer science based on this general construction. This example will be looked at in more detail later.

Certain other facets of the theory of sheaves are worth mentioning. The sheaves on a given space together with the morphisms between them form a category. This category has many of the properties of the category of sets and functions. In fact it is often thought of as a generalised model for set theory (see Johnstone [26], or Goldblatt [21]). The main difference is that it corresponds to a somewhat strange intuitionistic logic, whose truth values are the open sets of the base space. In this set theory, the role of the real numbers is taken by the sheaf of continuous or semi-continuous functions depending on how one forms the sheaf corresponding to the real numbers from that corresponding to the rationals within this set theory. An element in $F(X)$ for a sheaf F on X is called a *global section*. If F is the (semi-)continuous function sheaf then two global sections can be compared using the usual means of analysis, for instance metric space theory. It is also possible to do this purely categorically using ideas of Lawvere. (The following sketch uses a deeper level of category theory than the rest of this article and is only used at one place later on. It may therefore safely be omitted. A reference for the theory is Lawvere [31] and for a recent application in computer science [5].)

In an *enriched* category $\mathbf{C}$ the collections of morphisms between objects in $\mathbf{C}$ form objects in another category; for instance they may carry a topological or algebraic structure. Technically the category used for enriching the structure of $\mathbf{C}$ must be a *monoidal* category. In our case we need only consider the case when this monoidal category is the underlying monoid of the set $\mathbb{R}^+$ of non-negative real numbers with addition. If X is a metric space with metric d then we can form an $\mathbb{R}^+$-enriched category $\mathbf{C}$ whose set of objects is the set of points of X and where $\mathbf{C}(x, y) = d(x, y)$. The composition is given by the triangle inequality for the metric. It is now fairly obvious how to proceed to encode enriched limits etc. in this setting. The detailed theory of complete metric spaces from this viewpoint also involves the theory of enriched adjoint profunctors, that is the enriched version of the *distributeurs* (Bénabou, [3]), which are used by Bourn and Cordier to give a description of the shape category as a category of free algebras (Kleisli category). (This latter theory can be found in Cordier and Porter, [6].)

Other enrichments may possibly be of use. If $\mathbf{C}$ is chosen to have a richer algebraic structure, for instance that of some models of some data types, or automata, then these categories often have natural enrichments. Order-enriched categories have been also been considered in many recent papers on computer science and to some extent the enrichment chosen depends on the model of the storage and analysis of information being used. These issues are discussed briefly in [45].

The important intuition to retain is that a sheaf is obtained by gluing together local information, and a map between sheaves is obtained by gluing together locally defined maps.

Finally in this section, recall the way in which invariants of a sheaf are cal-

culated. Although probably not relevant directly, or in detail, to the problem of pattern recognition, sheaf cohomology may give guidance on how to proceed by way of analogy. One of the constructions of this cohomology consists of considering a covering of the space by open sets. Given a covering $\mathcal{U} = \{U_i : i \in I\}$ of X, one can form a polyhedron $\mathrm{Ner}(\mathcal{U})$, called the *nerve* of $\mathcal{U}$, built up from families of intersecting open sets in the family $\mathcal{U}$. For instance, the vertices of $\mathrm{Ner}(\mathcal{U})$ correspond to the open sets in $\mathcal{U}$; if U and V are in $\mathcal{U}$, there is an edge joining U to V in the polyhedron if $U \cap V$ is non-empty; if U, V, and W are in $\mathcal{U}$, there is a triangular face with vertices U, V, and W if $U \cap V \cap W$ is not empty, and so on. Using information in the sheaf over such finite subfamilies of $\mathcal{U}$ which have non-empty intersection, one builds an invariant of the space (with coefficients in the sheaf) by considering ever finer open covers. A process like this but without the sheaf, leads to one of the approaches to classical geometric shape theory. (See Mardešić and Segal [36], Segal [47], or for a detailed account linking it with the categorical approach, Cordier and Porter [6]. None of these sources attempts to look at the situation where a sheaf is present, and more work on interpretation and detailed modelling will be needed here before this can be directly applied to the recognition problem for grey-level images.)

5 Categorical Shape Theory

(The main reference for this section is Cordier and Porter, [6].)

The basic idea of categorical shape theory is that in any approximating situation, the approximations are what encode the only information that the system can analyse. Formally it is assumed that there exists a category $\mathbf{C}$ of *objects of interest* and a category $\mathbf{A}$ of *archetypes*, together with a functor $\mathbf{K} : \mathbf{A} \to \mathbf{C}$ that allows archetypes to be compared with objects. The importance of the category structure is that it requires one to specify what *transformations* of objects and archetypes are going to be available within the system.

5.1 Categories of Approximations

Suppose that we are given some functor $\mathbf{K} : \mathbf{A} \to \mathbf{C}$ as above, and an object X of $\mathbf{C}$.

Definition 6. An *approximation* to X is a pair (f, A) where A is an object of $\mathbf{A}$, hence an archetype, and $f : X \to \mathbf{K}A$. A morphism between approximations $u : (f, A) \to (g, A')$ is a morphism $u : A \to A'$ of the underlying archetypes, such that $\mathbf{K}(u)f = g$. The category of approximations to X will be denoted $(X, \mathbf{K})$.

This category contains the only information available to the system about the object X. The idea behind the definition of a morphism of approximations is that an approximation (f, A) informs the system of the comparison f between the object X and the archetype A. If u is as above, then the information given by f can be filtered through that given by g and so to some extent f might be considered to be redundant. This is only partially true as it is possible for there

also to be a morphism from (g, A') to (f, A). It would be tempting at this stage to take a limit of the diagram

$$\delta_X : (X, \mathbf{K}) \to \mathbf{A} \ ,$$

$$(f, A) \mapsto A \ ;$$

that is, of the A-component functor. If this limit existed, it would give an object of A that was a better approximation to X than any other one, and hence would assign a definite archetypal label to X. The problem is that such a limit may not exist. For instance, in the geometric form of shape theory, A is a category of polyhedra, and although one can take limits of these polyhedra as spaces, the result need not be a polyhedron. This to some extent explains the intuition about $(X, \mathbf{K})$. It acts as a formal limit of all approximations. This may be compared to the concept of a categorical neuron introduced by Ehresmann and Vanbremeersch, which being a formal colimit of lower-order neuronal patterns, represents an interacting system of lower-order information elements (see later).

5.2 The Shape Category of K

It has been suggested above that $(X, \mathbf{K})$ encodes the only information available to the system about the object X. The idea behind the shape category of the system K is that its morphisms should compare these categories of approximations for the various objects of interest; hence they should be functors, but not all functors are suitable. To gain some insight into which functors should be used note that if $\alpha : X \to Y$, then α *induces* a functor

$$(\alpha, \mathbf{K}) : (Y, \mathbf{K}) \to (X, \mathbf{K})$$

given by sending (f, A) in $(Y, \mathbf{K})$ to $(f\alpha, A)$ in $(X, \mathbf{K})$. These induced functors have two interesting properties:

(i) reversal of direction: the morphism α is from X to Y but $(\alpha, \mathbf{K})$ is from $(Y, \mathbf{K})$ to $(X, \mathbf{K})$;

(ii) stability of A-components: the A-components in (f, A) and in $(f\alpha, A)$ are the same, namely A.

These two properties will be abstracted to give the definition of the shape category of K.

Definition 7. The shape category ShK of the system K has as objects the objects of C, and from X to Y in ShK the morphisms are the functors $F : (Y, \mathbf{K}) \to (X, \mathbf{K})$ that preserve the A-component of objects, so if (f, A) is in $(Y, \mathbf{K})$, then $F(f, A)$ has the form (g, A) for the same A in A, and some $g : X \to \mathbf{K}A$ in C.

Two objects are said to have *the same* K*-shape* if they are isomorphic in ShK.

Intuitively this amounts to saying that the information available via $\mathbf{K}$ is not sufficient to tell the two objects apart, as the $\mathbf{K}$-approximation categories of the two objects are, in a precise sense, equivalent. To recognize an object is to assign an archetypal label to it. This amounts to saying that the given object X and the archetype A, say, corresponding to the label, have the same $\mathbf{K}$-shape, or that X and $\mathbf{K}(A)$ are isomorphic in $\mathsf{Sh}\mathbf{K}$. In this case that the object X is said to be $\mathbf{K}$-*stable* (or simply *stable*, if there is no possibility of confusion from such a shortened form).

For such objects X whose shape is recognizable in this way, the category of approximations $(X, \mathbf{K})$ has an *initial* object, that is, there is a *best* approximation, $f : X \to \mathbf{K}A$. Here *best* means that if $g : X \to \mathbf{K}B$ is some other approximation then there is a morphism $\alpha : A \to B$ of archetypes such that $g = \mathbf{K}(\alpha)f$, and moreover α is the only morphism with this property. Not all objects in $\mathbf{C}$ need be recognizable in this sense. For a given $\mathbf{K}$ it would seem to be important to decide which objects are stable by some characterization independent, if possible, of $\mathbf{K}$. (Hušek's paper, [27], in this volume, contains several examples of systems in which all objects are stable.)

Another pattern recognition problem is that of classification of shapes from the available information. It is clear that this corresponds to deciding when two objects have the same $\mathbf{K}$-shape, so that classifying shapes is equivalent in this sense to the problem of determining isomorphism types within $\mathsf{Sh}\mathbf{K}$. To attempt to do this one can hope to define shape invariants, so that non-isomorphic shapes will give different values for the invariant. The problem of defining such shape invariants is treated in the abstract case in Cordier and Porter [6].

6 Modelling Grey-levels and Colour

6.1 Observations and Objects

Ignoring grey-levels or colour for the moment, geometric objects will be modelled by topological spaces, typically specified as a closed subset of 2- or 3-dimensional space or 4-dimensional space-time if considering a moving or changing object. This is not necessarily a good model, but to examine a better alternative would necessarily involve a detailed discussion of the theory of observations, domain theory, the logic of assertions, and many other topics on the interface between mathematical logic, psychology, philosophy, and computer science (see Vickers, [49] and Barwise, [2]). One of the main points of that discussion, however, is that it questions the observational validity of the concept of point and as a topological space is made up of points, the question arises whether one is wise to model observed objects using concepts that are observationally invalid or at least questionable. There is some similarity between this querying and current models for visual perception mentioned elsewhere in this volume, and perhaps this resemblance is not coincidental.

In the model proposed here for the observation of a physical object, the points of the space are not important as such. The method used will give pointwise information only as a limiting case. As suggested by the example given earlier,

it will be assumed that there is an object A of attributes. This object may be assumed to have extra structure, for example, A may be a product of structures of very different types, some being a normed algebra so that the corresponding attribute might represent a light intensity function, others may be a discrete algebraic structure such as a Boolean algebra giving True-False-type information. Other possibilities might involve a component having a measure-theoretic or probabilistic nature. There is a large choice here and more complex structures will presumably give more detailed models. Several of the papers in this volume show facets of this. For example, Noest [43] considers orientation as being S^1-valued, velocity as being D^2-valued and disparity either D^2 or I^1-valued, whilst Zhang [50, 51] considers a non-Euclidean visual space. Schmitt [38] considers attributes such as a graph (the skeleton) with a function defined on it and then looks at the extent to which the attributes allowed for the reconstruction of the image. Given these examples, it seems be probable that it will be neccessary to replace a single sorted A by a more complex, many sorted algebra, so that an attribute might be a state in some finite state automaton or complex structured database, storing the possible observations made of a class of objects.

Definition 8. An observation is a function $f : U \to A$ from an open set U of the underlying spatial or space-time domain X. The *presheaf of observations* is formed by defining

$$\mathcal{O}(U) = \{f : U \to A \mid P(f)\} \ ,$$

where, as before, the proposition or relation P expresses some property of observation, embodying the laws that $\mathcal{O}$ is to satisfy. The elements of $\mathcal{O}(U)$ are called *local observations*. A *global observation* of the object is an element a of $\mathcal{O}(X)$.

To ensure that local observations glue together, it may be necessary to complete this presheaf $\mathcal{O}$. In this process one *does* obtain information at each point x of X by forming the object

$$\mathcal{O}_x = Colim\,\{\mathcal{O}(U) \mid x \in U\} \ ,$$

This construction thus corresponds to considering pointwise observations as the limit of local observations. It will be assumed from now on that the basic objects considered are sheaves and not just presheaves. An observation of such an object will be a global observation in the above sense, that is a global section of the sheaf $\mathcal{O}$. The category $\mathbf{B}$ of objects of interest will be the category of such global observations of attributes of spatial or space-time domains. Formally the definition will be :

Definition 9. The category of *global observations* has as objects triples $(X, \mathcal{O}, a)$ where X is a closed subset (of $\mathbb{R}^3$ or of $\mathbb{R}^4$), $\mathcal{O}$ is a sheaf of observations, and a is a global section of $\mathcal{O}$. A morphism $\Phi : (X, \mathcal{O}_X, a_X) \to (Y, \mathcal{O}_Y, a_Y)$ is a sheaf morphism $\Phi = (f, \phi)$, where, as in Definition 2, $f : X \to Y$ is a continuous function and $\phi : f^{-1}\mathcal{O}_X \to \mathcal{O}_Y$ is a sheaf morphism over Y, so that $\phi(Y)(a_X) = a_Y$.

6.2 Hierarchical Systems and Attributes

In a real-life recognition system, observations may be considered to be the complex synaptic patterns of neuronal impulses arriving in the brain from the retina caused by light from some real object. The archetypes could then be thought of as being memorized synaptic patterns (see von der Malsburg and Bienenstock [35]). Following Livingstone and Hubel [33], it seems that these objects and transformations of them may be built up from localized information coming from high-resolution analysis of detail and, in addition, from gluing information including information for figure-background discrimination, motion detection, and relative positions in space. Thus both local and global information is obtained.

The recent work of Ehresmann and Vanbremeersch [13] suggests a categorical model of such a system, compatible with the ideas suggested here. This theory combines a "localist" viewpoint (see Arbib [1] pp. 97–98 for a very brief description) with the conjectured models of biological neural networks in which the nodes of the network associated with concepts are interpreted as assemblies or virtual nodes; that is they are nodes corresponding to interacting families of neurons rather than actual physical neurons. Their model is phrased in the language of hierarchical systems. At the basic level, the model of an actual neural network is the category of paths on the graph whose vertices are the neurons and whose edges from a neuron N to a neuron N^* are the synapses with their pre-synaptic part in N and their post-synaptic part in N^*. Such a category of paths was considered earlier by Mink'o and Petunin [40].

Definition 10 (see Ehresmann and Vanbremeersch [8] p.29). A *hierarchical system* is a category $\mathbf{C}$ in which the class of objects is divided into levels, labelled with the natural numbers 0, 1, ... , p, such that each object of level $n+1$ for $n < p$ is the colimit in $\mathbf{C}$ of a diagram of objects of level n.

The neuron category would then represent level 0, with concepts appearing at higher levels as patterns of interacting neurons. The hierarchical system considered can evolve in time reacting to external stimuli, memorizing patterns and forming concepts and thus, in terms of categorical shape theory, archetypes.

In their paper [14], Ehresmann and Vanbremeersch adopt a modular theory of brain function, postulating the existence of distinct modules to treat specific features. "The modules treat objects and discriminate two objects according to a specific attribute without considering their resemblances or differences for other attributes."

The information stored in such a modularized system thus loosely corresponds to the use of the object of attributes earlier in this paper. Their modular hierarchical models are examples of what they call Memory Evolutive Systems (MES) and "the architecture of a MES is a compromise between a parallel-distributed processing (system) with a modular organization and a hierarchical associative network ... ." In other words, their modules are many sorted algebras modelling a computing system sometimes like an object-oriented database, sometimes like a neural network (e.g. a Hopfield net).

The geometry modules provide the geometric information combining local information with information on how it is to be glued to give a global picture. The colour modules handle information that is not involved directly in this determination of spatial form. The assignment of a colour attribute to a geometric feature provides the attribute function that is globally defined. The use of many-sorted algebras complicates this picture considerably as it is no longer simply the assignment of an element of some set with structure that is needed. (An indication of how one might get around this is suggested by Goguen [20] when he claims "Behaviour is limit", interpreting the limit of a diagram as its behaviour.) Ignoring this difficulty, it is clear that there is a similarity of approach inherent in this biological systems model. The fact that the type of category theory being used is also that used within theoretical computer science for the semantics of object-oriented programming languages, for modelling modular systems, and relational databases suggests that there is perhaps some hope of using insights from this theory for the development of new automatic pattern recognition languages.

An interesting question arises as to whether grey-level processing should be thought of as being distinct from orientation and structural processing. The assumption is sometimes made in mathematical morphology that the grey-level profile should be considered as a graph and then handled combinatorially (see Heijmans [22]). This assumes that they should be handled together; compare also the critique by Ronse [46].

6.3 The Categorical Model and Local Considerations

In this categorical model, the sheaf-theoretic interpretation might be questioned in as much as the neuronal pattern is not, in fact, a space, but is merely a colimit of simpler patterns in a hierarchical system. What is to be meant mathematically by *local* in such a situation? To answer this problem in detail would seem to be quite difficult. The answer may be to use a combination of the fact that *local* has a definite meaning in the object that is generating the pattern and also, as that pattern is a formal colimit of a diagram of lower order neurons, the linking between those neurons must presumably reflect that external local structure. Sheaf theory and category theory would suggest that the structure of a locale or an internal Grothendieck topology might be useful, but these ideas are not *clearly* those that are needed. A model roughly on the lines suggested here, but paying more attention to the local information processing, might replace sheaves of sets by fibred categories or indexed categories and would correspond more closely to that which would seem to be implied by the hierarchical systems approach of Ehresmann and Vanbremeersch. Such theories would involve *categories* of local information not just sets. Such a development may be necessary later and would correspond to the way in which categorical methods within computer science are evolving.

6.4 Modelling Archetypes

The archetypes are idealised or remembered patterns. Thus it will be assumed that they have similar structure. The actual processes of memory and the formation of concepts are not needed here (see von der Malsburg and Bienenstock [35] and again Ehresmann and Vanbremeersch [11]). The archetypes are thus either a memorised physical (coloured) object and thus modelled by a space, sheaf, and section as would be a real-life object, or alternatively one might attempt to model the neuronal pattern using the language of Ehresmann and Vanbremeersch [11]. Again the sheaf-theoretic picture would seem to work well enough for a first approximation at least.

6.5 Comparison Maps and Transformations

As defined earlier, in the simplest space-sheaf-section model of observations, the comparison maps or transformations will all be sheaf morphisms that send one global section to the other. If the object of attributes is endowed with a metric space structure, underlying perhaps an algebraic one, then it will be possible to consider an enrichment of the categorical model using the ideas of Lawvere sketched out at the end of Sect. 4. The formal categorical shape theory would not change in essence, but convergence within the sets of comparison maps would be able to be modelled categorically. This has not yet been done in detail.

In general, the archetypes and the transformations will form a category as will the patterns and their transformations. Even in the case of natural recognition systems, these categories will differ from person to person as the patterns involve filtered and preprocessed information, whilst the archetypes depend on memorized patterns and hence both depend on previous experience, cultural background, etc. In particular the transformations involved may be very different, as may be witnessed in the different speeds at which different people can identify deformed images, or can visually unmix images of knotted string. The increasing time delay corresponding to increased complexity of the deformation may be explainable if a generating set of potential transformations from which others are built by composition is postulated. It could still formally be assumed that the whole of the class of transformations was available just as in a study of a computer language, in which general formal results can apply to sentences that are so long that they could not be physically generated.

If the archetypes are memorized patterns (possibly simplified in some way), then the process of remembering gives a functor $\mathbf{K}$ from the category of archetypes to that of neuronal patterns. This functor is by its very nature unknown but properties of it could be investigated by studying the properties of this formal recognition system made up of the patterns, the archetypes, and $\mathbf{K}$.

In an automatic recognition system, the choice of generating transformations is clearly of importance. This may indicate a link with mathematical morphology, where thickening and thinning operations based on the basic operations of dilation and erosion are by Matheron's theorem, the basic building blocks of the transformations used; (see Heijmans [22] or [23] for an introduction to this

area). This link does not extend to the way in which grey-level images would be handled however.

The intuition of these transformations, and similarly of the comparison morphisms used to compare an input pattern with an archetypal pattern, is that they are basically geometric, being rotations, translations, or affine maps, at least at the spatial or structural level, and the categorical notation of an arrow is consistent with this as it is an extension of the notation often used for functions. However, as was mentioned earlier, modern theoretical computer science, in its use of categorical language, has emphasized several different interpretations of categories and therefore of the arrows that they contain. For instance an arrow may be interpreted as a proof, a transition between states, or a process. This again suggests that the categorical notion of morphism should not be too narrowly interpreted but rather that more detailed models might try to decide between possible interpretations by recourse to non-mathematical theories, experimentation, and data from simulations. One possible direction for research in categorical shape theory is the investigation of it as a formal language, with an eventual aim of developing a customized language for managing the processing of images and an implementation of that language that will integrate the activity of processing modules that handle various attributes of an image.

7 Observational Mechanisms

The categorical shape-theoretic model is formal. It does not presuppose a mechanism and work from there. This article has perhaps concentrated on the biological rather than the engineering context, but how realistic are the assumptions about the processing of sections within either context? In fact can a brain or a man-made neural network do any of the tasks potentially involved in the shape-theoretic description of the recognition process? (For an introduction to neural network theory at a readable level, see Arbib, [1]).

The classical result of Minsky and Papert [41] showed that a simple two-layer perceptron can only represent or approximate a very small class of functions; however they left open the possibility that a multilayer feed-forward net might be capable of doing better. Duda and Fossum [7] showed that any piecewise-linear decision region can be realized by a multilayer network. Clearly this is relevant to the problem of the approximation of polyhedra and thus to the approximation of more general shapes by polyhedra. Lippman [32] argued that arbitrary complex regions can be formed using four-layer networks. Hopfield's work [24] on associative memory then provides a model for the retrieval of archetypal patterns from the memory, involving the non-deterministic minimalization of an energy function. These results and ideas are sufficient to argue for the feasibility of a shape-theoretic model provided that no grey-levels or colour are involved, but can one justify the extension proposed here? The solution came in 1989 when Funahashi [16] and shortly after, Hornik *et al* [25] proved that feed-forward multilayer networks are capable of approximating arbitrarily closely any continuous function defined between compact sets, provided that there are enough hidden units.

(More recent results, again by Hornik *et al* [26], showed that with some restrictions, such networks can approximate an arbitrary function and its derivatives if they exist.) Although of course any brain or machine has only finitely many neurons available, these results indicate that even with the limited power of the machines available today, the hypotheses of shape theory that presume arbitrarily fine approximations to spaces and to locally defined (upper-semi-)continuous functions may not be as unreasonable as they might seem and hence that the simple sheaf-theoretic model for observations as proposed above may be able to be implemented with no great difficulty.

8 Conclusion

Without the development of categorical shape theory beyond its present stage (e.g. to use enriched category theory), the question posed as the title of this article depends mainly on the possibility or non-possibility of finding categories that can act as categories of objects and archetypes relevant to modelling grey-level images or more generally, coloured images or objects. A possible solution has been sketched here, namely that both the objects of interest and the archetypes be modelled by structures consisting of a space, a sheaf, and a section of that sheaf. To some extent, the spatial nature of the space may be in question, but to avoid that assumption would have needed a discussion of much more mathematics than space would allow. The use of a space does permit one to model the local/global interaction without too much difficulty, and sheaf theory then suggests a way of handling the integration of the local grey-level or colour information into the global picture. A further elaboration of this will be needed if the theory is to be rich enough to act as an adequate model.

The nature of the space of attributes will govern the applicability of the sheaf-theoretic approach. For instance, if developments in information modelling suggest, as looks very possible, that a better model would be a complex categorical structure such as an abstract data type, then there would probably be a need to enrich the sheaf-theoretic model; however the simple sheaf-theoretic intuition would remain usable as a first approximation.

Another advantage of a sheaf-theoretic model is that the theory handles comparison of objects fairly easily via the induced sheaf construction and the good properties of categories of sheaves on a fixed space. A categorical approach to pattern recognition does require that an analysis of objects includes an analysis of their allowable deformations, and this, in turn, puts demands on the type of model used. The model proposed here and the more structured categorical models that will perhaps be needed for finer modelling pass this test with flying colours.

References

1. Arbib, M.A. (1987). Brains, Machines and Mathematics, 2nd Edition, Springer-Verlag, Berlin.

2. Barwise, J. (1991). Information links in domain theory, Indiana University Logic Group, Preprint No. IULG-91-7.
3. Bénabou, J. (1973). Les distributeurs, Rapport 33, Inst. Math. Pure Appl. Univ. Louvain-la-Neuve.
4. Bourn, D., Cordier, J.-M., (1980). Distributeurs et théorie de la forme, Cahiers Top. et Géom. Diff. 21, pp. 161–189.
5. Cazley, R., Crew, R.F., Meseguer, J., Pratt, V. (1991). Temporal structures, Math. Struct. in Comp. Science 1, pp. 179–213
6. Cordier, J.-M., Porter, T. (1989). Shape Theory : Categorical Methods of Approximation, Ellis Horwood, Chichester, UK.
7. Duda, R.O., Fossum, H. (1966). Pattern classification by iteratively determined linear and piecewise linear discriminant functions. IEEE Transactions on Electronic Computers EC-15, pp. 220–232.
8. Ehresmann, A.C. , Vanbremeersch, J.-P. (1987). Hierarchical evolutive systems: a mathematical model for complex systems, Bull. Math. Biol. 49, pp. 13–50.
9. Ehresmann, A.C., Vanbremeersch, J.-P. (1987). A Mathematical model for complex systems, II: trial and error dynamics with hierarchical modulation. Preprint.
10. Ehresmann, A.C., Vanbremeersch, J.-P. (1989). Modèle d'interaction dynamique entre un système complexe et des agents, Revue Intern. Systémique 3, pp. 315–341.
11. Ehresmann, A.C. , Vanbremeersch, J.-P. (1989). Systèmes hiérarchiques évolutifs à mémoire auto-régulée synergie et cohérence dans les systèmes biologiques, Séminaire Transdisciplinaire, Centre Interuniversitaire Jussieu - St.Bernard, Mai.
12. Ehresmann, A.C. , Vanbremeersch, J.-P. (1990). Hierarchical evolutive systems, 8th International Conference of Cybernetics and Systems, New York.
13. Ehresmann, A.C., Vanbremeersch, J.-P. (1991). Un modèle pour des systèmes évolutifs avec mémoire basé sur la théorie des catégories, Revue Intern. Systémique, 5 (1), pp. 5–25.
14. Ehresmann, A.C., Vanbremeersch, J.-P. (1992). Semantics and communication for memory evolutive systems, 6^{th} International Conference on Systems Research, Informatics and Cybernetics, Baden-Baden.
15. Ehrich, H.-D., Goguen, J.A., Sernadas, A. (1990). A categorical theory of objects as observed processes, Proc. REX/FOOL, Noordwijkerhout.
16. Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks, Neural Networks 2, pp. 183–192.
17. Girard, J.-Y. (1987). Linear logic, Theoret. Comput. Sci. 50, pp. 1–102.
18. Girard, J.-Y. (1989). Towards a geometry of interaction, In: Gray, J.W. and Scedrov, A., (eds.), Categories in Computer Science and Logic, Boulder, June 1987, Vol. 92 of Contemporary Mathematics, American Mathematical Society, pp. 69–108.
19. Goguen, J.A. (1991). A categorical manifesto, Math. Struct. in Computer Science 1, pp. 49–67.
20. Goguen, J.A. (1992). Sheaf semantics for concurrent interacting objects, Math. Struct. in Comp. Science, 2, pp. 159–191.
21. Goldblatt, R. (1984). Topoi, the Categorical Analysis of Logic, North-Holland, Amsterdam.
22. Heijmans, H.J.A.M. (1987). Mathematical morphology: an algebraic approach, CWI Newsletter 14, pp. 7–27.
23. Heijmans, H.J.A.M. (1993). Mathematical morphology as a tool for shape description, this volume, pp. 147–176.

24. Hopfield, J.J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Nat. Acad. Sci. USA, 81, pp. 3088–3092.
25. Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators, Neural Networks 2, pp. 359–366.
26. Hornik, K., Stinchcombe, M., White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Networks 3, pp. 551–560.
27. Hušek, M. (1993). Introduction to categorial shape theory, with applications in mathematical morphology, this volume, pp. 91–110.
28. Johnstone, P.T. (1977). Topos Theory, Academic Press.
29. Lambek, J. (1968). Deductive systems and categories I, Math. Sys. Theory 2, pp. 287–318.
30. Lawvere, F.W. (1969). Adjointness in foundations, Dialectica 23, pp. 281–318.
31. Lawvere, F.W. (1973). Metric spaces, generalised logic and closed categories, in Rendiconti del Seminario Matematico e Fisico di Milano XLIII, Tipografia, Pavia.
32. Lipmann, R.P. (1987). An introduction to computing with neural nets, IEE ASSP Magazine 4, pp. 4–22.
33. Livingstone, M.S. , Hubel, D.H. (1988). Segregation of form, color movement and depth: anatomy, physiology, and perception, Science 240, pp. 740–749.
34. Mac Lane, S. (1971). Categories for the Working Mathematician, Grad. Texts in Mathematics 5, Springer-Verlag, Berlin.
35. Malsberg, C. van der , Bienenstock, E. (1986). Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain. In: Disordered Systems and Biological Organisation, NATO ASI Series F, vol.20, Springer-Verlag, Berlin.
36. Mardešić, S. , Segal, J. (1982). Shape Theory, the Inverse Systems Approach, North-Holland Mathematical Library vol 26, North-Holland, Amsterdam.
37. Martí-Oliet, N. , Meseguer, J. (1991). From Petri nets to linear logic, Math. Struct. in Comp. Science 1, pp. 69–101.
38. Mattioli, J., Schmitt, M. (1993). On information contained in the erosion curve, this volume, pp. 177–195.
39. Meseguer, J., Montanari, U. (1990). Petri nets are monoids, Information and Computation 88, pp. 105–155
40. Mink'o, A. A., Petunin, Y. (1981). Mathematical modeling of short-term memory, Kibernetika 2, pp. 282–297.
41. Minsky, M.L., Papert, S. (1969). Perceptrons: an Essay in Computational Geometry, MIT Press, Cambridge, MA.
42. Moggi, E. (1991). A category-theoretic account of program modules, Math. Struct. in Comp. Science 1, pp. 103–139.
43. Noest, A. (1993). Neural processing of overlapping shapes, this volume, pp. 383–392.
44. Pavel, M. (1991). Fundamentals of Pattern Recognition, Monographs and Textbooks in Pure and Applied Mathematics, 124, Marcel Dekker, New York.
45. Porter, T. (1993). Categorical shape theory as a formal language for pattern recognition, Annals of Maths. and Artificial Intelligence, special issue on "Mathematics in Pattern Recognition", to appear.
46. Ronse, C. (1989). Fourier analysis, mathematical morphology, and vision, PRLB, Working Document WD54, November.

47. Segal, J. (1993). Shape theory: an ANR-sequence approach, this volume, pp. 111–125.
48. Vanbremeersch, J.-P. , Ehresmann, A.C. (1969). A model for a neural system based on category theory, 2nd International Symposium on Systems Research, Informatics and Cybernetics.
49. Vickers, S. (1989). Topology via Logic, Cambridge Tracts in Theoretical Computer Science 5, Cambridge University Press, Cambridge, UK.
50. Zhang, J. (1993). Image representation using affine covariant coordinates, this volume, pp. 353–362.
51. Zhang, J., Wu, S. (1990). Structure of visual perception, Proc. Nat. Acad. Sci., 87, pp. 7819–7823.

Mathematical Morphology as a Tool for Shape Description*

Henk J.A.M. Heijmans

Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

Abstract. Mathematical morphology is an approach in image processing based on geometrical concepts such as transformation groups and metric spaces. As such it is well suited to the extraction of information about the shape of the various parts in a scene. This paper presents an overview of some known morphological techniques (e.g. skeletonisation, granulometric analysis) for the description and decomposition of shape.

Keywords: mathematical morphology, dilation, erosion, opening, closing, complete lattice, umbra transform, transformation geometry, distance geometry, translation-rotation group, granulometry, skeleton, shape decomposition.

1 Introduction

It has been frequently claimed in the literature that mathematical morphology is an approach well suited to the extraction of shape information from a scene. The aim of the present paper is to justify this claim by presenting a number of morphological tools for the description of shape.

Section 2 recalls briefly the basic concepts from mathematical morphology and discusses some elementary morphological operators for binary images. In Sect. 3 it will be explained how such operators can be extended to grey-scale images by means of the umbra transform. Two geometrical concepts lay the foundations of mathematical morphology, namely (i) *geometrical transformations* such as translations, rotations, reflections, perspective transformations, and (ii) *metric spaces and convexity*. Geometrical transformations form the basis for Sect. 4, in particular Sect. 4.1 where transformation-based morphology is discussed in a rather general context. It is shown how an arbitrary transformation group can be used as the basis for a family of morphological operators invariant under these transformations.

An alternative method of constructing morphological operators, discussed in Sect. 4.2, is based on the notion of distance. On any metric space one can define

* The author wishes to acknowledge Adri Steenbeek for implementing the decomposition algorithm.

morphological operators like dilations, erosions, openings, closings, etc. A class of metric spaces particularly important in the context of mathematical morphology is formed by the so-called Minkowski spaces. It turns out that for such spaces the transformation-based approach and the distance-based approach are closely related.

In Sect. 5 granulometries will be discussed. These can be viewed as the mathematical formalisation of a sieving process, and have been applied with success to many practical image analysis problems.

One of the most popular tools for shape description is formed by the skeleton (and its variants). The skeleton can be defined conveniently in terms of morphological operators. The morphological definition makes it rather easy to define skeletons based on a distance other than the Euclidean distance. As in Sect. 5, 'convexity' is the important word here.

Section 7 explains how to use morphological openings as a tool for shape decomposition. Actually, two decomposition algorithms will be discussed here, the first due to Pitas and Venetsanopoulos [22] and the second to Ronse [27, 29]. The exposition presented here is largely taken from the paper by Ronse [27] where a more general approach has been discussed.

Section 8 concludes with some additional remarks.

2 Basic Notions

There is considerable literature on mathematical morphology. Basic references are the monographs by Matheron [19] and Serra [30]. A basic account can also be found in [7]. A second volume, edited by Serra [31], treats a number of theoretical issues; a substantial part of this book is devoted to the theory of morphological filters. In this section some basic material will be presented. More details can be found in the references listed above.

The central idea of mathematical morphology is to examine the structural content of an image by matching it with small patterns at various locations in the image. By varying the size and the shape of the matching patterns, called *structuring elements*, one can obtain useful information about the shape of the image. Such an approach results in nonlinear image operators which are well suited to the analysis of the geometrical and topological structure of an image.

Originally, mathematical morphology was developed for binary images which can be represented mathematically as sets. The corresponding morphological operators essentially use only four ingredients from set theory, namely set intersection, union, complementation, and translation.

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of subsets of $\mathbb{R}^d$ and choose a structuring element $A \subseteq \mathbb{R}^d$. The Minkowski addition and subtraction are resp. defined as

$$X \oplus A = \bigcup_{a \in A} X_a \tag{1}$$

$$X \ominus A = \bigcap_{a \in A} X_{-a}, \tag{2}$$

where X_a is the translate of X along the vector a. Instead of (1) we can also write

$$X \oplus A = \{h \in \mathbb{R}^d \mid \check{A}_h \cap X \neq \emptyset\} \tag{3}$$

$$X \ominus A = \{h \in E \mid A_h \subseteq X\}, \tag{4}$$

where $\check{A}$ is the reflection of A with respect to the origin, that is, $\check{A} = \{-a \mid a \in A\}$.

Fig. 1. A set (left), and its dilation (middle) and erosion (right) with a disk.

Usually, one refers to the Minkowski addition (1) as the *dilation* by A, and to the Minkowski subtraction (2) as the *erosion* by A. Dilation and erosion are illustrated in Fig. 1. We introduce the notation $\delta_A(X) = X \oplus A$ and $\varepsilon_A(X) = X \ominus A$. In general, the operators δ_A and ε_A are not each other's inverses, that is, $(X \oplus A) \ominus A \neq X \neq (X \ominus A) \oplus A$. The operator

$$X \circ A = (X \ominus A) \oplus A \tag{5}$$

is called the *opening* by A. It is easy to show that

$$X \circ A = \bigcup \{A_h \mid h \in \mathbb{R}^d \text{ and } A_h \subseteq X\}.$$

In other words, $X \circ A$ is the union of all translates of the structuring element A which are contained in X. An example is given in Fig. 2.

The opening has the following properties: it is

- increasing, i.e., $X \subseteq Y$ implies that $X \circ A \subseteq Y \circ A$;
- translation invariant, i.e., $X_h \circ A = (X \circ A)_h$;
- anti-extensive, i.e., $X \circ A \subseteq X$;
- idempotent, i.e., $(X \circ A) \circ A = X \circ A$.

Note that the first two properties also hold for dilations and erosions. Every operator $\alpha : \mathcal{P}(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d)$ which is increasing, anti-extensive and idempotent is called an opening. If α_1, α_2 are openings on $\mathcal{P}(\mathbb{R}^d)$, then

$$\alpha_1 \leq \alpha_2 \iff \alpha_1\alpha_2 = \alpha_1 \iff \alpha_2\alpha_1 = \alpha_1.$$

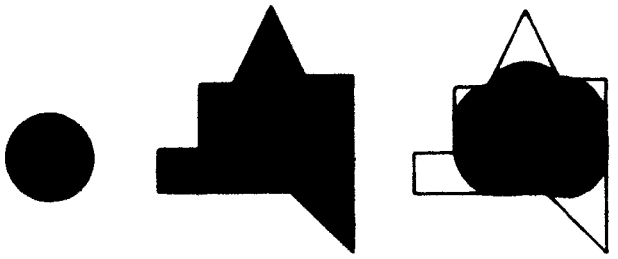


Fig. 2. The opening $X \circ A$ (right) of the set X (middle) with the disk A (left).

Here '$\alpha_1 \leq \alpha_2$' means that $\alpha_1(X) \subseteq \alpha_2(X)$ for every $X \in \mathcal{P}(\mathbb{R}^d)$, and '$\alpha_1\alpha_2$' is the composition of α_1 and α_2, i.e., $\alpha_1\alpha_2(X) = \alpha_1(\alpha_2(X))$. For the openings $X \to X \circ A$ it can be shown that $X \circ A \subseteq X \circ B$ for every X if and only if A is B-open. The latter means that $A \circ B = A$. For example, this condition holds if A is a square with sides 1 and B a line segment with length ≤ 1.

The operator given by

$$X \bullet A = (X \oplus A) \ominus A$$

is called the *closing* by A and has the same properties as the opening apart from the third: it is extensive instead of anti-extensive. The latter means that $X \subseteq X \bullet A$ for every $X \in \mathcal{P}(\mathbb{R}^d)$.

The observation that Minkowski addition and subtraction are not each other's inverses motivated Ghosh [6] to address himself to the problem of extending $\mathcal{P}(\mathbb{R}^d)$ with so-called *negative shapes* in such a way that the space becomes a group under Minkowski addition.

Recently, mathematical morphology has been extended to the framework of complete lattices. Recall that a complete lattice is a partially ordered set in which every subset has an infimum (greatest lower bound) $\bigwedge$ and supremum (smallest upper bound) $\bigvee$; see [1]. The space $\mathcal{P}(\mathbb{R}^d)$ with the inclusion order is a complete lattice. For a comprehensive account of the extension of mathematical morphology to complete lattices refer to [13, 28, 31] and [11].

Definition 1. Let $\mathcal{L}$ be a complete lattice. An operator $\delta : \mathcal{L} \to \mathcal{L}$ is called a *dilation* if it distributes over arbitrary suprema, that is,

$$\delta(\bigvee_{i \in I} X_i) = \bigvee_{i \in I} \delta(X_i),$$

for any family $\{X_i \mid i \in I\}$. Dually, an operator $\varepsilon : \mathcal{L} \to \mathcal{L}$ is called an *erosion* if it distributes over arbitrary infima, that is,

$$\varepsilon(\bigwedge_{i \in I} X_i) = \bigwedge_{i \in I} \varepsilon(X_i),$$

for any family $\{X_i \mid i \in I\}$.

A pair of operators (ε, δ), both mapping $\mathcal{L}$ into $\mathcal{L}$, is called an adjunction on $\mathcal{L}$ if for every $X, Y \in \mathcal{L}$,

$$\delta(X) \leq Y \iff X \leq \varepsilon(Y). \tag{6}$$

If (ε, δ) is an adjunction, then δ is a dilation and ε an erosion. Moreover, with every dilation δ one can associate a unique erosion ε so that (ε, δ) forms an adjunction. We say that ε and δ are *adjoint operators*. If (ε, δ) is an adjunction on $\mathcal{L}$ then $\delta\varepsilon$ is an opening and $\varepsilon\delta$ a closing.

The pair $(\varepsilon_A, \delta_A)$ introduced above forms an adjunction on $\mathcal{P}(\mathbb{R}^d)$. In Sect. 4 some other examples will be discussed.

3 Grey-scale Morphology

Many binary morphological operators can be extended to grey-scale images (modelled mathematically as functions). Denote by $\mathrm{Fun}(E)$ the space of functions mapping E into $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. It is easy to check that $\mathrm{Fun}(E)$ is a complete lattice. If $E = \mathbb{R}^d$, the Minkowski addition and subtraction of two functions F and G can be defined as

$$(F \oplus G)(x) = \bigvee_{h \in E} [F(x-h) + G(h)], \tag{7}$$

and

$$(F \ominus G)(x) = \bigwedge_{h \in E} [F(x+h) - G(h)]. \tag{8}$$

The opening is given by $F \circ G = (F \ominus G) \oplus G$, where G is called the *structuring function*.

A general approach to extend binary morphological operators to functions is provided by the *umbra transform*. For an extensive discussion refer to [10]. The key idea is to represent a function F on the space E by the set of points in $E \times \mathbb{R}$ on and below the graph of E. The resulting set is called an umbra.

Definition 2. Let E be an arbitrary set.
(a) A set $U \subseteq E \times \mathbb{R}$ is called an *umbra* if $(x, t) \in U$ if and only if $(x, s) \in U$ for every $s < t$.
(b) A subset $U \subseteq E \times \mathbb{R}$ is called a *pre-umbra* if $(x, t) \in U$ implies that $(x, s) \in U$ for every $s < t$.

For an illustration refer to Fig. 3.

The set of all umbras is denoted by $\mathrm{Umbra}(E)$. For a subset $X \subseteq E \times \mathbb{R}$ we define $\mathcal{U}_s(X)$ as the smallest umbra containing X. In other words, $\mathcal{U}_s(X)$ is the intersection of all umbras containing X; see Fig. 4. If F is a function mapping a set E (usually $\mathbb{R}^d$ or $\mathbb{Z}^d$) into $\overline{\mathbb{R}}$, then we define the umbra $\mathcal{U}_f(F)$ of F as

$$\mathcal{U}_f(F) = \{(x, t) \in E \times \mathbb{R} \mid t \leq F(x)\}; \tag{9}$$

see Fig. 4.

Fig. 3. A pre-umbra (left) and an umbra (right).

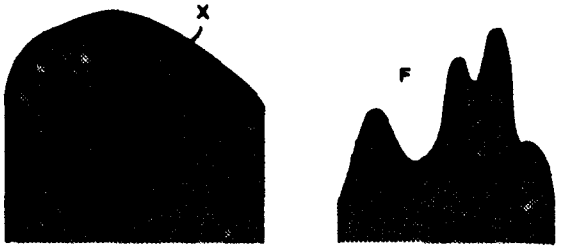


Fig. 4. The umbras $\mathcal{U}_s(X)$ of a set X (left) and $\mathcal{U}_f(F)$ of a function F (right).

The subscripts s and f refer to set and function respectively. The mapping $\mathcal{U}_f : \mathsf{Fun}(E) \to \mathsf{Umbra}(E)$ is called the *umbra transform*.

To every umbra U corresponds a unique function given by

$$[\mathcal{F}_s(U)](x) = \bigvee\{t \in \mathbb{R} \mid (x,t) \in U\}.$$

From now on $\mathcal{F}$ and $\mathcal{U}$ will be used to denote $\mathcal{F}_s$ and $\mathcal{U}_f$ respectively.

Proposition 3. $\mathsf{Umbra}(E)$ *with the set inclusion as partial order is a complete lattice with infimum and supremum of* U_i, $i \in I$, *respectively given by* $\bigwedge_{i\in I} U_i = \bigcap_{i\in I} U_i$ *and* $\bigvee_{i\in I} U_i = \mathcal{U}_s(\bigcup_{i\in I} U_i)$. *This lattice is isomorphic to* $\mathsf{Fun}(E)$ *with the isomorphism and its inverse respectively given by* $\mathcal{F}$ *and* $\mathcal{U}$.

Given a scalar $v \in \mathbb{R}$, the vertical translate of a set $X \subseteq E \times \mathbb{R}$ and a function $F \in \mathsf{Fun}(E)$ are respectively defined as $X^v = \{(x, t+v) \mid (x,t) \in X\}$ and $F^v(x) = F(x) + v$. It is obvious that $X \to X^v$ maps a (pre-)umbra onto a (pre-)umbra. Furthermore $\mathcal{F}(U^v) = [\mathcal{F}(U)]^v$ and $\mathcal{U}(F^v) = [\mathcal{U}(F)]^v$.

Lemma 4. *Let* $U \subseteq E \times \mathbb{R}$.

(a) U *is a pre-umbra if and only if* $U \subseteq U^v$ *for every* $v > 0$.
(b) U *is an umbra if and only if* $U = \bigcap_{v>0} U^v$.
(c) *If* U *is a pre-umbra then* $\mathcal{U}_s(U) = \bigcap_{v>0} U^v$.

For a proof refer to [10].

The umbra transform is used to map operators on $\mathcal{P}(E \times \mathbb{R})$ to operators on $\mathsf{Fun}(E)$. Assume that ψ is an increasing operator on $\mathcal{P}(E \times \mathbb{R})$ which is invariant under vertical translations. The latter means that $\psi(X^v) = [\psi(X)]^v$ for $X \subseteq E \times \mathbb{R}$. If U is a pre-umbra, then $U \subseteq U^v$ for $v > 0$ and hence

$$\psi(U) \subseteq \psi(U^v) = [\psi(U)]^v;$$

in other words, $\psi(U)$ is a pre-umbra as well. The operator $\tilde{\psi}$ on $\mathcal{P}(E \times \mathbb{R})$ is defined as

$$\tilde{\psi}(X) = \bigcap_{v>0} [\psi(X)]^v = \bigcap_{v>0} \psi(X^v). \tag{10}$$

From Lemma 4 it follows that $\tilde{\psi}(U) = \mathcal{U}_s(\psi(U))$ if U is a pre-umbra, and thus $\tilde{\psi}(U)$ is an umbra. This shows in particular that $\tilde{\psi}$ leaves $\mathsf{Umbra}(E)$ invariant.

Theorem 5. *Given an increasing operator ψ on $\mathcal{P}(E \times \mathbb{R})$ which is invariant under vertical translations, the operator Ψ given by*

$$\Psi = \mathcal{F} \circ \tilde{\psi} \circ \mathcal{U} \tag{11}$$

defines an increasing operator on $\mathsf{Fun}(E)$ invariant under vertical translations.

Obviously, if ψ is invariant under translations in E, then the same holds for Ψ.

Consider, as an example, the Minkowski addition for functions. Given a function G on $\mathbb{R}^d$, the operator $F \to F \oplus G$ may be derived from the above construction with $\psi(X) = X \oplus \mathcal{U}(G)$.

Define the domain $\mathsf{dom}(F)$ of a function F as the set of all $x \in \mathbb{R}^d$ for which $F(x) > -\infty$. If G is a function which assumes the value 0 on its domain A and $-\infty$ elsewhere, then the resulting dilation (erosion, etc.) is called a *flat dilation* (erosion, etc.). More generally, a flat function operator can be defined as follows. Given an increasing binary operator ψ_0 on $\mathcal{P}(E)$, define an increasing operator ψ on $\mathcal{P}(E \times \mathbb{R})$ by putting

$$\pi_t \psi(X) = \psi_0(\pi_t X),$$

for $X \subseteq E \times \mathbb{R}$ and $t \in \mathbb{R}$. Here π_t is the operator given by $\pi_t X = \{x \in E \mid (x,t) \in X\}$. In other words, $\psi(X)$ is the set obtained by applying ψ_0 to every cross section $\pi_t X$. The extension of ψ given by (11) yields a grey-scale operator Ψ. If the threshold sets of F are defined as $X(F,t) = \{x \in E \mid F(x) \geq t\}$ then $\Psi(F)$ is given by

$$\Psi(F)(x) = \sup\{t \in \mathbb{R} \mid x \in \psi_0(X(F,t))\}.$$

The operator Ψ is called the *flat extension* of ψ_0 to $\mathsf{Fun}(\mathbb{R}^d)$, and inherits most properties of ψ_0; for instance, if ψ_0 is an opening then Ψ is such as well. Refer to [9] for more details.

4 Morphology Versus Geometry

Generally speaking, morphological operators are defined by moving a small test pattern over the image, checking at all positions how it relates to the image and using the outcome to define an output image. In classical translation morphology, 'moving' means 'translating'. But one can think of situations where translation is not appropriate, or even worse, not possible. We mention some examples.

In certain applications, for example, radar imaging, rotation symmetry comes in naturally. In such cases one has to include rotations in the group of permitted

motions. Similar remarks apply to situations where perspective transformations play a role; think, for instance, of the problem of monitoring the traffic on a highway with a camera at a fixed position. It is apparent that in this case the detection algorithms must take into account the distance between the camera and the cars.

If the underlying support space is not just the Euclidean space $\mathbb{R}^d$ or a regular grid, but rather a manifold (e.g. the sphere), translation has to be understood in the sense of parallel transport along geodesics, as Roerdink explains in [26]. In general the motion of a pattern along geodesics is quite troublesome. However, for some specific examples such as the sphere, it is possible to obtain concrete results [25, 24].

Another class of images which requires reflections about ways of matching patterns is formed by the graph-based images. In a number of applications a graph provides the appropriate mathematical structure to model an image. This occurs when the image contains a large amount of relatively small objects (e.g. cells in an electron microscopy image). In such cases the edges of the graph can be used to model the spatial relationships between the objects. Heijmans *et al.* [12, 14] use the notion of a structuring graph to define morphological operators on such graphs. A more direct approach, based on distance, was given by Vincent [35]. In the latter approach the central idea is to define a pattern at every position (a ball with given radius) rather than moving around a given pattern using a given group of transformations.

In this section both approaches will be discussed in a more general context. For the sake of exposition the discussion is restricted to binary images, or, to stay within mathematical terms, to the space $\mathcal{P}(E)$, where E is the support space. This can be the Euclidean space $\mathbb{R}^d$, the discrete space $\mathbb{Z}^d$, a manifold, a graph, etc. In the transformation approach it is assumed that we are given a transformation group on E. In practical cases the choice of this group is often determined by the underlying mathematical structure of E. In the distance approach it is merely assumed that E is a metric space.

4.1 Transformation-based Morphology

This section outlines how basic morphological operators such as dilation, erosion, opening, and closing can be extended to general geometric spaces. A first observation is that translation morphology is rather special since translations define a simply transitive abelian group (definitions below).

Consider the Minkowski addition $X \to X \oplus A$ where $A \subseteq \mathbb{R}^2$. In general, this operation is not invariant under rotations, that is $(R_\varphi X) \oplus A \neq R_\varphi(X \oplus A)$. Here R_φ is the rotation around 0 over an angle φ. In fact, one can show that the operation is rotation invariant if and only if A is, that is, $R_\varphi A = A$ for every $\varphi \in [0, 2\pi]$. Below, this simple example will be put into a more general algebraic framework.

The following account is based on previous work of Heijmans and Ronse [8, 13, 28] and Roerdink [25, 24]. For a comprehensive account refer to [11]. Assume

that E is a set and that $\mathbf{T}$ is a transformation group on E; here transformation means bijective mapping. $\mathbf{T}$ is *transitive* on E if for every $x, y \in E$ there exists a transformation $\tau \in \mathbf{T}$ such that $\tau x = y$. If this transformation is unique for every pair x, y, then $\mathbf{T}$ is called *simply transitive* on E. It is easy to show that for an abelian transformation group, transitivity implies simple transitivity. The group of isometries on $\mathbb{R}^2$ is transitive but not simply transitive. Recall that an isometry is a transformation which preserves distances [23]. The translations form an abelian, simply transitive transformation group on $\mathbb{R}^2$. The rotations around 0 yield an abelian group which, however, is not transitive.

The operator $\psi : \mathcal{P}(E) \to \mathcal{P}(E)$ is called a $\mathbf{T}$-operator if

$$\psi \circ \tau = \tau \circ \psi, \quad \text{for every } \tau \in \mathbf{T}.$$

A $\mathbf{T}$-operator which is a dilation will be called a $\mathbf{T}$-dilation, etc. For $X \subseteq E$ and $\tau \in \mathbf{T}$ define $\tau X = \{\tau x \mid x \in X\}$.

Following the expression $X \oplus A = \bigcup_{a \in A} X_a$, one might attempt to define a $\mathbf{T}$-dilation as follows:

$$\delta_{\mathbf{A}}(X) = \bigcup_{\tau \in \mathbf{A}} \tau X; \tag{12}$$

here $\mathbf{A}$ is an arbitrary subset of $\mathbf{T}$. It is apparent that $\delta_{\mathbf{A}}$ is a dilation in the sense of Definition 1. The adjoint erosion is given by

$$\varepsilon_{\mathbf{A}}(X) = \bigcap_{\tau \in \mathbf{A}} \tau^{-1} X. \tag{13}$$

If the transformation group $\mathbf{T}$ is abelian then $\delta_{\mathbf{A}}$ is a $\mathbf{T}$-dilation and $\varepsilon_{\mathbf{A}}$ a $\mathbf{T}$-erosion. If, in additional, $\mathbf{T}$ is transitive, then every $\mathbf{T}$-dilation and $\mathbf{T}$-erosion on $\mathcal{P}(E)$ are of this form. We consider this case in more detail. First, fix an origin $o \in E$. To every $x \in E$ there corresponds a unique transform $\tau_x \in \mathbf{T}$ which carries o to x, $\tau_x o = x$ (compare this with the relation between affine spaces and vector spaces). For $x \in E$ we define $A(x) \subseteq E$ as

$$A(x) = \{\tau x \mid \tau \in \mathbf{A}\}.$$

Obviously, $A(x) = \tau_x A(o)$, so $A(x)$ can be interpreted as the 'translate' of the structuring element $A(o)$. Then

$$\delta_{\mathbf{A}}(X) = \bigcup_{x \in X} A(x), \quad \varepsilon_{\mathbf{A}}(X) = \{x \in E \mid A(x) \subseteq X\}. \tag{14}$$

The opening $\delta_{\mathbf{A}} \varepsilon_{\mathbf{A}}$ is given by

$$\delta_{\mathbf{A}} \varepsilon_{\mathbf{A}}(X) = \bigcup \{A(h) \mid h \in E \text{ and } A(h) \subseteq X\}. \tag{15}$$

In fact, in this case one obtains all the results also known from the translation invariant case.

Proposition 6. *Let* $\mathbf{T}$ *be an abelian, (simply) transitive transformation group on* E. *Fix* $o \in E$ *and, for* $x \in E$, *let* τ_x *be the unique transformation in* $\mathbf{T}$ *which carries* o *to* x. *For every* $A \subseteq E$ *the pair*

$$\delta(X) = \bigcup_{x \in X} \tau_x(A), \quad \varepsilon(X) = \{h \in E \mid \tau_h(A) \subseteq X\}$$

forms a $\mathbf{T}$*-adjunction on* $\mathcal{P}(E)$. *Moreover, every* $\mathbf{T}$*-adjunction is of this form.*

If the translation group $\mathbf{T}$ is not abelian then neither $\delta_{\mathbf{A}}$ nor $\varepsilon_{\mathbf{A}}$ are $\mathbf{T}$-invariant in general. To achieve $\mathbf{T}$-invariance $\delta_{\mathbf{A}}$ is rewritten in the following way:

$$\delta_{\mathbf{A}}(X) = \bigcup_{x \in X} A(x), \tag{16}$$

where $A(x) = \{\tau x \mid \tau \in \mathbf{A}\} = \delta_{\mathbf{A}}(\{x\})$. This expression for $\delta_{\mathbf{A}}$ corresponds to the intuitive idea of moving a structuring element A over all points $x \in X$. The operator given by (16) is a dilation for every mapping $A : E \to \mathcal{P}(E)$. It can easily be shown that $\delta_{\mathbf{A}}$ is $\mathbf{T}$-invariant if and only if

$$A(\tau x) = \tau A(x) \tag{17}$$

for every $x \in E$ and $\tau \in \mathbf{T}$. Assume from now on that $\mathbf{T}$ is transitive on E. Fix an origin $o \in E$. Let Σ be the subgroup of $\mathbf{T}$ containing all τ which leave o invariant,

$$\Sigma = \{\tau \in \mathbf{T} \mid \tau o = o\}.$$

Σ is sometimes called the *stabilizer* of o. For example, if $\mathbf{T}$ is the group on $\mathbb{R}^2$ consisting of all rotations and translations, then the stabilizer of a point consists of all rotations around that point. If $\mathbf{T}$ is simply transitive then Σ contains only the identity mapping. Clearly, (17) implies that

$$A(o) = \Sigma A(o), \tag{18}$$

where $\Sigma X = \{\tau x \mid \tau \in \Sigma,\ x \in X\}$. It follows from (17) that

$$A(x) = \tau_x A(o),$$

where τ_x is a transformation carrying o to x.

Proposition 7. *Let* $\mathbf{T}$ *be a transitive transformation group on* E *and let* $o \in E$ *be fixed. The dilation* δ *given by*

$$\delta_A(X) = \bigcup_{x \in X} A(x), \tag{19}$$

where $A : E \to \mathcal{P}(E)$, *is* $\mathbf{T}$*-invariant if and only if* $A(\tau x) = \tau A(x)$ *for every* $\tau \in \mathbf{T}$ *and* $x \in E$. *This implies in particular that* $A(o) = \Sigma A(o)$. *Moreover, every* $\mathbf{T}$*-dilation is of this form under the given assumptions. The adjoint erosion is given by*

$$\varepsilon_A(X) = \{x \in E \mid A(x) \subseteq X\}. \tag{20}$$

Essentially, this result says that the only way to obtain T-adjunctions is as follows. Take a structuring element $A \subseteq E$ and define $A(x) = \tau_x \Sigma A$. Then the pair $(\varepsilon_A, \delta_A)$, where δ_A, ε_A are given by (19) and (20) respectively, defines a T-adjunction. In fact, in the introductory example of this section dealing with the translation-rotation invariant case, was reached a similar conclusion.

We conclude with some remarks about the corresponding openings. The opening $\delta_A \varepsilon_A$ given by

$$\delta_A \varepsilon_A(X) = \{\tau \Sigma A \mid \tau \in \mathbf{T} \text{ and } \tau \Sigma A \subseteq X\},$$

is T-invariant. However, the opening

$$\alpha_A(X) = \bigcup \{\tau A \mid \tau \in \mathbf{T} \text{ and } \tau A \subseteq X\}$$

is T-invariant as well. It is easy to see that

$$\alpha_A \geq \delta_A \varepsilon_A,$$

where the equality holds if and only if A is Σ-symmetric, that is $A = \Sigma A$. The difference between these two openings will be illustrated for the translation-rotation invariant case. Let $X \circ A$ be given by (5), with $A \subseteq \mathbb{R}^2$. Then

$$\delta_A \varepsilon_A(X) = X \circ \Big(\bigcup_{0 \leq \varphi \leq 2\pi} R_\varphi A \Big),$$

(observe that in this case $\Sigma A = \bigcup_{0 \leq \varphi \leq 2\pi} R_\varphi A$), and

$$\alpha_A(X) = \bigcup_{0 \leq \varphi \leq 2\pi} X \circ R_\varphi A.$$

For instance if A is a line segment with length l and centre 0, then α_A preserves all line segments with length $\geq l$ regardless of their orientation, whereas $\delta_A \varepsilon_A$ preserves only disks with diameter $\geq l$; see Fig. 5.

4.2 Distance-based Morphology

In this section, the construction of morphological operators on $\mathcal{P}(E)$, when E is equipped with a notion of distance, will be explained. Recall that a function $d : E \times E \to \mathbb{R}_+$ is called a *metric* or *distance function* if for $x, y, z \in E$

(D1) $d(x, y) = 0 \iff x = y$;
(D2) $d(x, y) = d(y, x)$;
(D3) $d(x, z) \leq d(x, y) + d(y, z)$.

The last property is called the *triangle inequality*. If d is a metric on E then we say that (E, d) is a *metric space*. The ball with radius r centred at x is given by

$$B(x, r) = \{y \in E \mid d(x, y) \leq r\}.$$

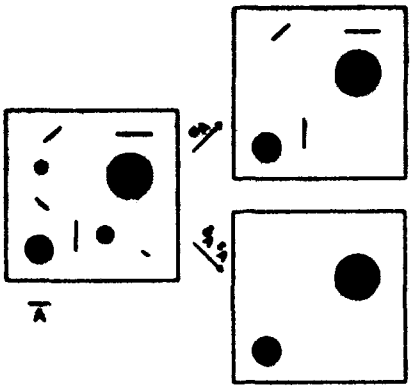


Fig. 5. The opening α_A preserves all line segments with length $\geq l$, whereas the opening $\delta_A \varepsilon_A$ preserves only disks with diameter $\geq l$.

One can define morphological operators on $\mathcal{P}(E)$ by taking these balls as structuring elements. More specifically, one can define a family of dilations δ^r, $r \geq 0$, as follows:

$$\delta^r(X) = \bigcup_{x \in X} B(x, r). \tag{21}$$

The adjoint erosion is given by

$$\varepsilon^r(X) = \{h \in E \mid B(h, r) \subseteq X\}. \tag{22}$$

One can easily show that

$$\delta^r \delta^s \leq \delta^{r+s}, \quad r, s \geq 0. \tag{23}$$

In fact, this relation is a consequence of the triangle inequality.

An important instance of a metric space is the so-called *Minkowski space*; this can be defined as a finite-dimensional normed vector space [23]. If $\|\cdot\|$ is a norm on E then $d(x, y) = \|x - y\|$ defines a metric. Besides the axioms (D1)–(D3), this metric satisfies

(D4) $d(x + h, y + h) = d(x, y)$,
(D5) $d(\lambda x, \lambda y) = |\lambda| d(x, y)$,

for $\lambda \in \mathbb{R}$, and $x, y, h \in E$. The best known example of a Minkowski space is of course the Euclidean space. In a Minkowski space the balls are of the special form $B(x, r) = (rB)_x$, the translate of rB along the vector x. Here B is the unit ball centred at the origin. This unit ball is compact, convex, contains 0 in its interior, and is reflection symmetric with respect to 0. Since the equality

$$rA \oplus sA = (r + s)A. \quad r, s \geq 0 \tag{24}$$

holds if and only if A is convex, the dilations δ^r given by (21) satisfy

$$\delta^r \delta^s = \delta^{r+s}, \quad r, s \geq 0, \tag{25}$$

if E is a Minkowski space. It follows that this semigroup relation is of great importance for the construction of granulometries.

It is worthwhile pointing out that the distance-based approach and the transformation-based approach are not complementary but rather alternative formulations of the same idea. To make this point clear consider the case $E = \mathbb{R}^d$. Let the structuring element $A \subseteq \mathbb{R}^d$ have the same properties as the unit ball in a Minkowski space. That is, A is compact, convex, contains 0 in the interior, and is reflection symmetric with respect to 0. Then there is a norm $\|\cdot\|_A$ on $\mathbb{R}^d$ for which the unit ball is A, namely,

$$\|x\|_A = \inf\{t > 0 \mid \frac{1}{t}x \in A\}.$$

In this case the dilation δ^r is given by

$$\delta^r(X) = X \oplus rA.$$

This identity says that the transformation-based dilation $X \to X \oplus rA$ corresponds to a distance-based dilation with a suitably chosen metric.

For a comprehensive discussion of the role of metric spaces in mathematical morphology refer to [11]; see also [31, Sections 1.6 and 2.4].

5 Granulometries

In certain image-analysis problems one is interested in the size distribution of the various objects in the scene. Many sizing techniques are based on the intuitive notion of a sieving process. Consider a binary image consisting of a finite number of isolated particles. Pass these particles through a stack of sieves with decreasing mesh widths and measure the number or the total volume of the particles remaining on a particular sieve. This results in a histogram which may be interpreted as a size distribution. Such an intuitive approach immediately raises a number of questions. A first objection is that objects are not classified according to their size but rather according to the property that they can or cannot pass a certain mesh opening. Furthermore, one has to decide which motions (translation, rotation, reflection) should be allowed in order to force a particle through a certain sieve. Another problem is that particles in an image may overlap and will be classified as one large particle. Thus we are led to the conclusion that the intuitive characterization of a size distribution as the outcome of a sieving process is too vague and too restricted.

Matheron [19] first realized that the concept of an opening in the morphological sense should underlie a formal definition of a size distribution. This approach is not only general but also attractive from a mathematical point of view. This section presents Matheron's definition of *granulometry* (meaning a tool to 'measure the grains'). In the first part of the section the discussion is restricted to the binary image space $\mathcal{P}(\mathbb{R}^d)$. At the end some recent results for grey-scale images will be discussed.

Definition 6. A granulometry on $\mathcal{P}(\mathbb{R}^d)$ is a one-parameter family of openings $\{\alpha_r \mid r > 0\}$ such that

$$\alpha_s \leq \alpha_r \text{ if } s \geq r. \tag{26}$$

It is obvious that property (26) is equivalent to

$$\alpha_r \alpha_s = \alpha_s \alpha_r = \alpha_s, \quad \text{if} \quad s \geq r. \tag{27}$$

If $\alpha_r(X)$ is defined as the union of all connected components of X whose volume is not less than Cr^d, where $C > 0$ is a constant, then $\{\alpha_r\}$ is a granulometry. This granulometry satisfies the additional properties

- α_r is translation invariant;
- $\alpha_r(X) = r\alpha_1(r^{-1}X)$, $r > 0$.

A granulometry on $\mathcal{P}(\mathbb{R}^d)$ with these properties is called a *Minkowski granulometry*. Note that the terminology 'Euclidean granulometry' is more common in the morphological literature [19]; however, this name will be reserved for a more specific example (see below).

If $\{\delta^r \mid r > 0\}$ is a family of dilations which have the semigroup property

$$\delta^r \delta^s = \delta^{r+s}, \quad r, s > 0, \tag{28}$$

then the openings $\alpha_r = \delta^r \varepsilon^r$ form a granulometry. Here ε^r is the erosion adjoint to δ^r. In particular, if d is a metric on $\mathbb{R}^d$ then the balls with radius r, $B(r) = B(0, r)$, satisfy

$$B(r) \oplus B(s) \subseteq B(r + s). \tag{29}$$

This is a direct consequence of the triangle inequality. If equality in (29) holds, that is,

$$B(r) \oplus B(s) = B(r + s), \tag{30}$$

then the dilations $\delta^r(X) = X \oplus B(r)$ have the semigroup property (28) and in this case the openings $\alpha_r(X) = X \circ B(r)$ form a granulometry. Since every opening involves only one structuring element, this granulometry is called a *structural granulometry*. To avoid confusion note the following two facts: (i) in general the family $B(r)$ satisfies only (29) and not (30); (ii) for $\alpha_r(X) = X \circ B(r)$ to be a granulometry, (30) is sufficient but by no means necessary. More specifically, $X \circ B(r)$ defines a granulometry if and only if $B(s)$ is $B(r)$-open for $s \geq r$. There are many families $B(\cdot)$ which satisfy this condition but not (30). However, this is no longer true if we assume in addition that the resulting granulometry is of Minkowski-type. Namely, suppose that α_r is of Minkowski-type and that $\alpha_1(X) = X \circ B$. Then $\alpha_r(X) = r\alpha_1(r^{-1}X) = r(r^{-1}X \circ B) = X \circ rB$. This shows that $B(r) = rB$ for some $B \subseteq \mathbb{R}^d$. In order that the openings $X \circ rB$ define a granulometry the structuring element must satisfy the condition that sB is rB-open for $s \geq r$, or equivalently

$$rB \text{ is } B\text{-open for } r \geq 1.$$

The following result is due to Matheron [19].

Theorem 9. *Let $B \subseteq \mathbb{R}^d$ be compact. Then rB is B-open for $r \geq 1$ if and only if B is convex.*

Note that for convex B the relation $rB \oplus sB = (r+s)B$ holds.

Corollary 10. *Assume that $B(r)$ is compact for $r \geq 0$. The openings $\alpha_r(X) = X \circ B(r)$ define a Minkowski granulometry if and only if $B(r) = rB$ for some B which is compact and convex.*

Remark. For completeness note that the openings

$$\alpha_r(X) = \bigcup_{s \geq r} X \circ sB$$

define a Minkowski granulometry for any structuring element $B \subseteq \mathbb{R}^d$. In applications the infinite union is undesirable. To get rid of it, it must be assumed that B is convex.

In the previous section it was shown that, for a compact, convex subset B of $\mathbb{R}^d$ which contains 0 in its interior and is reflection symmetric with respect to 0, there is a unique norm $\|\cdot\|_B$ on $\mathbb{R}^d$ such that the unit ball $\{x \in \mathbb{R}^d \mid \|x\|_B \leq 1\}$ coincides with B. The corresponding metric space is a Minkowski space; in other words, every Minkowski space is uniquely determined by a set B with the given properties.

To close the circle of arguments let δ^r be the dilation given by $\delta^r(X) = X \oplus rB$, where B is convex. Then δ^r obeys the semigroup property (28) and $\delta^r \varepsilon^r(X) = X \circ rB$.

As was note earlier, granulometries which are translation invariant and scale-compatible are usually called 'Euclidean granulometries'. However, the adjective 'Euclidean' is used for the granulometry given by the openings $X \circ rB$ where B is the unit ball in the Euclidean metric (obviously, this is also a Minkowski granulometry).

The granulometries mentioned in Corollary 10 and the subsequent remark are not the only ones which are of Minkowski-type. It is not hard to show that the union of an arbitrary collection of Minkowski granulometries is again a Minkowski granulometry. In fact, one can prove the following result.

Theorem 11. *Every Minkowski granulometry on $\mathcal{P}(\mathbb{R}^d)$ is of the form*

$$\alpha_r(X) = \bigcup_{B \in \mathcal{B}} \bigcup_{s \geq r} X \circ sB,$$

where $\mathcal{B}$ is an arbitrary collection of subsets of $\mathbb{R}^d$.

Consider the granulometry given by the openings $X \circ rB$, where B is convex. Suppose that Y is a component of X which contains at least one translate of rB, that is, $Y \circ rB \neq \emptyset$. In $\alpha_r(X) = X \circ rB$ the whole component Y will not be preserved, but only the subset $Y \circ rB$, which may be much smaller in practice. In some applications (e.g. in the case where one measures the total area of $\alpha_r(X)$)

one would prefer to retain the whole particle Y if its opening by rB is non-void. This can be achieved by defining the following modification:

$$\check{\alpha}_r(X) = \rho(X \circ rB; X), \tag{31}$$

where $\rho(X; M)$ is the reconstruction of X within the mask set M, that is, the union of all connected components of M which intersect X; see Fig. 6.

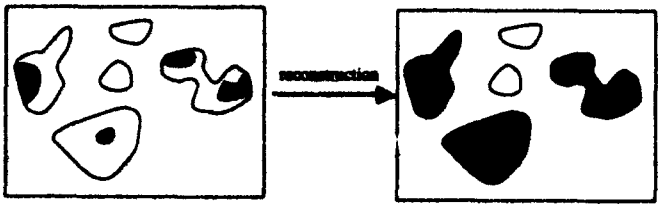


Fig. 6. Geodesic reconstruction.

In fact, a more general result can be established.

Proposition 12. *If $\{\alpha_r\}$ is a granulometry and if $\check{\alpha}_r$ is given by $\check{\alpha}_r(X) = \rho(\alpha_r(X); X)$, then $\{\check{\alpha}_r\}$ defines a granulometry as well.*

This section is concluded with some results for grey-scale granulometries recently obtained by Kraus *et al.* [15]. As in the binary case, a granulometry on $\mathrm{Fun}(\mathbb{R}^d)$ is defined as a collection of openings α_r on $\mathrm{Fun}(\mathbb{R}^d)$ which satisfy (26), or equivalently (27). It is easy to show that the extension of binary granulometries to grey-scale functions by thresholding yields grey-scale granulometries. In order to extend the notion of a Minkowski granulometry we must define translations and scalings for grey-scale functions. Concerning translations one may either restrict attention to translations in the domain (H-translations) or allow translations in the grey-level space as well (together called T-translations). A T-translation of F is given by $(F_h^v)(x) = F(x-h) + v$, where $h \in \mathbb{R}^d$ and $v \in \mathbb{R}$. Here only the second alternative will be considered. A similar choice has to be made for scalings. Either one can choose the so-called *umbral scaling*

$$(\lambda \cdot F)(x) = \lambda F(x/\lambda) \tag{32}$$

(here the adjective 'umbral' expresses that this operation scales the umbra, the points on and below the graph of the function) or the *spatial scaling*

$$(\lambda \cdot F)(x) = F(x/\lambda). \tag{33}$$

Both cases are illustrated in Fig. 7.

It is apparent that the choice between these two scalings has an enormous impact on the kind of shape information extracted by the granulometry. Here the spatial scaling, also referred to as *H-scaling*, is chosen.

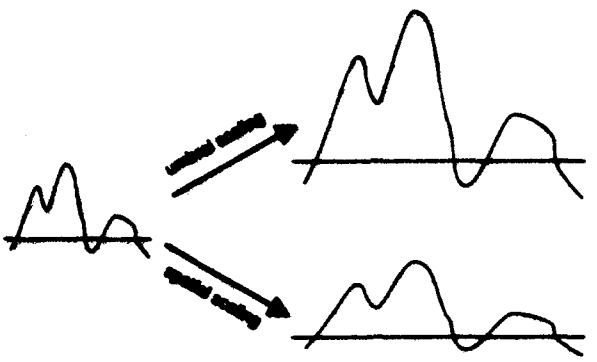


Fig. 7. Umbral scaling versus spatial scaling.

A granulometry α_r on $Fun(\mathbb{R}^d)$ is a *(T,H)-Minkowski granulometry* if it is invariant under T-translations (i.e., $\alpha_r(F_h^v) = [\alpha_r(F)]_h^v$) and compatible with H-scalings (i.e., $\alpha_r(F) = r \cdot \alpha_1(r^{-1} \cdot F)$) where '$\cdot$' denotes the scaling defined by (33). It is easy to show that, for every structuring function G, the openings

$$\alpha_r(F) = \bigvee_{s \geq r} F \circ (s \cdot G)$$

define a (T,H)-Minkowski granulometry. To eliminate the outer supremum the function G must satisfy

$$(r \cdot G) \circ G = r \cdot G \text{ for } r \geq 1. \tag{34}$$

If G is upper semi-continuous and has compact domain, then condition (34) is satisfied if and only if the domain of G is convex, and G is constant there. If B is the domain of G, then $\alpha_r(F) = F \circ (r \cdot G)$ is the flat extension of the binary Minkowski granulometry $X \to X \circ rB$. For more details refer to [15]. Related results can be found in [5].

6 Skeletons

Skeletonization algorithms have become enormously important in image processing. A first systematic study of skeletons was undertaken by Blum [2, 3] in the context of models for visual perception. However, the underlying ideas can be traced back to the work of Motzkin [20, 21, 32]. Blum introduced the prairie fire model to visualize his ideas. Think of the set X as a dry prairie and suppose that all the grass at the boundary of X is set on fire at the same moment. The resulting fires propagate at constant speed according to Huygen's principle. The *skeleton* $\Sigma(X)$ (or *medial axis* as Blum called it; later he introduced the term *symmetric axis* [3]) is the set of quench points where fire fronts coming from different directions extinguish each other. Shortly after the publication of Blum's first paper [2] there appeared an influential paper by Calabi and Hartnett [4] which carried the mathematical theory of skeletons much further. Lantuejoul [16]

was the first to write down an explicit expression for the skeleton using morphological transformations; see [30, Chapter XI] and [17]. Recently, Matheron [31, Chapter 11,12] derived a number of interesting topological results about the skeleton. Note, however, that both Serra and Matheron restrict themselves to open sets.

In the orignal work of Blum it was assumed that the fire spreads at a constant speed in all directions. In the formal discussion below this restriction will not be made; instead the speed is allowed to be non-isotropic.

Assume that $B \subseteq \mathbb{R}^d$ is a compact convex set and, moreover, that B contains more than one point. The *regular* and *singular* parts of a set X (with respect to B) are respectively defined by

$$\mathrm{reg}_B(X) = \bigcup_{r>0} X \circ rB, \tag{35}$$

$$\mathrm{sing}_B(X) = X \setminus \mathrm{reg}(X). \tag{36}$$

The first expression means that a point lies in the regular part of X if it is contained in $(rB)_h$ for some $r > 0$ which lies completely inside X. Apparently,

$$X = \mathrm{reg}_B(X) \cup \mathrm{sing}_B(X).$$

It is easy to show that $\mathrm{reg}_B(\cdot)$ defines a translation invariant opening on $\mathcal{P}(\mathbb{R}^d)$, and that $X^\circ \subseteq \mathrm{reg}_B(X)$. Here X° denotes the interior of X. Consequently, $\mathrm{sing}_B(X) \subseteq \partial X$, the boundary of X.

Definition 13. Assume that $(rB)_h$ is contained in X. Then $(rB)_h$ is a *maximal B-shape* in X if $(rB)_h \subseteq (r'B)_{h'} \subseteq X$ implies that $r' = r$ and $h' = h$.

Furthermore, define the rth B-skeleton subset by

$$\Sigma_{B,r}(X) = \mathrm{sing}_B(X \ominus rB). \tag{37}$$

Lemma 14. *$(rB)_h$ is a maximal B-shape in X if and only if $h \in \Sigma_{B,r}(X)$.*

Proof. Use the property that $h \in \Sigma_{B,r}(X)$ if and only if

(i) $h \in X \ominus rB$, and
(ii) $h \notin (X \ominus rB) \circ \varepsilon B$, for every $\varepsilon > 0$.

'If': assume that (i) and (ii) are satisfied. We show that $(rB)_h$ is a maximal B-shape in X. Suppose that $(rB)_h \subseteq [(r+\varepsilon)B]_h \subseteq X$. Then $h \in (\varepsilon B)_h \subseteq X \ominus rB$, which yields that $h \in (X \ominus rB) \circ \varepsilon B$, a contradiction.
'Only if': assume that $(rB)_h$ is a maximal B-shape in X. Then $h \in X \ominus rB$, i.e. (i) holds. Assume that $h \in (X \ominus rB) \circ \varepsilon B$ for some $\varepsilon > 0$. Then $h \in [(r+\varepsilon)B]_h \ominus rB = (\varepsilon B)_h \subseteq (X \ominus rB)$ which implies that $(rB)_h \subseteq [(r+\varepsilon)B]_h \subseteq X \circ rB \subseteq X$. But this means that $(rB)_h$ is not a maximal B-shape, a contradiction. □

It is obvious that $\Sigma_{B,r}(X) \cap \Sigma_{B,s}(X) = \emptyset$ if $r \neq s$. The *B-skeleton* $\Sigma_B(X)$ of X is defined as the (disjoint) union of all $\Sigma_{B,r}(X)$,

$$\Sigma_B(X) = \bigcup_{r \geq 0} \Sigma_{B,r}(X). \tag{38}$$

Theorem 15. *The B-skeleton of a set has empty interior.*

Proof. Assume $h \in \Sigma_{B,r}(X)$. We show that h does not lie in the interior of $\Sigma_{B,r}(X)$. If $r = 0$ then $h \in \partial X$ and therefore it cannot lie in the interior of $\Sigma_B(X)$. We assume that $r > 0$. The set $(rB)_h$ must intersect the boundary ∂X in a point $y \neq h$.

We restrict attention to the 2-dimensional case; for higher dimensions one can use similar arguments. Suppose first that $B^\circ \neq \emptyset$. If $0 \in B^\circ$ then the assertion is trivial; otherwise choose a point p so that $0 \in (B_p)^\circ$ and use (40) below. Suppose next that $B^\circ = \emptyset$. Then B is a line segment. It is obvious that a maximal line segment in X must intersect ∂X in at least two points. This proves the assertion.

We show that a point $k \in (h, y]$ cannot be contained in $\Sigma_B(X)$. Suppose namely that $k \in \Sigma_{B,s}(X)$. Then, by Lemma 14, $(sB)_k$ is a maximal B-shape inside X. There is a $\lambda \in [0, 1)$ so that $k = \lambda h + (1 - \lambda)y$. It is easy to check that $s \leq s_0$ where s_0 is the solution of $k + s_0/r(y - h) = y$. A straightforward calculation shows that $s_0 = \lambda r$ and we conclude that $s \leq \lambda r$. But it is not difficult to show that $(\lambda r B)_k \subseteq (rB)_k$ which contradicts our assumption that $(sB)_k$ is a maximal B-shape. □

The skeleton contains information about the shape of an object. In a sense it expresses how the shape of a set X relates to the shape of the structuring element B. Although it is tempting to make this assertion more concrete this matter is not pursued and consideration is limited to the example depicted in Fig. 8.

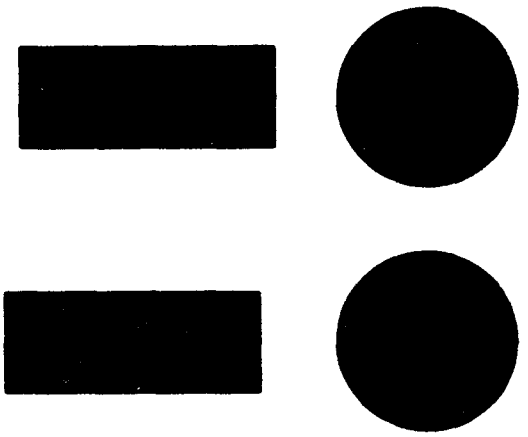


Fig. 8. The B-skeleton of a rectangle and a disk when B is a disk (top) and a square (bottom).

Here we compute the B-skeleton of a rectangle and a disk when B is a disk and a square respectively. Note that in the first case the speed of the prairie fire is uniform in all directions, whereas in the case where B is a square, the prairie fire has the highest speed ($\sim \sqrt{2}$) in the diagonal directions and the lowest speed

(~ 1) in the horizontal and vertical directions. The *quench function* q_B is defined as

$$q_B(X,h) = r \quad \text{if} \quad h \in \Sigma_{B,r}(X).$$

Note that $q_B(X,\cdot)$ has domain $\Sigma_B(X)$ for every set X.

If $h \in \Sigma_B(X)$ and $r = q_B(X,h)$ then $(rB)_h \subseteq X$. This yields the result that

$$\bigcup_{h \in \Sigma_B(X)} [q_B(X,h)B]_h = \bigcup_{r \geq 0} \Sigma_{B,r}(X) \oplus rB \subseteq X.$$

If equality holds then the original set X can be reconstructed from the data of the B-skeleton and the associated quench function. It is not difficult to see that this holds if every $x \in X$ is contained in at least one maximal B-shape.

Theorem 16. *Let X be closed. The equality*

$$\bigcup_{r \geq 0} \Sigma_{B,r}(X) \oplus rB = X \tag{39}$$

holds in each of the following cases:

(i) X is bounded;
(ii) B is the closed unit ball and X contains no half-spaces;
(iii) B is a finite line segment and X contains no half-lines with the same orientation;
(iv) B is a square and X contains no quarter-spaces with the same orientation.

The B-skeleton depends on the position of the origin (which we always assume to be contained within B). Moreover, a translation of B does not only induce a translation of the skeleton. In fact, one can show that a translation $B \to B_p$ has the following effect:

$$\Sigma_{B_p}(X) = \{h - q_B(X,h)p \mid h \in \Sigma_B(X)\}$$
$$q_{B_p}(X, h - q_B(X,h)p) = q_B(X,h).$$

The erosion $X \ominus tB$ obeys the expression:

$$\begin{aligned} \Sigma_r(X \ominus tB) &= \mathsf{sing}((X \ominus tB) \ominus rB) \\ &= \mathsf{sing}(X \ominus (r+t)B) \\ &= \Sigma_{r+t}(X). \end{aligned}$$

This yields the result that

$$\Sigma_r(X \ominus tB) = \Sigma_{r+t}(X) \quad \text{and} \quad X \ominus tB = \bigcup_{r \geq 0} \Sigma_{r+t}(X) \oplus rB,$$

and hence that

$$\Sigma(X \ominus tB) = \bigcup_{r \geq t} \Sigma_r(X).$$

Consider the opening $X \circ tB$. If $r \geq t$ then

$$\begin{aligned}(X \circ tB) \ominus rB &= \Big(((X \ominus tB) \oplus tB) \ominus tB\Big) \ominus (r-t)B \\ &= (X \ominus tB) \ominus (r-t)B \\ &= X \ominus rB.\end{aligned}$$

From this identity one derives that

$$\Sigma_r(X \circ tB) = \Sigma_r(X) \text{ if } r \geq t.$$

Unfortunately it is not possible to make general statements about $\Sigma_r(X \circ tB)$ for $r < t$. Even though $X \circ tB$ is a union of B-shapes of radius $\geq t$, it is possible that $X \circ tB$ contains maximal B-shapes of radius $\leq t$. However, it is possible to reconstruct $X \circ tB$ from the rth skeleton subsets $\Sigma_r(X)$, $r \geq t$. For that purpose the expression for $X \ominus tB$ derived above is used. One obtains

$$\begin{aligned}X \circ tB &= (X \ominus tB) \oplus tB \\ &= \Big[\bigcup_{r \geq 0} \Sigma_{r+t}(X) \oplus rB\Big] \oplus tB \\ &= \bigcup_{r \geq 0} \Sigma_{r+t}(X) \oplus (r+t)B;\end{aligned}$$

that is,

$$X \circ tB = \bigcup_{r \geq t} \Sigma_r(X) \oplus rB. \tag{40}$$

This observation suggests a family of morphological transformations, called *quench function transformations* by Serra [30, Exercise XI-I.8]. Let $f : \mathbb{R}_+ \to \mathbb{R}$; define the transformation ψ_f on $\mathcal{P}(\mathbb{R}^d)$ as

$$\psi_f(X) = \bigcup_{r \geq 0} \Sigma_{B,r}(X) \oplus f(r)B,$$

where $f(r)B = \emptyset$ if $f(r) < 0$. Dilation, erosion and opening by tB are examples of such transformations.

Finally, it is pointed out that there is a relation between the B-skeleton and the Minkowski granulometry $\alpha_r(X) = X \circ rB$ discussed in the previous section. If B is convex then $X \circ sB \subseteq X \circ rB$ if $s \geq r$. If $X \circ rB$ is substantially larger than $X \circ (r+\varepsilon)B$, where $\varepsilon > 0$ is small, then it may be concluded that X has components with B-size r; here the phrase 'Y has B-size r' means that Y contains a B-shape with radius r but not with radius $> r$.

Theorem 17. *For any set X,*

$$\Sigma_{B,r}(X) = \emptyset \iff X \circ rB = \bigcup_{\varepsilon > 0} X \circ (r+\varepsilon)B$$

if $r \geq 0$.

Proof. '$\Rightarrow$': assume that $\Sigma_{B,r}(X) = \mathrm{sing}_B(X \ominus rB) = \emptyset$. This means that $X \ominus rB = \bigcup_{\varepsilon>0}(X \ominus rB) \circ \varepsilon B$. Then

$$\begin{aligned} X \circ rB &= (X \ominus rB) \oplus rB \\ &= [\bigcup_{\varepsilon>0}(X \ominus rB) \circ \varepsilon B] \oplus rB \\ &= \bigcup_{\varepsilon>0}[((X \ominus rB) \circ \varepsilon B) \oplus rB] \\ &= \bigcup_{\varepsilon>0}((X \ominus rB) \ominus \varepsilon B) \oplus \varepsilon B \oplus rB \\ &= \bigcup_{\varepsilon>0} X \circ (r+\varepsilon)B. \end{aligned}$$

'$\Leftarrow$': assume that $X \circ rB = \bigcup_{\varepsilon>0} X \circ (r+\varepsilon)B$. We get

$$\begin{aligned} X \ominus rB &= (X \circ rB) \ominus rB \\ &= [\bigcup_{\varepsilon>0} X \circ (r+\varepsilon)B] \ominus rB \\ &= [\bigcup_{\varepsilon>0}(X \ominus (r+\varepsilon)B) \oplus \varepsilon B] \oplus rB \ominus rB \\ &\subseteq [\bigcup_{\varepsilon>0}(X \ominus (r+\varepsilon)B) \oplus \varepsilon B] \\ &= \bigcup_{\varepsilon>0}(X \ominus rB) \circ \varepsilon B. \end{aligned}$$

The reverse inclusion is trivial. Hence $\Sigma_{B,r}(X) = \mathrm{sing}_B(X \ominus rB) = \emptyset$. □

7 Morphological Shape Decomposition

An important problem in image analysis is the decomposition of an object into simpler parts. Such decompositions can, for example, be used for object recognition tasks. In the literature one can find a multitude of techniques for shape decomposition. In this section two approaches based on morphological openings are briefly described. A first approach, described in Sect. 7.1, was given by Pitas and Venetsanopoulos [22]. In their approach the simplest possible shape is a disk B (or any other convex structuring element). Starting with an object X one finds the largest radius r_1 for which $X \circ r_1 B \neq \emptyset$ and defines

$$X_1 = X \circ r_1 B,$$

the first-order approximation in the shape decomposition. Subsequently, one computes the largest radius r_2 for which

$$(X \setminus X_1) \circ r_2 B \neq \emptyset.$$

The second-order approximation is given by

$$X_2 = X_1 \cup (X \setminus X_1) \circ r_2 B.$$

Thus this decomposition algorithm is described by the recursion formula

$$\begin{cases} X_0 = \emptyset \\ X_{t+1} = X_t \cup (X \setminus X_t) \circ r_{t+1} B, \ t \geq 0, \end{cases}$$

where r_t is the radius of the maximal inscribable ball rB in $X \setminus X_{t-1}$. Note that $(X \setminus X_{t-1}) \circ r_t B$ is of the form $L \oplus r_t B$, where L is the part of the skeleton of $X \setminus X_{t-1}$ where the quench function is maximal. A set of the form $L \oplus rB$, where L is an arc, is called the *Blum ribbon* [22] An example is depicted in Fig. 9, where X is a binary image on the hexagonal grid. The structuring element is the elementary hexagon consisting of 7 points. This figure depicts respectively the original image X and its decompositions X_1 ($r_1 = 29$), X_6 ($r_6 = 17$), X_{11} ($r_{11} = 11$), X_{15} ($r_{15} = 7$), X_{19} ($r_{19} = 3$). In this example 21 iterations are required to recover the original image.

Recently Ronse [27] has developed a very general theory for morphological shape description and decomposition; see also [29]. His theory, which applies to a large class of partially ordered sets, is based upon notions as *toggles of openings, choice functions* and *open-condensations*. The latter concept will appear again. Besides the Pitas–Venetsanopoulos decomposition, which yields a union of disjoint components, Ronse also studied a decomposition in which the building components are not necessarily disjoint.

Section 7.1 describes the Pitas–Venetsanopoulos algorithm, and Sect. 7.2 the algorithm due to Ronse. Both sections are based on Ronse's work [27], in particular on his notion of open-condensation; this notion will be discussed below. Throughout the remainder of this section only the binary case is considered. However, the results can easily be extended to grey-scale images.

Definition 18. An operator ψ on $\mathcal{P}(E)$ is said to be *condensing* if $X \subseteq Y \subseteq Z$ and $\psi(X) = \psi(Z)$ implies that $\psi(X) = \psi(Y) = \psi(Z)$. If ψ is anti-extensive, idempotent and condensing then it is called an *open-condensation*.

The condensation property is slightly more general than monotonicity: every increasing or decreasing operator is condensing. Similarly the concept of open-condensation extends that of an opening; in particular, every opening is an open-condensation. Refer to [27] for a number of basic results.

7.1 Pitas–Venetsanopoulos Decomposition

This section describes a morphological decomposition algorithm originally due to Pitas and Venetsanopoulos [22]. The present treatment, however, is based on the work of Ronse [27].

Suppose a finite collection of openings

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$$

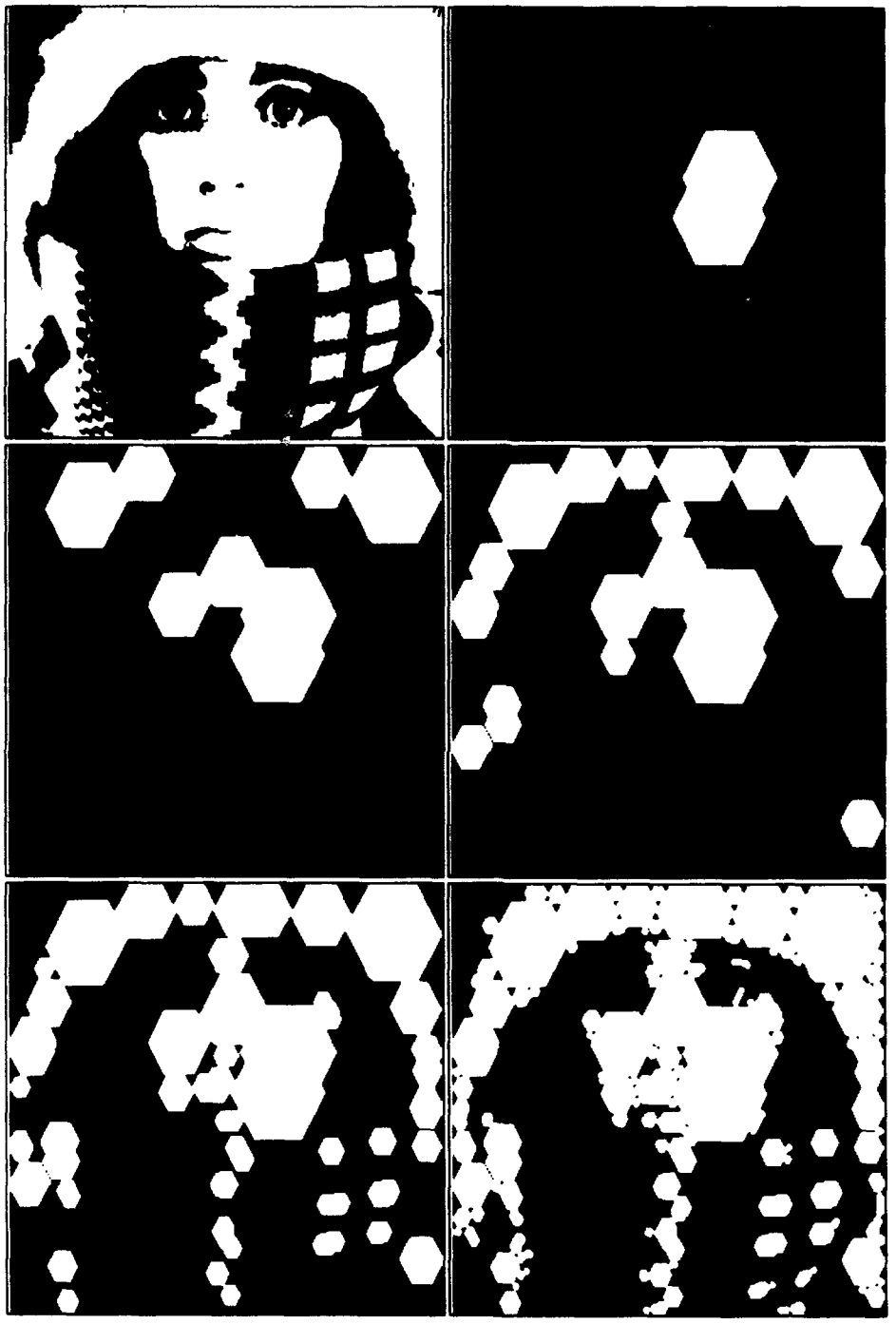


Fig. 9. Pitas–Venetsanopoulos decomposition.

is given. In the original theory of Pitas and Venetsanopoulos these openings correspond to openings with balls with decreasing radius.

Let $X \subseteq E$ and define $i(X)$ as the smallest index i such that $\alpha_i(X) \neq \emptyset$; put $i(X) = n+1$ if such index does not exist. Define

$$\theta(X) = \alpha_{i(X)}(X)$$

with $\alpha_{n+1}(X) = \emptyset$. In general, θ is not increasing. However, the following result can be proved.

Lemma 19. *θ is an open-condensation.*

Proof. It is obvious that θ is anti-extensive and idempotent. We show that θ has the condensation property. Suppose $X \subseteq Y \subseteq Z$ and $\theta(X) = \theta(Z)$. It is obvious that $i(X) \geq i(Z)$. On the other hand,

$$\alpha_{i(Z)}(Z) = \alpha_{i(Z)}(\alpha_{i(Z)}(Z)) = \alpha_{i(Z)}(\alpha_{i(X)}(X)) \subseteq \alpha_{i(Z)}(X),$$

which gives that $i(X) \leq i(Z)$. Therefore equality holds. But then also $i(Y) = i(X)$ and the result follows. □

Consider the following decomposition algorithm:

$$\begin{cases} X_0 = \emptyset \\ X_{t+1} = X_t \cup \theta(X \setminus X_t), \ t \geq 0. \end{cases}$$

It is not difficult to see that the corresponding index sequence $i(X \setminus X_t)$ is increasing. As soon as $i(X \setminus X_t)$ reaches the value $n+1$ the recursion is stopped. Define the operator γ_t as

$$\gamma_t(X) = X_t.$$

Proposition 20. *The operator γ_t is an open-condensation.*

To prove this the following result is required.

Lemma 21. *Assume that ψ is an open-condensation. Then γ defined by $\gamma(X) = \psi(X) \cup \theta(X \setminus \psi(X))$ is an open-condensation as well.*

Proof. It is evident that γ is anti-extensive. We show that γ is idempotent. Since $\psi(X) \subseteq \gamma(X) \subseteq X$, $\psi(\psi(X)) = \psi(X)$ and ψ is an open-condensation, we obtain that $\psi\gamma(X) = \psi(X)$. Therefore,

$$\gamma^2(X) = \psi(X) \cup \theta(\gamma(X) \setminus \psi(X)).$$

Use $\gamma(X) \setminus \psi(X) = \theta(X \setminus \psi(X)) \setminus \psi(X) = \theta(X \setminus \psi(X))$ and obtain $\gamma^2(X) = \gamma(X)$.

Finally we show that γ is condensing. Suppose that $X \subseteq Y \subseteq Z$ and $\gamma(X) = \gamma(Z)$. Since $\psi\gamma = \psi$ we get $\psi(X) = \psi(Z)$. Since ψ is an open-condensation, this means that $\psi(X) = \psi(Y) = \psi(Z)$. Then

$$\theta(X \setminus \psi(X)) = \gamma(X) \setminus \psi(X) = \gamma(Z) \setminus \psi(Z) = \theta(Z \setminus \psi(Z))$$

and

$$X \setminus \psi(X) \subseteq Y \setminus \psi(Y) \subseteq Z \setminus \psi(Z).$$

But θ is an open-condensation (cf. Lemma 19) and we may conclude that $\theta(X \setminus \psi(X)) = \theta(Y \setminus \psi(Y))$. Therefore $\gamma(X) = \gamma(Y)$. □

Proposition 20 follows from this result by induction. Namely $\gamma_0(X) = \emptyset$ defines an open-condensation. Suppose that γ_t is an open-condensation. Then $\gamma_{t+1}(X) = \gamma_t(X) \cup \theta(X \setminus \gamma_t(X))$, and by Lemma 21 this is an open-condensation as well.

Note that the above decomposition is quite different from the 'classical' morphological multiscale description given by the sequence $X_t = \alpha_t(X)$.

7.2 Ronse Decomposition

In this section a variant of the decomposition algorithm given by Ronse in [27] is discussed. An application can be found in [29]. The underlying idea is captured by the following example. Suppose that one has a collection of structuring elements $B_1, B_2, \cdots, B_n$, and that one wants to approximate an object X by the openings $X \circ B_1, X \circ B_2, \cdots X \circ B_n$. It depends on the shape of X which B_i suits best. Consider the convex polygon X which has 8 equal edges of length l as depicted in Fig. 10, let D be a disk and S a square both with area A. If $l^2 \geq A$ then $X_S = X$ whereas X_D is a strict subset of X. Therefore X_S is a better approximation than X_D for such values of A. This changes if A is increased. In particular, if $A > 4l^2$ then $X_S = \emptyset$; however $X_D = \emptyset$ if and only if $A > \frac{\pi}{4}(3+2\sqrt{2})l^2$. Inside X always fits a disk with radius $l(2\tan\frac{\pi}{8})^{-1} = \frac{1}{2}l(\sqrt{2}+1)$. Note that $\frac{\pi}{4}(3+2\sqrt{2}) > 4$.

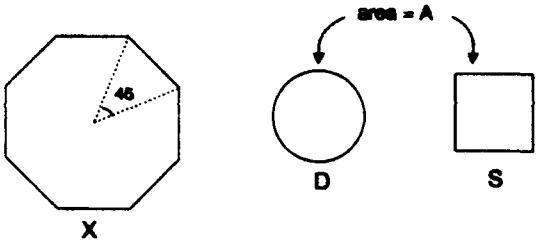


Fig. 10. A convex polygon which has 8 equal edges, and structuring elements D and S both with area A.

How can such a simple observation be put into a formal mathematical framework? For that purpose Ronse [27] introduced the general notion of a choice function. In the discussion hereafter an attempt will be made to be more specific and it will be assumed that there is a distance function on the object space. This distance function is used to choose, at every step, between the different alternatives.

Assumption. Let $\mathcal{L} \subseteq \mathcal{P}(E)$ and let D be a function which maps a pair $X, Y \in \mathcal{L}$ with $X \subseteq Y$ onto a positive real; assume moreover that D has the following properties:

- $D(X,Y) = 0$ iff $X = Y$;
- $D(X,Y) \leq D(X,Y')$ if $X \subseteq Y \subseteq Y'$.

Assume that the object space $\mathcal{L}$ is closed under finite unions. As a concrete example consider the case where $\mathcal{L}$ is the space of compact sets in $\mathbb{R}^d$ and D is the Hausdorff metric. Let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be a finite collection of openings which map $\mathcal{L}$ into $\mathcal{L}$. Furthermore, let ψ be an open-condensation. Define, for $X \subseteq Y$, $i(X, Y)$ as the smallest index i such that $D(X, X \cup \alpha_i(Y))$ is maximal. If, however, this expression is 0 for every i (which is the case iff $\alpha_i(Y) \subseteq X$ for every i) then $i(X, Y) = n + 1$. Furthermore, define $\alpha_{n+1}(X) = \emptyset$ for every X.

Lemma 22. *Given an open-condensation ψ, the operator γ given by*

$$\gamma(X) = \left(\psi \vee \alpha_{i(\psi(X),X)}\right)(X)$$

is an open-condensation as well.

Proof. Assume that $\gamma(X) \subseteq Y \subseteq X$. We show that $\gamma(X) = \gamma(Y)$. It is clear that $\psi(X) \subseteq Y \subseteq X$. Since ψ is an open-condensation we get that $\psi(X) = \psi(Y)$. Let $i_0 = i(\psi(X), X)$, then $\alpha_{i_0}(X) \subseteq Y \subseteq X$. The fact that α_{i_0} is an opening implies that

$$\alpha_{i_0}(X) \subseteq \alpha_{i_0}(Y) \subseteq \alpha_{i_0}(X),$$

and therefore $\alpha_{i_0}(X) = \alpha_{i_0}(Y)$. Since $\alpha_j(Y) \subseteq \alpha_j(X)$ for all j we conclude that

$$D(\psi(Y), \psi(Y) \cup \alpha_j(Y)) = D(\psi(X), \psi(X) \cup \alpha_j(Y))$$

is maximal if $j = i_0$. This shows that $i(\psi(Y), Y) = i_0$, and hence that $\gamma(X) = \gamma(Y)$.

It is obvious that γ is anti-extensive. Thus, substituting $Y = \gamma(X)$, we find that γ is idempotent. We show that γ is condensing. Assume $X \subseteq Y \subseteq Z$ and $\gamma(X) = \gamma(Z)$. Then $\gamma(Z) = \gamma(X) \subseteq X \subseteq Y \subseteq Z$. From the considerations above it follows that $\gamma(Y) = \gamma(Z)$. This concludes the proof. □

Consider the algorithm:

$$\begin{cases} X_0 = \emptyset, \\ X_{t+1} = X_t \cup \alpha_{i(X_t,X)}(X). \end{cases}$$

It is obvious that $X_0 \subseteq X_1 \subseteq \cdots \subseteq X$. In fact, this sequence is strictly increasing until it reaches the final result $\bigcup_{i=1}^{n} \alpha_i(X)$ (after at most n iterations). Putting

$$\gamma_t(X) = X_t,$$

the following conclusion can be derived from Lemma 22.

Proposition 23. *γ_t is an open-condensation for every $t \geq 0$.*

The approach outlined above is less general than the original approach of Ronse [27]. The algorithm described by Ronse and Macq [29] uses structuring elements of different scales. It first checks if there is a structuring element in the largest size scale which yields an increment in the approximation; within that scale class it chooses the shape giving the largest increment. If there do no longer exist structuring elements which give increments, only then is the scale decreased and structuring elements in the next, smaller size class are considered.

Finally, observe that in both approaches the decomposition operators γ_t are translation invariant if every α_i is translation invariant.

8 Final Remarks

Mathematical morphology is a branch of image analysis particularly suited to shape description. In fact, the concept of a structuring element enables one to infer shape-related information from a scene. It has been explained that there are essentially two different ways to conceive of a structuring element. Firstly, it can be regarded as a subset of a certain transformation group. This point of view is quite flexible in that one may choose those transformations which are best suited to the given application. Alternatively, one can use the notion of distance to define structuring elements. In this approach, the notion of convexity plays an important role.

This paper has presented a bird's eye view of the different morphology-based methods for shape description, including granulometric analyses, skeletonization techniques, and morphological decomposition algorithms.

A concept very closely related to the morphological granulometry is the so-called *pattern spectrum* introduced by Maragos [18]. For a compact set $X \subseteq \mathbb{R}^d$ the pattern spectrum relative to a convex structuring element $B \subseteq \mathbb{R}^d$ is defined as

$$PS_X(r, B) = -\frac{d}{dr}\text{Area}(X \circ rB), \quad r \geq 0,$$
$$PS_X(-r, B) = \frac{d}{dr}\text{Area}(X \bullet rB), \quad r > 0.$$

For discrete images one defines $nB = B \oplus \cdots \oplus B$ (n times) and

$$PS_X(n, B) = \text{Area}(X \circ nB \setminus X \circ (n+1)B), \quad n \geq 0,$$
$$PS_X(-n, B) = \text{Area}(X \bullet nB \setminus X \bullet (n-1)B), \quad n \geq 1.$$

Analogous definitions can be given for grey-scale images. Maragos uses the pattern spectrum as a shape-size descriptor. Furthermore he points out connections between the pattern spectrum and the skeleton transform (cf. Theorem 17).

Van den Boomgaard and Smeulders [33, 34] have initiated a theory which can be regarded as the morphological analogue of the Gaussian scale space. The basic idea is to dilate (or erode) an image with a parameterized family of quadratic structuring elements. They show that resulting images satisfy a nonlinear PDE related to Burger's equation.

References

1. Birkhoff, G. (1984). Lattice Theory, 3rd edition, Am. Math. Soc. Coll. Publ., 25, Providence, RI.
2. Blum, H. (1967). A transformation for extracting new descriptors of shape. In: Wathen-Dunn, W. (ed.), Models for the Perception of Speech and Visual Forms, MIT Press.
3. Blum, H. (1973). Biological shape and visual sciences (Part I), J. Theor. Biol. 38, pp. 205–287.

4. Calabi, L., Hartnett, W.E. (1968). Shape recognition, prairie fires, convex deficiencies and skeletons, Am. Math. Monthly 75, pp. 335–342.
5. Dougherty, E.R. (1992). Euclidean gray-scale granulometries: representation and umbra inducement, J. Math. Imaging Vision 1, pp. 7–21.
6. Ghosh, P.K. (1992). On negative shape, this volume, pp. 225–248.
7. Giardina, C.R., Dougherty, E.R. (1988). Morphological Methods in Image and Signal Processing, Prentice Hall, Englewood Cliffs, NJ.
8. Heijmans, H.J.A.M. (1987). Mathematical morphology: an algebraic approach, CWI Newsletter 14, pp. 7–27.
9. Heijmans, H.J.A.M. (1991). Theoretical aspects of gray-scale morphology, IEEE Trans. Pattern Anal. Mach. Intell. 13, pp. 568–582.
10. Heijmans, H.J.A.M. (1993). A note on the umbra transform in mathematical morphology, Pattern Recognition Letters, to appear.
11. Heijmans, H.J.A.M. (1993). Morphological Image Operators, Academic Press, in preparation.
12. Heijmans, H.J.A.M., Nacken, P. Toet, A., Vincent, L. (1992). Graph morphology, J. Visual Comm. Image Repr. 3, pp. 24–38.
13. Heijmans, H.J.A.M., Ronse, C. (1990). The algebraic basis of mathematical morphology. Part I: dilations and erosions, Comp. Vision Graph. Image Process. 50, pp. 245–295.
14. Heijmans, H.J.A.M., Vincent, L. (1992). Graph morphology in image analysis. In: Dougherty, E.R. (ed.), Mathematical Morphology in Image Processing, Marcel Dekker, New York, Chapter 6, pp. 171–203.
15. Kraus, E.J., Heijmans, H.J.A.M., Dougherty, E.R. (1992). Gray-scale granulometries compatible with spatial scalings, CWI Report BS-R9212, Amsterdam.
16. Lantuejoul, C.(1980). Skeletonization in quantitative metallography. In: Haralick, R.M., Simon, J.C. (eds), Issues in Image Processing, Sijthoff and Noordhof, Groningen, pp. 107-135.
17. Maragos, P. (1986). Morphological skeleton representation and coding of binary images, IEEE Trans. Acoustics, Speech and Signal Process. 34, pp. 1228–1244.
18. Maragos, P. (1989). Pattern spectrum and multiscale shape representation, IEEE Trans. Pattern Anal. Mach. Intell. 11, pp. 701–716.
19. Matheron, G. (1975). Random Sets and Integral Geometry, J. Wiley & Sons, New York.
20. Motzkin, T.S. (1935). Sur quelques propriétés caractéristique des ensembles convexes, Rend. Reale Acad. Lincei, Classe Sci. Fis., Mat. Nat. 21, pp. 562–567.
21. Motzkin, T.S. (1935). Sur quelques propriétés caractéristique des ensembles bornés non convexes, Rend. Reale Acad. Lincei, Classe Sci. Fis., Mat. Nat. 21, pp. 773–779.
22. Pitas, I., Venetsanopoulos, A.N. (1990). Morphological shape decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 12, pp. 38–45.
23. Rinow, W. (1961). Die Innere Geometrie der Metrischen Räume, Springer, Berlin.
24. Roerdink, J.B.T.M. (1992). Mathematical morphology with non-commutatitive symmetry groups. In: Dougherty, E.R. (ed.), Mathematical Morphology in Image Processing, Marcel Dekker, New York, Chapter 7, pp. 205–254.
25. Roerdink, J.B.T.M. (1993). On the construction of translation and rotation invariant morphological operators, Haralick, R.M. (ed.), Morphology: Theory and Hardware, Oxford University Press, to appear.
26. Roerdink, J.B.T.M. (1993). Manifold shape: from differential geometry to mathematical morphology, this volume, pp. 209–223.

27. Ronse, C. (1992). Toggles of openings, and a new family of idempotent operators on partially ordered sets. In: Applicable Algebra in Engineering, Communication and Computing, to appear.
28. Ronse, C., Heijmans, H.J.A.M. (1991). The algebraic basis of mathematical morphology. Part II: openings and closings, Comp. Vision Graph. Image Process: Im. Underst. 54, pp. 74–97.
29. Ronse, C., Macq, B. (1991). Morphological shape and region description, Signal Processing 25, pp. 91–105.
30. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.
31. J. Serra, (ed) (1988). Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances, Academic Press, London.
32. Valentine, F.A. (1964). Convex Sets, McGraw-Hill, New York.
33. van den Boomgaard, R. (1992). Mathematical Morphology: Extensions towards Computer Vision, PhD Thesis, University of Amsterdam.
34. van den Boomgaard, R., Smeulders, A.W.M. (1992). Towards a morphological scale space theory, this volume, pp. 631–640.
35. Vincent, L. (1990). Algorithmes Morphologiques a Base de Files d'Attente et de Lacets. Extension aux Graphes, PhD Thesis, Ecole Nationale Superieure des Mines de Paris, Fontainebleau.

On Information Contained in the Erosion Curve

Juliette Mattioli[1] *and Michel Schmitt*[2]

[1] L.C.R, Thomson-CSF, Domaine de Corbeville, 91404 Orsay-Cedex, France
CEREMADE, Université Paris Dauphine, 75775 Paris-Cedex, France
[2] L.C.R, Thomson-CSF, Domaine de Corbeville, 91404 Orsay-Cedex, France

Abstract. An erosion curve can be associated with any binary planar shape. This curve is the function which maps a given radius to the area of the shape eroded by a sphere with this radius. Note the analogy with the approach whereby a shape is quantified by granulometry. Under some regularity conditions the erosion curve of a given shape can be expressed as an integral of the quench function along the skeleton of this shape. This paper describes the relationship between sets with the same erosion curve. It is shown that this curve is not affected if the arcs of the skeleton are bent: the erosion curve quantifies "soft" shapes. In the generic case, there are five possible cases of behaviour of the second derivative of the erosion curve: each case corresponds to a different behaviour of the skeleton and the associated quench function. Finally, it is shown how to reconstruct the family of shapes with a given erosion curve.

Keywords: mathematical morphology, erosion curve, granulometry, skeleton, quench function, shape index, isoperimetrical deficiency index, elongation index, concavity index, stretching index, spectral function.

1 Introduction

Shape description is a very important problem in pattern analysis. It provides descriptions of objects according to their shape, which can be used for pattern recognition.

The principle [7, 3, 27] is to synthesize the information contained in a shape into a curve, called the "erosion curve". The *erosion curve* of a subset X of $\mathbb{R}^2$ is defined by $\Psi_X(r) = A(X \ominus rB)$, for $r \geq 0$. Here $A(X)$ stands for the area of X and B is the unit ball of $\mathbb{R}^2$. The erosion curve is translation and rotation invariant, and gives global information about the shape.

This paper deals with the following questions:

- Having an erosion curve, what can be said about the original shape? How can one reconstruct a shape from only knowledge of the erosion curve?
- What information is lost during the computation of the erosion curve?

After recalling some basic notions of mathematical morphology, we give properties of the skeleton of a compact planar shape and its links with erosions by disks. Then, we study the erosion curve Ψ_X and show that its second derivative gives information about shape and characterizes classes of shape. Finally, we present a method for building a shape from knowledge of only Ψ_X.

2 Notions of Mathematical Morphology and Shape Index

2.1 Notions of Mathematical Morphology

Morphological shape analysis uses the idea of Boolean operations to make comparisons between an arbitrary reference shape called the *structuring element* and the image.

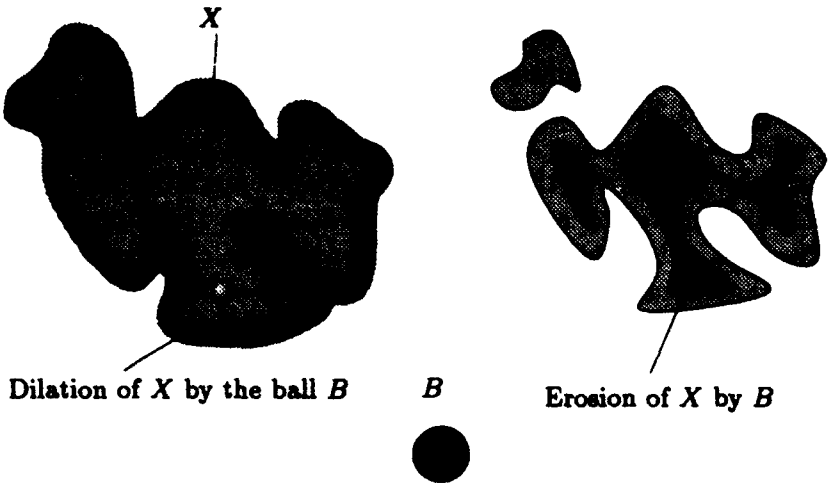


Fig. 1. Dilation and erosion by a disk B

If we consider an isotropic structuring element, a ball of radius r centreed at the origin, denoted by rB, the *eroded set* of X with respect to rB is given by: $X \ominus rB = \{x,\ (rB)_x \subset X\}$, where B_x stands for the translation of B at point x.

According to the usual duality principle with respect to the complementation, the *dilation* is expressed by $X \oplus rB = \{x,\ (rB)_x \bigcap X \neq \emptyset\}$, and we have $(X^c \ominus rB)^c = X \oplus rB$.

The *opening* of X by rB is the domain swept out by all translates of rB which are included in X: $X_{rB} = \bigcup\{rB_x,\ rB_x \subset X\}$, and the *closing* of X by rB, $X^{rB} = ((X^c)_{rB})^c$.

2.2 Shape Indices

Since the shape index successfully gives information on the shape, it must be translation, rotation, and scale invariant. We recall four shape indices:

1. *The isoperimetrical deficiency index* of a planar compact set X is defined by [20]:

$$FP(X) = 1 - \frac{4\pi A(X)}{P(X)^2} \ , \tag{1}$$

where $A(X)$ stands for the area of X and $P(X)$ is the perimeter of X. It is a very popular index which ranges from 0 (for a disk) to 1 (for an object with null surface area); indeed this index is very sensitive to noise on the perimeter.

2. *The elongation index*, that is the lengthening index related to the inscribed radius, is defined by:

$$FE(X) = \frac{\pi R(X)^2}{A(X)} \ , \tag{2}$$

where $R(X)$ is the maximal radius of the inscribed disks. We can compute it by erosion:
$R = \max\{r,\ X \ominus rB \neq \emptyset\}$, where B is the unit disk. This index is maximal and equal to 1 for a disk and it is a good characteristic for elongated sets.

3. *The concavity index in surface* is defined by [3]:

$$FC(X) = \frac{A(X)}{A(\overline{co}(X))} \ , \tag{3}$$

where $\overline{co}(X)$ is the Euclidean convex hull of X. This ratio is equal to 1 for a convex set and it gives information on the crevices of X.

4. *The stretching index*, that is the geodesic lengthening index, is defined by [6]:

$$FS(X) = \frac{\pi L(X)^2}{4A(X)} \ , \tag{4}$$

where $L(X)$ is the geodesic diameter of X, that is the length of the longest geodesic path included in X. This index ranges from 1 (for a disk) to infinity (for an object with null area) and is more robust than the isoperimetric deficit index.

2.3 Granulometry and Related Transformations

Shape indices give global measures of X. We need a deeper analysis of crevice size repartition. The "granulometry" principle [7] is to transform the binary planar shape X into a curve which is translation, rotation, and scale invariant by a family of morphological transformations (ϕ_r) depending on r, which typically is the size of the structuring element. Then, we build a map Ψ_X which associates r with the Lebesgue measure of $\phi_r(X)$.

For example, if $\phi_r(X) = X \ominus rB$ for $X \subset \mathbb{R}^2$, then $\Psi_X(r) = A(X \ominus rB)$ is called *the erosion curve*, with $r \geq 0$, where B is the unit ball of $\mathbb{R}^2$.

It is obvious that this curve is not sensitive to a translation and rotation of the shape.

3 Spectral Function

The scale invariance of the curve Ψ_X is obtained by a normalization number which is the area of the limit of the closings of X by the increasing family $(\lambda B)_{\lambda \geq 0}$ [24].

Definition 1. The *spectral function* s_X of the set X is defined by:

$$s_X(u) = \frac{A(X \ominus uRB)}{A\left(\lim_{\lambda \to \infty} X^{\lambda B}\right)} , \qquad u \geq 0 , \tag{5}$$

where $R = \max\{\lambda,\ \lambda B \subset X\}$ is the maximal radius of the inscribed disks.

Since X will totally disappear after an opening by a structuring element of size bigger than the radius R of the maximal inscribed disk R, we have $s_X(u) = 0$ for $u > 1$.

s_X is a decreasing mapping from 1 to 0, and left continuous. Note that the computation of the spectral function does not use any knowledge of spectral theory or of Fourier analysis;

s_X is translation, rotation, and scale invariant with respect to X.

The following theorem characterizes the limit $\lim_{\lambda \to +\infty} X^{\lambda K}$:

Theorem 2. *[7, 26] If K is a compact convex set with non-empty interior, and if K admits a finite curvature at each point of its boundary, then for every closed set X, we have:*

$$\overline{co}(X) = \lim_{\lambda \to +\infty} X^{\lambda K} ,$$

where $\overline{co}(X)$ is the Euclidean convex hull of X.

4 Study of the Erosion Curve in Higher Dimensions

We first study the erosion curve in $\mathbb{R}^n$ defined by $\Psi_X(r) = V^{(n)}(X \ominus rB)$ where $V^{(n)}$ stands for the hyper-volume measure, and we give the expression of its first derivative. We must define the notion of surface area measure and we use the distributional derivative.

Definition 3. Let $X \subset \mathbb{R}^n$ be a Borel set. We define the n-surface area measure of X by:

$$S^{(n)}(X) = \sup \left\{ \begin{array}{l} \displaystyle\int_{\mathbb{R}^n} \mathbf{1}_X(x) div(f(x)) dx \ , \quad f \text{ is } C^1 \text{ on } \mathbb{R}^n \text{ with} \\ \text{compact support and} \quad \forall x, \|f(x)\| \leq 1 \end{array} \right\} ,$$

where $\mathbf{1}_X(x) = 1$ if $x \in X$ and $\mathbf{1}_X(x) = 0$ if $x \notin X$,

and where $div(f(x)) = \displaystyle\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}$ if $f(x) = (f_1(x), f_2(x), \ldots, f_n(x))$.

In fact, if X is a compact set in $\mathbb{R}^n$ such that its boundary is C^1 almost everywhere, this n-surface area definition is the same as the usual one:

Proposition 4. *[14] Let X be a compact set of $\mathbb{R}^n$ such that its boundary ∂X is C^1 almost everywhere. Then:*

$$S^{(n)}(X) = \int_{\partial X} dh \ ,$$

where dh is the differentiable n-surface area measure of the boundary ∂X of X.

Ψ_X is almost everywhere differentiable because Ψ_X has bounded variation [19]. In the convex case, the derivative of Ψ_X is now well known:

Lemma 5. *[8, 10] If X is a compact convex set such that X is open with respect the unit ball B (i.e. $X = X_B$) then*

$$\frac{d}{dr} V^{(n)}(X \ominus rB) = -S^{(n)}(X \ominus rB) \ , \quad \forall r \in [0,1].$$

The following proposition shows that in the distributional sense the previous formula remains true under much more general conditions.

Proposition 6. *[14] Let X be a compact set (not necessarily convex and open with respect to B) such that for all $r \in [O, R]$ ($R = \sup\{r,\ X \ominus rB \neq \emptyset\}$) the eroded set $X \ominus rB$ is measurable. Then, the mapping $\Psi_X : r \to V^{(n)}(X \ominus rB)$ is differentiable in the distributional sense, and its distributional derivative is equal to $r \to -S^{(n)}(X \ominus rB)$:*

$$\frac{d\left(V^{(n)}(X \ominus rB)\right)}{dr} = -S^{(n)}(X \ominus rB) \ .$$

5 Properties of the Skeleton

In the following, we restrict our study of Ψ_X to X in the plane ($X \subset \mathbb{R}^2$). The erosion curve is the function $\Psi_X : r \to A(X \ominus rB)$. As the properties of the function Ψ_X are based on the skeleton of X, we first recall its definitions and properties and its links with erosions by disks.

The *skeleton* $Sk(X)$ of a compact planar set X is the locus of the maximal inscribed closed balls in X [2] (see Fig. 2).

If we denote by $c_r(X)$ the set of centres of the maximal balls of radius $r > 0$ then $Sk(X) = \bigcup_{r \geq 0} c_r(X)$ and the object reconstruction is given by the formula:

$$X = \bigcup_{r \geq 0} c_r(X) \oplus rB \ .$$

The datum of the set X is equivalent to that of its skeleton together with the maximal radius r associated with each point of $Sk(X)$ [26, 11]. This maximal radius function, called the *quench function*, denoted by q_X, is defined by: $q_X(x) =$

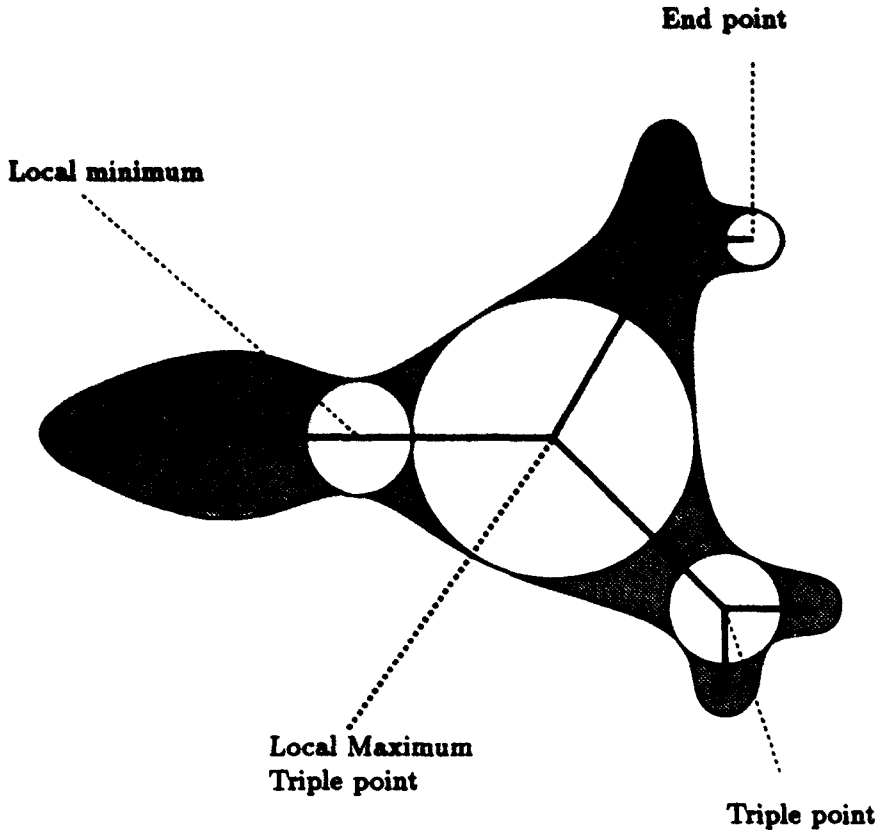


Fig. 2. Skeleton of a planar compact shape

$d(x, X^c)$, for $x \in Sk(X)$, where d is the Euclidean distance, and X^c is the complement of X.

The expression of the skeleton of $X \ominus r_0 B$ is

$$Sk(X \ominus r_0 B) = \bigcup_{r \geq r_0} c_r(X) \quad \text{and} \quad X \ominus r_0 B = \bigcup_{r \geq r_0} c_r(X) \oplus (r - r_0)B \ . \quad (6)$$

In other words, the skeleton of the eroded set $X \ominus r_0 B$ is composed of points x of the skeleton of X where $q_X(x) \geq r_0$.

Note that, unfortunately, there is no similar formula for the skeleton of the open set $X_{r_0 B}$, (contrary to [26, p. 377]) because in general $Sk(X_{r_0 B}) \not\subset Sk(X)$. We have only the reconstruction formula:

$$X_{r_0 B} = \bigcup_{r \geq r_0} c_r(X) \oplus rB \ . \quad (7)$$

The skeleton of X is not necessarily a finite graph [11, 12]. Nevertheless, if the boundary ∂X of a compact planar set X is a finite union of C^3 arcs, then $Sk(X)$ is a connected finite graph with simple arcs [18, 17].

Thus, if the boundary of X is a finite union of C^3 arcs, the skeleton $Sk(X)$ of X has a finite number of end points and multiple points. We consider each

edge of $Sk(X)$ as a curve $\gamma(s)$ where s is a skeleton parameterization with its arc length (see Fig. 3). The skeleton parameterization with arc length goes over the skeleton by following the boundary parameterization. Note that all points which are not end points have many abscissae (for example, a triple point of the skeleton has three abscissae because it is the centre of a maximal inscribable ball with three contact points with the boundary). We finally define on each edge the function q by $q(s) = q_X(\gamma(s)) = d(\gamma(s), X^c)$.

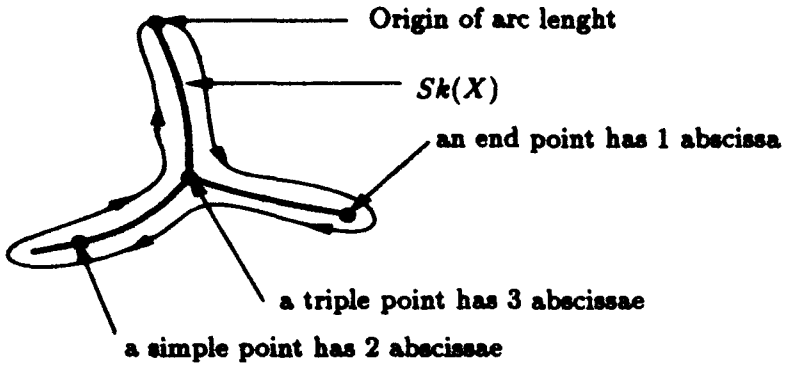
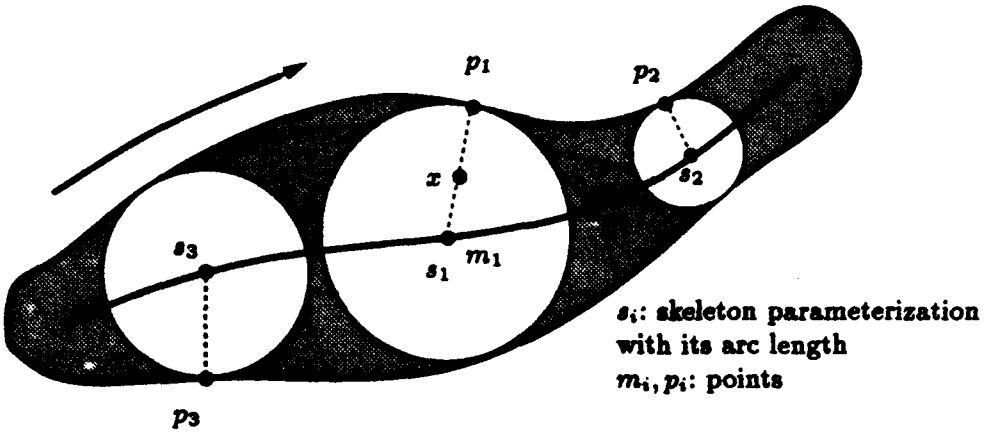


Fig. 3. Skeleton parameterization with arc length



$$Av(x) = [x, p_1],\ Am(x) = [m_1, x],\ Ar(x) = [m_1, p_1]$$

Fig. 4. The skeleton parameterization with arc length goes over the skeleton by following the boundary parameterization

We define the *downstream* of a point $x \in X$ (resp. the *upstream*), denoted by $Av(x)$ (resp. $Am(x)$), to be the set of $y \in X$ that satisfy the relation [11]:

$$d(y, X^c) = d(x, X^c) - d(x, y) \quad (\text{resp. } d(y, X^c) = d(x, X^c) + d(x, y)).$$

If $x \in Sk(X)$ then its upstream is reduced to the point x itself and conversely:

$$x \in Sk(X) \Leftrightarrow Am(x) = \{x\}.$$

We define the *edge* of x, denoted by $Ar(x)$, to be the union of the upstream and the downstream of x: $Ar(x) = Am(x) \bigcup Av(x)$.

Two non-identical edges are either disjoint or cross at one point $x \in Sk(X)$, which must be their common upstream extremity.

6 Study of the Erosion Curve in Two Dimensions

Let Γ be a finite union of C^3 arcs [2, 18, 17]. Let X be the connected compact set defined by $X = \Gamma_{int} \bigcup \Gamma$ where Γ_{int} is the interior of Γ (Jordan's theorem [4]). By hypothesis, X has no holes. The boundary of X is $\partial X = \Gamma$. We say that a point m of Γ is a *critical point of the curvature* if there exists an open neighbourhood $\mathcal{V}(m)$ of m such that the curvature is constant and strictly positive on $\mathcal{V}(m) \bigcap \Gamma$.

The right framework for our theorem is the following: *suppose that for all $r \in \mathbb{R}^+$, the number of connected components of $Sk(X \ominus rB)$ is finite and that the boundary ∂X of X is a finite union of C^3 arcs without critical point.* This class of shapes is a very wide one including polygons and avoids pathological cases.

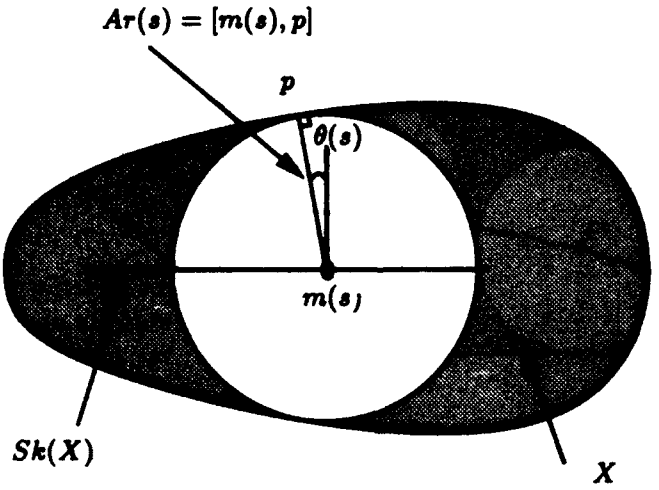


Fig. 5. Skeleton parameterization

Proposition 7. *For all s such that s is a skeleton parameterization with arc length of a point x in the skeleton $Sk(X)$ which is not an end point or a multiple point (i.e. $\gamma(s) = x$), then the function $q : s \to d(\gamma(s), X^c)$ is differentiable and we have*

$$\frac{dq}{ds}(s) = -\sin\theta(s) \ ,$$

where $\theta(s)$ is the angle between the normal on s at $Sk(X)$ and $Ar(x)$.

If $\gamma(s)$ is the skeleton parameterization with arc length s of a point $x \in Sk(X)$ which is an end point (but not critical) then the function q is right and left differentiable and

$$\left(\frac{dq}{ds}\right)_r(s) = -\left(\frac{dq}{ds}\right)_l(s) = -1 \ , \tag{8}$$

with $\left(\frac{dq}{ds}\right)_r(s) = \lim_{\varepsilon \to 0^+} \frac{q(s+\varepsilon) - q(s)}{\varepsilon}$, *the right derivative, and*

$\left(\frac{dq}{ds}\right)_l(s) = \lim_{\varepsilon \to 0^-} \frac{q(s+\varepsilon) - q(s)}{\varepsilon}$, *the left derivative.*

Formula (8) means that $\theta^+(s) = -\theta^-(s) = \frac{\pi}{2}$.

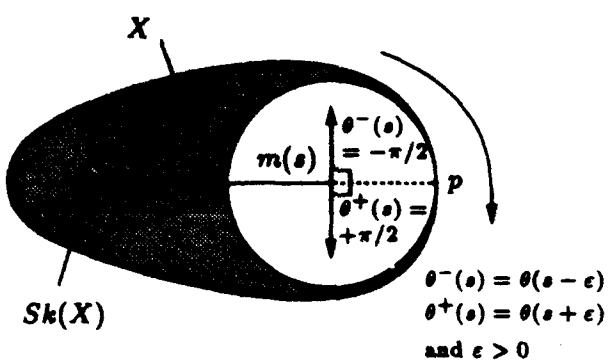


Fig. 6. End point case

If $x \in Sk(X)$ is a multiple point, let $(s_i)_{i=1}^n$ be the different curvilinear abscissae of x, then the function q is right and left differentiable at each s_i and

$$\left(\frac{dq}{ds}(s_i)\right)_r = -\sin\theta^+(s_i) \ , \quad \left(\frac{dq}{ds}(s_i)\right)_l = -\sin\theta^-(s_i) \ ,$$

with $\sum_{i=1}^n (\theta^+(s_i) - \theta^-(s_i)) = (n-2)\pi$.

The proof of this proposition and all other propositions and theorems are given in [15, 13].

In fact q is Lipschitz and twice differentiable except for s such that $\gamma(s)$ is an end point or a multiple point of $Sk(X)$.

Using remark 5, Proposition 7 and the parameterization of the skeleton $Sk(X)$ and of the boundary ∂X of X, we are able to express the surface area of X and

of $X \ominus rB$ for all $r < R$ where $R = \max\{r \mid rB \subset X\}$ is the size of the greatest disk contained in X, as an integral on $Sk(X)$ or on $Sk(X \ominus rB)$ of a function of q and of its first and second derivative:

$$A(X \ominus rB) = \int_{Sk(X \ominus rB)} \frac{q(s) - r}{2\sqrt{1 - q'^2(s)}} \left[2\left(1 - q'^2(s)\right) - q(s)q''(s) + rq''(s)\right] ds \quad (9)$$
$$\forall 0 \leq r < R \ .$$

The first consequence is that all these integrals depend only on the skeleton parameterisation with arc length s of the skeleton and not on the actual shape of the skeleton.

Theorem 8. *Two shapes having the same topology of skeleton, the same edge length and the same q function have the same erosion curve (see Fig. 7).*

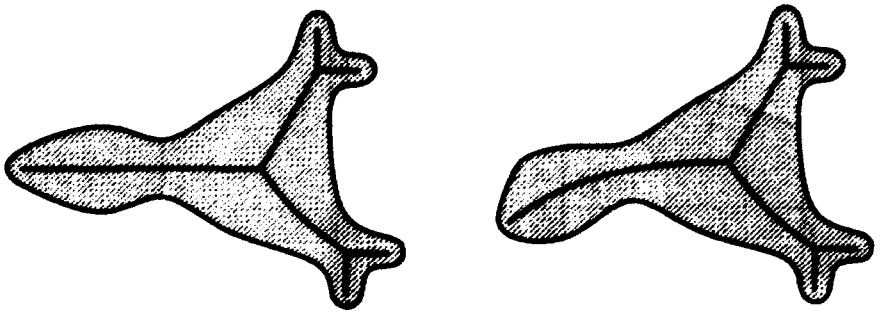


Fig. 7. Two shapes having the same erosion curve

This result may be expressed thus: *the erosion curve quantifies soft shapes* because bending the skeleton does not change the erosion curve. But note that two shapes having the same erosion curve do not necessarily have skeletons which are topologically equivalent.

Let us study the erosion curve in more detail.

Lemma 9. *The function $\Psi_X : r \in \mathbb{R}^+ \to A(X \ominus rB)$ is continuous, decreasing and differentiable for all $r \in]0, R[$ with $R = \max\{r,\ rB \subset X\}$, and we have*

$$\Psi'_X(r) = -P(X \ominus rB) \ , \quad (10)$$

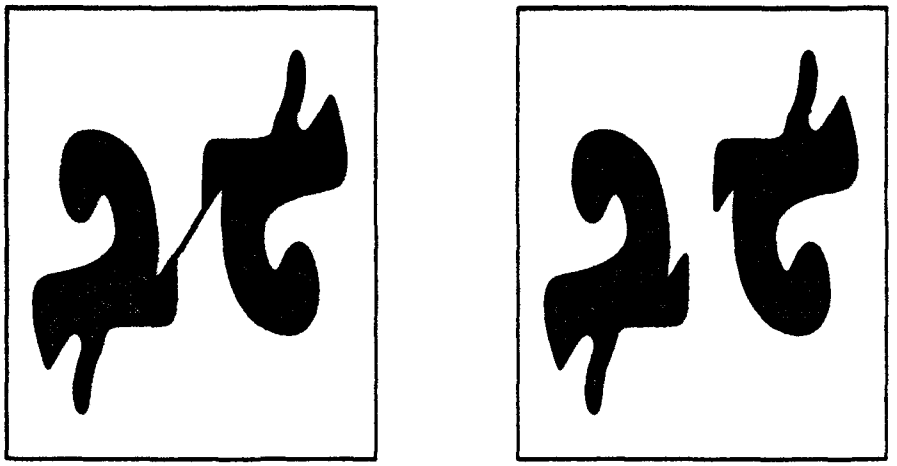
where $P(\cdot)$ represents the perimeter measure [13].

We recall that X is defined by $\Gamma_{int} \bigcup \Gamma$ where Γ is a finite union of C^3 arcs without critical point. This hypothesis means that each point of $Sk(X)$ is the centre of a maximal inscribable disk with a finite number of contact boundary points (i.e. an end point (resp. a triple point, ...) is the centre of a maximal

inscribable circle with one (resp. three, ...) contact boundary points). If we assume only that Γ is a finite union of C^3 arcs, then (10) becomes

$$\lim_{\varepsilon \to 0^+} \frac{\Psi_X(r+\varepsilon) - \Psi_X(r)}{\varepsilon} = -P\Big(\lim_{\varepsilon \to 0^+} \overline{\overset{\circ}{X}} \ominus (r+\varepsilon)B\Big) ,$$
$$\lim_{\varepsilon \to 0^-} \frac{\Psi_X(r+\varepsilon) - \Psi_X(r)}{\varepsilon} = -P\Big(\lim_{\varepsilon \to 0^-} X \ominus (r+\varepsilon)B\Big) .$$

Figure 8 shows a problem arising when X has parts of zero thickness ($\overline{\overset{\circ}{X}} \neq X$).



X $\qquad$ $\overline{\overset{\circ}{X}}$

Fig. 8. The set X has parts with zero thickness, that is $\overline{\overset{\circ}{X}} \neq X$.

In the convex case, we have Miles' formulae [8]. For all compact convex sets $X \subset \mathbf{R}^2$ such that $X = X_B$, and for all $r \in [0,1]$:

$$\begin{cases} P(X \ominus rB) = P(X) - 2\pi r , \\ A(X \ominus rB) = A(X) - P(X)r + \pi r^2 . \end{cases}$$

We examine what happens during an infinitesimal erosion of $X \ominus rB$ by εB for $\varepsilon \to 0^+$. Five cases are possible.

1. The erosion is simple: the number of end points, multiple points, and connected components of $Sk(X \ominus rB)$ and $Sk(X \ominus (r+\varepsilon)B)$ are equal.
2. There is a disconnection: $X \ominus rB$ has less connected components than $X \ominus (r+\varepsilon)B$.
3. A connected component of $X \ominus rB$ containing a multiple point of $Sk(X)$ which is a local maximum of the quench function q_X vanishes.
4. A connected component of $X \ominus rB$ not containing a multiple point of $Sk(X)$ vanishes.
5. A multiple point of $Sk(X)$ vanishes.

We define the right and the left second derivatives of Ψ_X by

$$\begin{cases} \Psi_l''(r) = \lim\limits_{\varepsilon \to 0^+} \dfrac{1}{\varepsilon} [\Psi'(r+\varepsilon) - \Psi'(r)], \\ \Psi_r''(r) = \lim\limits_{\varepsilon \to 0^-} \dfrac{1}{\varepsilon} [\Psi'(r+\varepsilon) - \Psi'(r)], \end{cases}$$

Theorem 10. *If the five cases occur at different sizes of erosions, then they can be distinguished on the erosion curve. The discontinuities of the second derivative of Ψ_X characterize them.*

All results are summarized in the Table 1.

But note the following points:

- If two cases occur at the same r, nothing can be said, and the equivalences are only implications.
- The order of multiplicity of the multiple points cannot be computed. For example, in the convex polygonal case [13], it can be shown that the number of edges cannot be computed from the erosion curve.

More precisely:

Proposition 11. *Given an erosion curve Ψ_X of a convex polygon circumscribed on a circle, there exists $n_0 \geq 0$ such that for each $n \geq n_0$, a convex polygon with n sides can be constructed that has Ψ_X as erosion curve.*

In fact, for a convex polygon inscribed in a circle, the only information contained in the granulometric curve is $\sum_{i=1}^{n} \tan(\delta_i/2)$ where δ_i is the angle of the polygon at vertex i.

7 Principle of Reconstruction

We now tackle the problem of constructing shapes from the knowledge of erosion curves Ψ_X. To stay in the same conditions as before, the skeleton of X will be a finite union of C^3-arcs whose multiple points are only triple points. We will read the erosion curve with decreasing r from R to 0.

At each step, we point out the degrees of freedom in the reconstruction process. Note that the reconstruction process gives us all the shapes having the same erosion curve.

Suppose that for a given $r \in]0, R]$, the set $X \ominus rB$ is already constructed. The skeleton $Sk(X \ominus rB)$ and the function $q_r : s \in Sk(X \ominus rB) \to q(s) - r$ are also known. We propose a method for building a set Y having the same erosion curve, such that there exists $\varepsilon > 0$ independent of r, $Y \ominus \varepsilon B = X \ominus rB$ (ε is smaller than the difference of two successive values of r where Ψ_X'' is discontinuous).

Table 1. Theorem 10

Fundamental theorem			
Infinitesimal erosion	**Erosion curve**	$X \ominus (r-\varepsilon)B \Leftrightarrow X \ominus (r+\varepsilon)B$	**Axes of freedom for reconstruction**
1 The erosion is simple: The number of end points, multiple points, and connected components of $Sk(X \ominus rB)$ and of $Sk(X \ominus (r+\varepsilon)B)$ is equal.	$\Psi''_{X_r}(r) = \Psi''_{X_l}(r) < +\infty$		Lengths of arcs of the skeleton
2 There is a disconnection: The number of connected components of $X \ominus rB$ are greater than those of $X \ominus (r+\varepsilon)B$.	$\Psi''_{X_l}(r) < \Psi''_{X_r}(r) = +\infty$		The choice of arbitrary connected components
3 A connected component of $X \ominus rB$ containing a multiple point of $Sk(X)$ which is a local maximum of the quench function qx vanishes.	$\Psi''_{X_r}(r) < \Psi''_{X_l}(r) < +\infty$		The order and the geometry of the multiple point
4 A connected component of $X \ominus rB$ not containing a multiple point of $Sk(X)$ vanishes.	$\Psi''_{X_r}(r) < \Psi''_{X_l}(r) = +\infty$		The choice of edges of the skeleton which will be connected
5 A multiple point of $Sk(X)$ vanishes	$\Psi''_{X_l}(r) < \Psi''_{X_r}(r) < +\infty$		The order and the geographical location of the multiple point

- *If r is not a discontinuity of Ψ''_X* (Case 1), then at each extremity α_i of $Sk(X \ominus rB)$, we extend the segment line of a length equal to $\eta_i = \varepsilon / q'_r(\alpha_i)$. The values of $q'_{r-\varepsilon}(\alpha_i + \eta_i)$ are constrained only by one equation involving $\Psi''_X(r-\varepsilon)$. This infinity of solutions gives the first degrees of freedom in the reconstruction process.
  We first draw parallel exterior arcs of the boundary of $X \ominus rB$ at distance ε and add tips at each extremity such that the total surface area of these tips is equal to a suitable constant induced by $\Psi_X(r)$ (see Fig. 9).

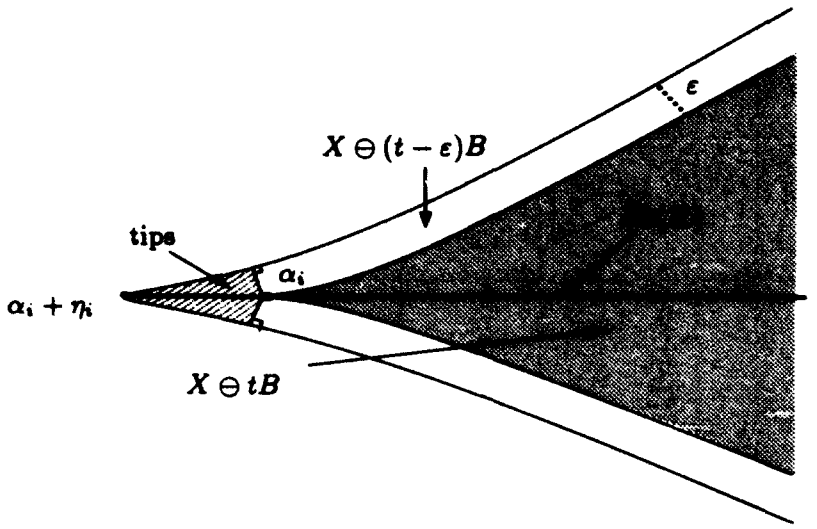


Fig. 9. Reconstruction of $X \ominus (t-\varepsilon)B$ from $X \ominus rB$ when r is not a discontinuity of Ψ''_X.

- *If r is a discontinuity of Ψ''_X* then
  - *If $\Psi''_{X_r}(r) - \Psi''_{X_l}(r) > 0$* (Case 5), we arbitrarily choose an end point s_0 of $Sk(X \ominus rB)$ that we transform into a triple point (see Fig. 10).
    If δ_0, δ_1 are the angles between the new line segments at s_0 with the old line segment, we have

$$\begin{cases} \tan(\delta_0 + \delta_1) - \tan\delta_0 - \tan\delta_1 = -\frac{1}{2}\left[\Psi''_{X_r}(r) - \Psi''_{X_l}(r)\right] \\ \cos(\delta_0 + \delta_1) = -q'(s_0) \end{cases}, \tag{11}$$

    We extend the skeleton of $\eta_i = \varepsilon / q'_r(\alpha_i)$ for all other extremities $\alpha_i \neq s_0$ and of respectively $\varepsilon / \sin\delta_0$ and $\varepsilon / \sin\delta_1$ for the new branch of the skeleton. In general, the system (11) has two symmetrical solutions in δ_0, δ_1. We implicitly infer $q'_{r-\varepsilon}(\alpha_i + \eta_i)$ from the value of $\Psi''_X(r-\varepsilon)$ and compute in the same way as the boundary of $X \ominus (r-\varepsilon)B$.
  - *If $\Psi''_{X_r}(r) - \Psi''_{X_l}(r) < 0$* (Case 3), a new connected component is created, the skeleton of which has a triple point. This triple point is a local maximum of q_X and if $\delta_0, \delta_1, \delta_2$ are angles between the three edges of

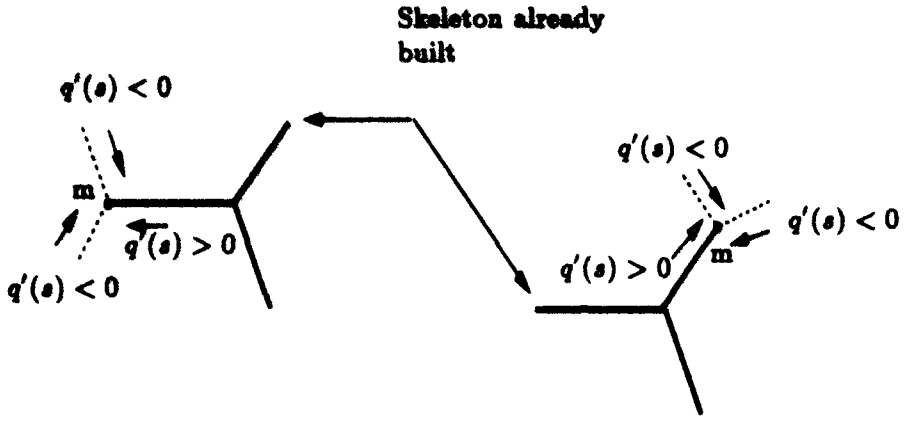


Fig. 10. The arbitrary chosen end point s_0 of $Sk(X \ominus rB)$ is transformed into a triple point.

this new piece of $Sk(X)$ we have

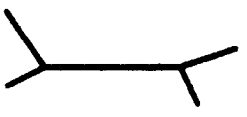
$$\sum_{i=0}^{2} \tan \delta_i = \frac{1}{2} \left[\Psi_{X_r}''(r) - \Psi_{X_l}''(r) \right] \quad \text{with} \quad \sum_{i=0}^{2} \delta_i = 2\pi \ .$$

In general, this equation has infinitely many solutions, with one degree of freedom, for example δ_0.

- *If* $\Psi_{X_r}''(r) = +\infty$ (Case 2), we have a narrow part on X. We reconnect two arbitrary different components of $Sk(X \ominus rB)$, i.e. the union of two end points belonging to two different components.
- *If* $\Psi_{X_l}''(r) = +\infty$ (Case 4), one adds a new connected component, the skeleton of which is an edge with a local maximum of q_X on it.

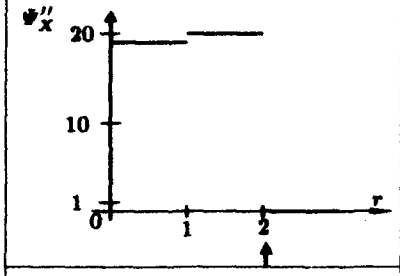
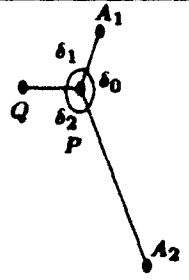
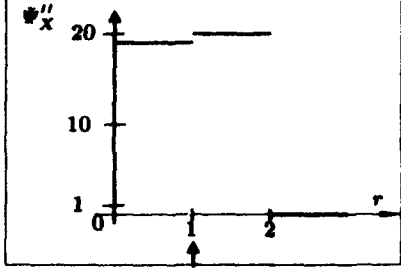
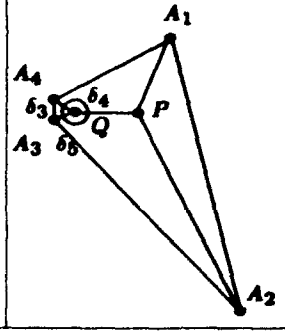
Note that at each step, the curvature of the skeleton is not derived from the erosion curve.

Let us now illustrate the use of these rules on a real example of an erosion curve Ψ. Table 2 shows the different steps. The discontinuities of Ψ'' are located at $r = 1$ and $r = 2$. As Ψ'' is a staircase function, we can choose X to be a polygon. The Ψ'' curve shows that the skeleton of X is topologically equivalent to



Then, we have only one degree of freedom in the reconstruction process, namely δ_0 (see Table 2). Figure 11 shows other reconstruction polygons with different δ_0.

Table 2. An example of reconstruction process. Different solutions are built in Fig. 11

Reconstruction Process				
Ψ''_X and current value of r	Discontinuity	Corresponding Equations	Reconstruction Process	Example of Numerical Results
	$\Psi x''_l(2)-\Psi x''_r(2)<0$ A triple point, local maximum of qx, is created (Case 3).	$\sum_{i=0}^{2}\tan\delta_i=\frac{1}{2}\left[\Psi x''_r(2)-\Psi x''_l(2)\right]$ and $\sum_{i=0}^{2}\delta_i=2\pi$		$\delta_0 = 120°$, $\delta_1 = 97.5°$, $\delta_2 = 142.5°$, $PA_1 = 2.52$, $PA_2 = 15.24$, $PQ = 2$.
	$\Psi x''_l(1)-\Psi x''_r(1)>0$ A new triple point at the end of an edge of $Sk(X)$ is created (Case 5).	$\sum_{i=3}^{5}\tan\delta_i=\frac{1}{2}\left[\Psi x''_r(1)-\Psi x''_l(1)\right]$ and $\cos(\delta_4+\delta_5)=-\frac{1}{PQ}=-0.5$		$\delta_3 = 60°$, $\delta_4 = 137.5°$, $\delta_5 = 162.5°$, $QA_3 = 1.05$, $QA_4 = 1.36$, $PQ = 2$.

8 Conclusion

The spectral function is not only of mathematical interest but also a powerful tool in the discrimination of shapes as illustrated by the following application [13, 16]. We want to discriminate the morphological extracted shape features of four types of hand-drawn shapes (stars, almost circular objects, elongated shapes, and objects consisting of many blobs) by neural methods.

A set of 100 patterns is used in discrimination experiments. For each pattern, we first compute the associated spectral function and we sample the curve at the 26 values:

$$u \in \{0.0, 0.4, 0.8, \ldots, 0.92, 0.96, 1.0\} .$$

The learning set is composed randomly of 2/3 of the examples. The results obtained with a multi-layer perceptron ((26-8-4) totally connected), averaged on 10 random experiments are as follows:

Spectral function	
Learning	100.0%
Generalization	99.0%

The study of $\Psi_X : r \in \mathbb{R}^+ \to A(X \ominus rB)$ has shown that Ψ_X is continuous and differentiable, and the discontinuities of its second derivative give characteristic shape information. There are three types of discontinuities, each one characteristic of an unambiguous event. For a given r:

1. If $\Psi_{X_r}''(r) = +\infty$, then the shape X under study has a narrow part (Case 2).
2. If $\Psi_{X_l}''(r) = +\infty$, then X has a swell (Case 4).
3. If $\Psi_{X_l}''(r)$ and $\Psi_{X_r}''(r)$ exist but are different, then there is a multiple point on the skeleton of X (Cases 3 and 5). Nevertheless, it is impossible to know the order of this multiple point.

If we suppose that the five cases of Table 1 could not occur at the same time, then the number of narrow parts and swells on the shape and the number of multiple points in the skeleton are information contained in the erosion curve; otherwise we have only a lower bound for these numbers. Besides, the polygonal feature is not measurable from the erosion curve, because bending the skeleton does not change the erosion curve, and if we are sure that the shape is a polygon, we have only a lower bound on the number of its vertices.

As to opening curve, (i.e. $\Phi : r \to A(X_{rB})$), the results are very similar. The opening curve extracts the same kind of information about X, located in the discontinuities of Φ, Φ', and Φ''. The study of the opening curve together with the closing curve, i.e.

$$\Omega : r \to \begin{cases} A(X^{-rB}), r < 0 \\ A(X_{rB}), \quad r \geq 0 \end{cases}$$

(where $(X^{rB})^c = X^c_{rB}$ and X^c represents the complementation of X), gives much more information about X but its study is not yet undertaken.

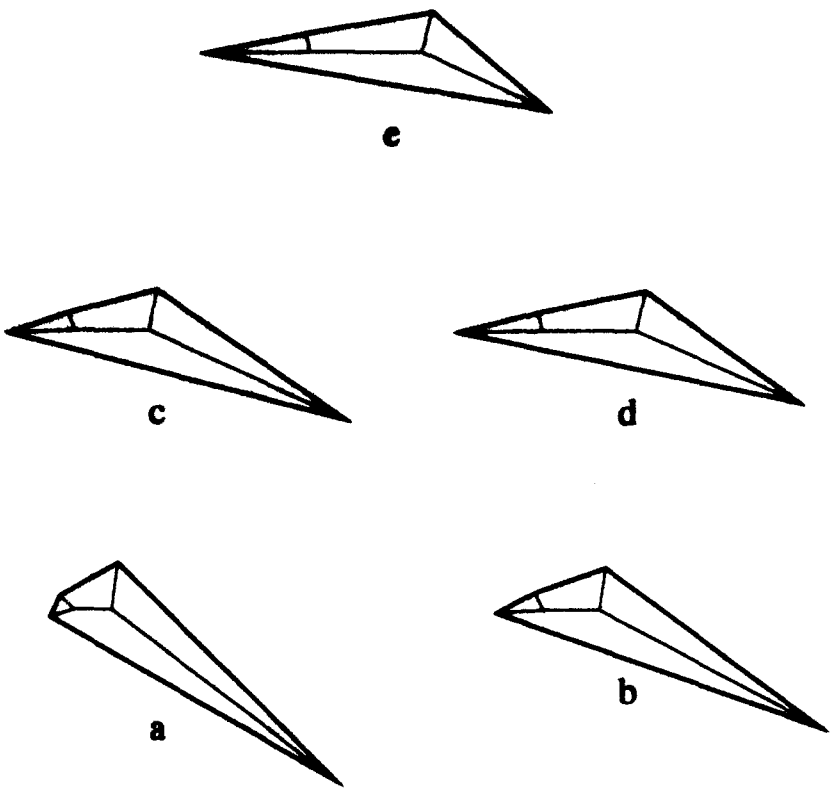


Fig. 11. Examples of polygons having the same function Ψ_X given in Table 2. (a) PQ=2, (b) PQ=3, (c) PQ=4, (d) PQ=5, (e) PQ=6.

References

1. Aubin J-P., Frankowska, H. (1990). Set-Valued Analysis, Birkhauser, Boston.
2. Calabi, L., Riley, J.A. (1967). The skeletons of stable plane sets, Technical Report AF 19 (628–5711), Parke Math. Lab. Inc., Massachusetts.
3. Coster, M., Chermant, J.L. (1985). Précis d'Analyse d'Images, CNRS Ed., Paris.
4. Dieudonné, J. (1969). Eléments d'Analyse, volume I, Gauthier-Villars, Paris.
5. Lantuéjoul, Ch. (1978). La squelettisation et son application aux mesures topologiques des mosaiques polycristallines, Thèse Ecole des Mines de Paris.
6. Lantuéjoul, Ch., Maisonneuve, F. (1984). Geodesic methods in quantitative image analysis, Pattern Recognition, 17, pp. 177–187.
7. Matheron, G. (1975). Random Sets and Integral Geometry, John Wiley and Sons, New York.
8. Matheron, G. (1977). La formule de Steiner pour les érosions, Technical Report 496, Centre de Géostatistique, Ecole des Mines de Paris.
9. Matheron, G. (1978). The infinitesimal erosions. In: Miles, R.E., Serra, J. (eds.), Lecture Notes in Biomathematics. 23, Geometrical Probability and Biological Structures: Buffon's 200th Anniversary, Springer Verlag, Berlin.

10. Matheron, G. (1978). Quelques propriétés topologiques du squelette, Technical Report 560, Centre de Géostatistique, Ecole des Mines de Paris.
11. Matheron, G. (1988). Examples of topological properties of skeletons. In: Serra, J. (ed.), Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances, Academic Press, London.
12. Matheron, G. (1988). On the negligibility of the skeleton and the absolute continuity of erosions. In: Serra, J. (ed.), Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances, Academic Press, London.
13. Mattioli, J. (1991). Squelette, érosion et fonction spectrale par érosion d'une forme binaire planaire, Technical Report ASRF-91-8, Thomson-CSF, L.C.R., Orsay.
14. Mattioli, J. (1992). Etude de la fonction $\Psi_X : r \rightarrow V^{(n)}(X \ominus rB)$ pour $X \subset \mathbb{R}^n$ compact simplement connexe, Technical Report ASRF-92-3, Thomson-CSF, L.C.R., Orsay.
15. Mattioli, J., Schmitt, M. (1993). Inverse problems for granulometries by erosion, Journal of Mathematical Imaging and Vision, to appear.
16. Mattioli, J., Schmitt, M., Pernot, E., Vallet, F. (1991). Shape discrimination based on mathematical morphology and neural networks. In: Proc. Int. Conf. on Artificial Neural Networks, Helsinki, pp. 112-117.
17. Riley, J. (1965). Plane graphs and their skeletons, Technical Report 60429, Park Math. Lab. Inc., Massachusetts.
18. Riley, J., Calabi, L. (1964). Certain properties of circles inscribed in simple closed curves, Technical Report 59281, Park Math. Lab. Inc., Massachusetts.
19. Rudin, W. (1966). Real and Complex Analysis, Mc Graw-Hill, New York.
20. Santalo, L.A. (1976). Integral Geometry and Geometric Probability, Addison Wesley, London.
21. Schmitt, M. (1989). Des algorithmes morphologiques à l'intelligence artificielle. Thèse Ecole des Mines de Paris.
22. Schmitt, M. (1990). Connexité du squelette d'un compact convexe. Technical Report ASRF-90, Thomson-CSF, L.C.R., Orsay.
23. Schmitt, M. (1991). On two inverse problems in mathematical morphology. In: Dougherty, E.R. (ed.) Mathematical Morphology in Image Processing. Marcel Dekker. Inc., New York.
24. Schmitt, M., Mattioli, J. (1991). Shape recognition combining mathematical morphology and neural networks. In: SPIE: Application of Artificial Neural Network, Orlando.
25. Schmitt, M., Vincent, L. (1993). Morphological Image Analysis: A Practical and Algorithmic Handbook, Cambridge University Press, to appear.
26. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.
27. Stoyan, D., Kendall, W.S., Mecke, J. (1987). Stochastic Geometry and Its Applications, John Wiley and Sons, New York.

Morphological Area Openings and Closings for Grey-scale Images*

Luc Vincent

Xerox Imaging Systems, 9 Centennial Drive, Peabody MA 01960, USA

Abstract. The filter that removes from a binary image the components with area smaller than a parameter λ is called area opening. Together with its dual, the area closing, it is first extended to grey-scale images. It is then proved to be equivalent to a maximum of morphological openings with all the connected structuring elements of area greater than or equal to λ. The study of the relationships between these filters and image extrema leads to a very efficient area opening/closing algorithm. Grey-scale area openings and closings can be seen as transformations with a structuring element which locally adapts its shape to the image structures, and therefore have very nice filtering capabilities. Their effect is compared to that of more standard morphological filters. Some applications in image segmentation and hierarchical decomposition are also briefly described.

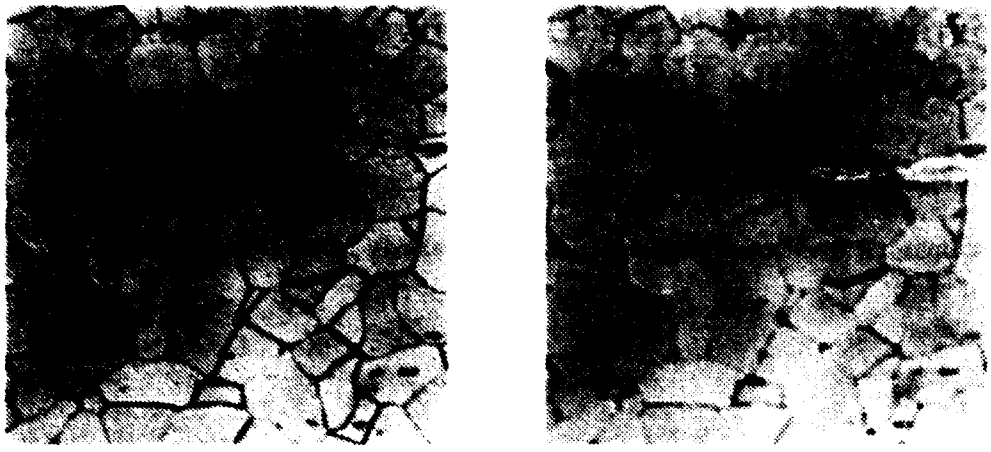
Keywords: area opening, extrema, filtering, opening and closing, mathematical morphology, shape.

1 Introduction

A classic image analysis preprocessing problem consists of filtering out small light (respectively dark) particles from grey-scale images without damaging the remaining structures. Often, simple morphological openings (respectively closings) [10, 11] with disks or approximations of disks like squares, hexagons, octagons, etc., are good enough for this task. However, when the structures that need to be preserved are elongated objects, they can be either completely or partly removed by such an operation.

Let us consider for example Fig. 1a, representing a microscopy image of a metallic alloy. It is "corrupted" by some black noise that one may wish to remove (note that part of what is called noise here is the intra-grain texture!). As shown in Fig. 1b, a closing of this image with respect to the elementary ball of the

* The author is grateful to Henk Heijmans, Christian Lantuéjoul, Ben Wittner, and Gilles Leborgne for several useful suggestions and fruitful discussions.



(a) (b)

Fig. 1. Microscopic image of a metallic alloy (a) and its morphological closing by an elementary (9 pixels) square (b).

8-connected metric (i.e. a square of 9 pixels) severely damages most of the inter-grain lines, while still preserving some of the largests bits of noise (like the blobs in the bottom right and left corners).

This is the reason why, in this context, openings and closings with line segments are widely used. In this paper, the morphological opening and closing by a structuring element B are denoted by γ_B and ϕ_B respectively (see, e.g., [10, 11, 12]). Denote also by l_n^0, l_n^1, l_n^2, l_n^3 the line segments of length n and respective orientation 0°, 45°, 90° and 135°. The following operations

$$\Gamma_n^l = \vee_{i\in[0,3]}\gamma_{l_n^i} \quad \text{and} \quad \Phi_n^l = \wedge_{i\in[0,3]}\phi_{l_n^i}$$

are respectively an algebraic opening and an algebraic closing [10, 11, 12]. They tend to preserve elongated structures better than their disk-based counterparts (see also [11, pp. 110–112]). However, they are still far from being ideal: indeed, they are first very computationally intensive, since they involve a series of expensive operations. Furthermore, as illustrated by Fig. 2, they may remain unsatisfactory in some cases; when n is small, some of the noise fragments are still present, and with increasing values of n, the inter-grain lines tend to be damaged.

The remedy to this last problem is to increase the number of orientations of the used line segments, but this in turn increases the computational complexity of the algorithm. In addition, even with a large number of orientations, very thin lines might still end up broken. As will be seen in Sect. 4, the classic solution to this involves a transformation called *grey-scale reconstruction* [4, 2, 16, 17]. In this paper, an even better and more systematic technique is proposed: use *all* possible connected structuring elements of a given size (number of pixels). This will lead to the introduction of the area openings and closings.

The paper is organized as follows: in the next section, area openings and closings are defined and some of their properties are reviewed. Their relations

(a) (b) (c)

Fig. 2. Maxima of linear openings of increasing size of Fig. 1a.

with image extrema are studied and are at the basis of a very efficient algorithm. Lastly, Sect. 4 illustrates their usefulness for some filtering, segmentation, and hierarchical decomposition applications.

2 Area Openings and Closings: Definitions and Properties

2.1 Definition in Terms of Areas

Throughout the paper, the sets X under study will be constrained to be subsets of a connected compact set $\mathbf{M} \subset \mathbb{R}^2$ called the "mask". All the notions and algorithms introduced easily generalize to arbitrary dimensions.

The definitions proposed below for area openings and closings are based on the so-called connected openings [11, 12]:

Definition 1 Connected opening. The connected opening $C_x(X)$ of a set $X \subseteq \mathbf{M}$ at point $x \in \mathbf{M}$ is the connected component of X containing x if $x \in X$ and $\emptyset$ otherwise.

On binary two-dimensional images (i.e., on subsets of the mask $\mathbf{M}$), the area opening γ_λ^a is defined as follows:

Definition 2 Binary area opening. Let $X \subset \mathbf{M}$ and $\lambda \geq 0$. The area opening of parameter λ of X is given by

$$\gamma_\lambda^a(X) = \{x \in X \mid \mathsf{Area}\,(C_x(X)) \geq \lambda\} \ . \tag{1}$$

More intuitively, if $(X_i)_{i \in I}$ denotes the connected components of X, it becomes clear that $\gamma_\lambda^a(X)$ is equal to the union of the connected components of X with area greater than λ:

$$\gamma_\lambda^a(X) = \bigcup \{X_i \mid i \in I, \mathsf{Area}\,(X_i) \geq \lambda\} \ . \tag{2}$$

By area is meant the *Lebesgue* measure in $\mathbb{R}^2$.

Obviously, γ_λ^a is *increasing, idempotent,* and *anti-extensive.* It is therefore legitimate to call it an opening [7, 11, 12]. By duality, the binary area closing can be defined as follows:

Definition 3. The area closing of parameter $\lambda \geq 0$ of $X \subset \mathbf{M}$ is given by:

$$\phi_\lambda^a(X) = [\gamma_\lambda^a(X^C)]^C.$$

where X^C denotes the complement of X in $\mathbf{M}$, i.e. the set $\mathbf{M} \setminus X$ ($\setminus$ denoting the set difference operator). As the dual of the area opening, the area closing fills in the holes of a set whose areas are strictly smaller than the size parameter λ.

The growth of these transformations makes it possible to extend them straightforwardly to grey-scale images [12], i.e., to mappings from $\mathbf{M}$ to $\overline{\mathbb{R}}$:

Definition 4 Grey-scale area opening. For a mapping $f : \mathbf{M} \longrightarrow \overline{\mathbb{R}}$, the area opening $\gamma_\lambda^a(f)$ is given by:

$$(\gamma_\lambda^a(f))(x) = \sup\{h \leq f(x) \mid \text{Area}\,(\gamma_x(T_h(f))) \geq \lambda\} \quad (3)$$

$$= \sup\{h \leq f(x) \mid x \in \gamma_\lambda^a(T_h(f))\}\ . \quad (4)$$

In this definition, $T_h(f)$ stands for the threshold of f at value h, i.e:

$$T_h(f) = \{x \in \mathbf{M} \mid f(x) \geq h\}\ . \quad (5)$$

In other words, to compute the area opening of f, all the possible thresholds $T_h(f)$ of f are first considered and their area openings $\gamma_\lambda^a(T_h(f))$ are found. Since γ_λ^a is increasing, $Y \subseteq X \implies \gamma_\lambda^a(Y) \subseteq \gamma_\lambda^a(X)$. Thus, the $\{\gamma_\lambda^a(T_h(f))\}_{h \in \mathbb{R}}$ are a decreasing sequence of sets which by definition constitute the threshold sets of the transformed mapping $\gamma_\lambda^a(f)$.

By duality, one similarly extends the concept of area closing to mappings from $\mathbf{M}$ to $\overline{\mathbb{R}}$. These area openings and closings for grey-scale images are typical examples of flat increasing mappings (also called stack mappings) [12, 19, 13]. Their geometric interpretation is relatively simple: a grey-scale area opening basically removes from the image all the light structures which are "smaller" than the size parameter λ, whereas the area closing has the same effect on dark structures. It is stressed that the word *size* exclusively refers here to an area (or number of pixels in the discrete case). Theorem 10 below will provide a more refined interpretation of this intuitive interpretation.

2.2 Second Approach to Area Openings and Closings

In this section, it is shown that area openings can be obtained through maxima of classic morphological openings with connected structuring elements. Recall that γ_B denotes the morphological opening by structuring element B.

Lemma 5. *Let $B \subset \mathbf{M}$. $\gamma_B \leq \gamma_\lambda^a$ if and only if B is a finite union of connected components of area greater or equal to λ.*

Proof. If $B = \cup_{i=1}^n B_i$ with $\forall i \in [1, n]$, B_i connected and Area $(B_i) \geq \lambda$, then for any i, $\gamma_{B_i} \leq \gamma_\lambda^a$. Thus, $\gamma_B \leq \gamma_\lambda^a$. Conversely, if $\gamma_B \leq \gamma_\lambda^a$, then $\gamma_B(B) \leq \gamma_\lambda^a(B)$, i.e. $B \leq \gamma_\lambda^a(B)$. Since γ_λ^a is anti-extensive, this implies that $B = \gamma_\lambda^a(B)$. Thus, by definition of γ_λ^a, this implies that every connected component of B is of area $\leq \lambda$. Since we operate in domain M these components are in finite number. □

The following theorem can now be stated:

Theorem 6. *Denoting by $\mathcal{A}_\lambda$ the class of the subsets of $\mathbf{M}$ which are connected and whose area is greater than or equal to λ, the following equation holds:*

$$\gamma_\lambda^a = \bigcup_{B \in \mathcal{A}_\lambda} \gamma_B \ . \tag{6}$$

Proof. γ_λ^a being a translation-invariant algebraic opening, a famous result by Matheron [7] states that it is the supremum of all the morphological openings γ_B that are smaller than or equal to γ_λ^a:

$$\gamma_\lambda^a = \bigcup \{\gamma_B \mid \gamma_B \text{ morphological opening}, \gamma_B \subseteq \gamma_\lambda^a\}.$$

Thus, applying lemma 5,

$$\gamma_\lambda^a = \bigcup \{\gamma_B \mid B = \textstyle\bigcup_{i=1}^n B_i,\ B_i \text{ connected}, \text{Area}(B_i) \geq \lambda\}.$$

Obviously, for each of these Bs, $\gamma_B \subseteq \gamma_{B_i}$, $\forall i$. The above union can thus be reduced to the *connected* sets B of area $\geq \lambda$:

$$\gamma_\lambda^a = \bigcup \{\gamma_B \mid B \text{ connected}, \text{Area}(B) \geq \lambda\},$$

which completes the proof. □

Similarly, it can be proved that the area-closing of parameter λ is equal to the infimum of all the closings with connected structuring elements of area greater or equal to λ.

In the discrete domain, any connected set of area greater or equal to $\lambda \in \mathbb{N}$ contains a connected set of area equal to λ. The theorem can thus be made more specific as follows:

Corollary 7. *Let $\mathbb{Z}^2$ be the discrete plane equipped with e.g., 4- or 8-connectivity. For $X \in \mathbb{Z}^2 \cap \mathbf{M}$ and $\lambda \in \mathbb{N}$,*

$$\gamma_\lambda^a(X) = \bigcup \{\gamma_B(X) \mid B \in \mathbb{Z}^2 \text{ connected}, \text{Area}(B) = \lambda\}.$$

Theorem 6 can now be extended to grey-scale:

Proposition 8. *Let $f : \mathbf{M} \longrightarrow \overline{\mathbb{R}}$, be an upper semi-continuous mapping [10, pp. 425–429]. The area opening of f is given by:*

$$\gamma_\lambda^a(f) = \bigvee_{S \in \mathcal{A}_\lambda} \gamma_S(f) \ . \tag{7}$$

Note that to extend Theorem 6 to grey-scale, we need to apply it to the threshold sets $T_h(f)$. They thus have to be compact, and this is why upper semi-continuity of f is required. A dual proposition can be stated for grey-scale area closings, which now requires lower semi-continuity for f.

The previous proposition leads to a different understanding of area openings (respectively closings). As a maximum of openings with all possible connected elements of a minimal size, it can be seen as *adaptive:* at every location, the structuring element adapts its shape [1] to the image structure so as to "remove as little as possible".

3 Relation with Extrema, Algorithm

This section exclusively deals with openings, the dual case of the closings being easy to derive from the results. We first recall the notion of *maximum* on a mapping [10, page 445].

Definition 9 Regional maximum. Let f be an upper semi-continuous (u.s.c.) mapping from $\mathbf{M}$ to $\overline{\mathbf{R}}$. A (regional) maximum of f at level $h \in \overline{\mathbf{R}}$ is a connected component M of $T_h(f)$ such that

$$\forall h' > h, \quad T_{h'} \cap M = \emptyset \ . \tag{8}$$

The following theorem can now be stated:

Theorem 10. *Let f be a u.s.c. mapping from $\mathbf{M}$ to $\overline{\mathbf{R}}$, $\lambda \geq 0$. Denoting $\mathcal{M}_\lambda$ the class of the u.s.c. mappings $g : \mathbf{M} \longmapsto \overline{\mathbf{R}}$ such that any maximum M of g is of area greater than or equal to λ,*

$$\gamma_\lambda^a(f) = \sup\{g \leq f \mid g \in \mathcal{M}_\lambda\} \ . \tag{9}$$

Proof. Let $g \in \mathcal{M}_\lambda$, $g \leq f$, and let $h \in \overline{\mathbf{R}}$. Let A be an arbitrary connected component of $T_h(g)$. Since g is u.s.c., A is a compact set and therefore, there exists $x \in A$ such that $g(x) = \max\{g(y) \mid y \in A\}$. Let $h' = g(x)$ and $B = C_x(T_{h'}(g))$. B is obviously a maximum of g at altitude h'. Indeed, if there existed a $y \in B$ such that $g(y) > h'$, we would have $y' \notin A$ (the maximal value of g on A is h), and thus $A \subset A \cup B \subseteq T_h(g)$. Furthermore, $A \cup B$ is connected as the union of two connected sets with non-empty intersection, which would be in contradiction with the fact that A is a connected component of $T_h(g)$. B is therefore a maximum at altitude h' of g and $B \subseteq A$. Since by hypothesis, $\text{Area}(B) \geq \lambda$, we therefore have $\text{Area}(A) \geq \lambda$.

Thus, for every $h \in \overline{\mathbf{R}}$, $\gamma_\lambda^a(T_h(g)) = T_h(g)$. Besides, $T_h(g) \subseteq T_h(f)$. Therefore, by growth of γ_λ^a, $\gamma_\lambda^a(T_h(g)) = T_h(g) \subseteq \gamma_\lambda^a(T_h(f))$. This being true for every threshold, we conclude that $g \leq \gamma_\lambda^a(f)$.

Conversely, $\forall h \in \overline{\mathbf{R}}$, any connected component A of $T_h(\gamma_\lambda^a(f))$ is of area $\geq \lambda$. Thus, all the maxima of $\gamma_\lambda^a(f)$ are of area $\geq \lambda$. It follows that $\gamma_\lambda^a(f) \in \mathcal{M}_\lambda$ and (anti-extensivity) $\gamma_\lambda^a(f) \leq f$, which completes the proof. □

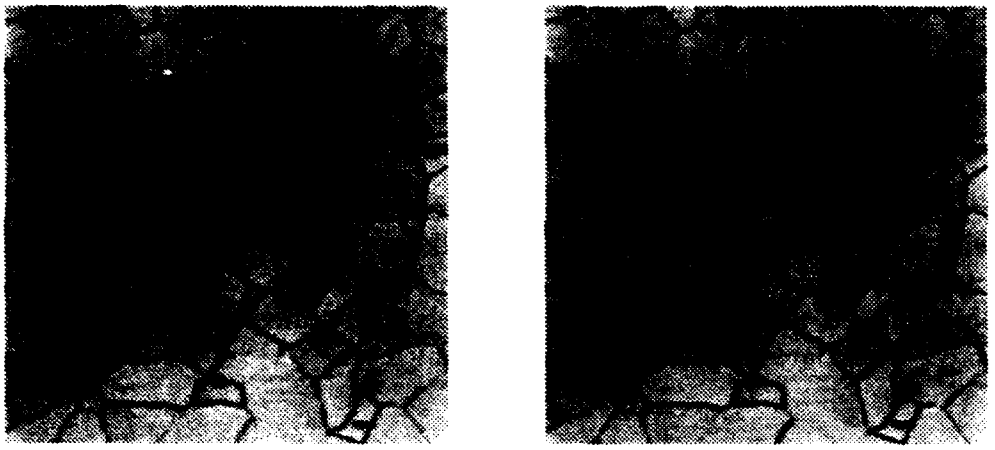
This theorem provides a third interpretation of grey-scale area openings useful for implementation purposes. Indeed:

- Obviously, applying Definition 2 and computing $\gamma_\lambda^a(I)$ for every threshold of the original grey-scale image I then "piling up" the resulting binary images is a much too computationally expensive operation.
- Similarly, computing all the possible openings with all the possible connected structuring elements of λ pixels (see Sec. 2.2) becomes an impossible task as soon as λ is greater than 4 or 5. Indeed, the number of possible structuring elements becomes tremendous! Note however that an *approximate* algorithm based on such principles has been proposed for $\lambda \leq 8$ [1]. It is however still very slow and inaccurate, and the constraint $\lambda \geq 8$ does not leave enough filtering power for most applications.

The algorithm developed for this study is based on theorem 10 and Corollary 7. Its first step is to extract (and label) the regional maxima of image I under study (for this step, refer to [15, 18, 2]). Then, to each maximum are progressively added its neighboring pixels, starting with those with largest value (in other words, the local threshold around the maximum is progressively lowered). As soon as the area of the current broadened maximum becomes larger than λ, the process stops and value v is assigned to all pixels of the broadened maximum. The next maximum is then considered, etc. Implementation of this procedure on a *Sun Sparc Station 2* allows us to compute area openings of size 100 on a 256×256 image in less than 3 seconds on average! Adapting it to area closings is straightforward.

4 Applications

4.1 Grains Image Filtering Problem Revisited



(a) (b)

Fig. 3. (a) After dual reconstruction of Fig. 1a from Fig. 2c. (b) Area closing of Fig. 1a.

It was mentioned in the introduction that *grey-scale reconstruction* helps in this image filtering task. As illustrated by Fig. 3a, dual reconstruction (refer to [15, 16, 17]) of Fig. 1a from the minimum of closings of Fig. 2c yields a very clean image. However, the area closing introduced in this paper performs even better: Fig. 3b represents an area closing of size 40 of Fig. 1a.

One can see that, while the overall cleanliness is relatively similar in Fig. 3a and Fig. 3b, the latter does a better job of preserving the inter-grain separations, especially those whose orientation is not one of the four orientations used in the original minimum of closings with line segments. This is illustrated by Fig. 4, which is the thresholded algebraic difference between Fig. 3a and Fig. 3b. Note that on the contrary to classic morphological openings and closings, both the

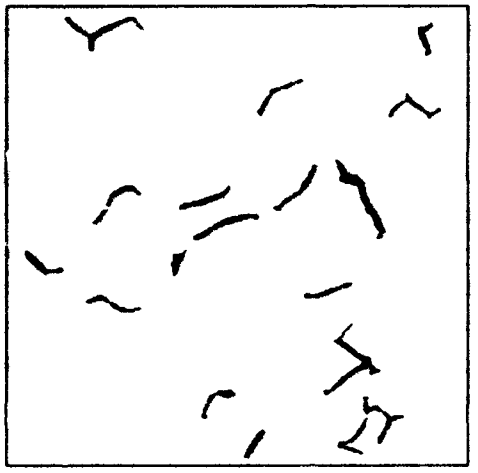
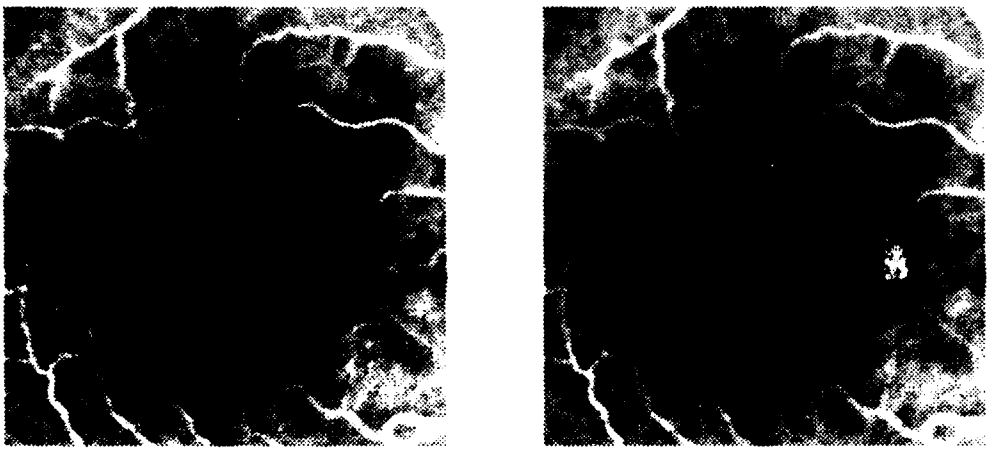


Fig. 4. Areas where the area closing performs substantially better than the filter of Fig. 3a at preserving the thin dark lines between grains.

reconstruction method and the present area openings/closings yield filtered images where no roughness due to the shape of the chosen structuring elements may be observed.

4.2 Use for Image Segmentation

Just as with classic openings and closings, one can very well perform top-hats [8] with area openings and closings. This allows the straightforward extraction of small light or dark structures regardless of their shape. As an example, let us consider Fig. 5a, an image of eye blood vessels where microaneurisms have to be detected. These are small light structures which are

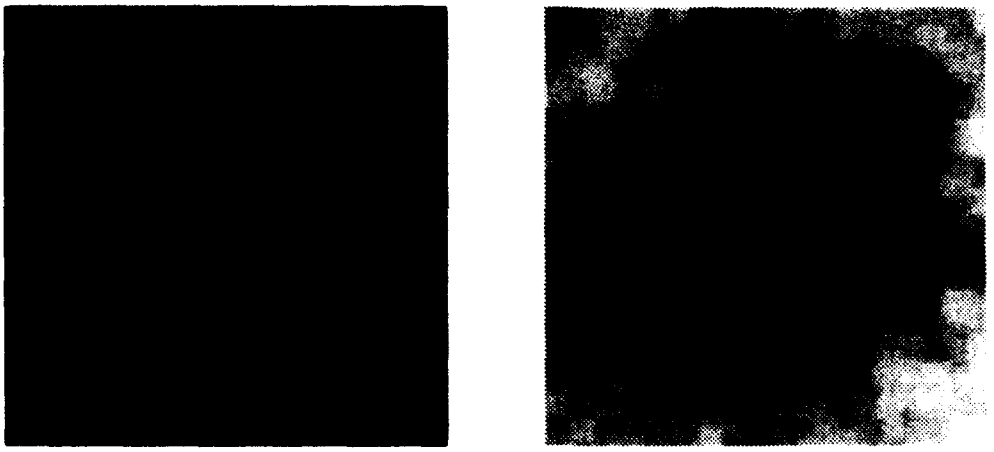


(a) (b)

Fig. 5. (a) Original image (angiography) of eye blood vessels with microaneurisms; (b) area opening of size 60.

- disconnected from the network of the blood vessels,
- predominantly located on the dark areas of the image, i.e. here, the central region.

A direct area opening of size larger than any possible aneurism yields Fig. 5b and its subtraction from the original image (area top-hat) is shown in Fig. 6a. The aneurisms are clearly visible but some other small structures not located on



(a) (b)

Fig. 6. (a) Pixelwise algebraic difference between Fig. 5a and Fig. 5b (area top-hat); (b) morphological opening of Fig. 5a by a large square.

the dark image areas are also present.

Now, by computing an opening of Fig. 5a with respect to a large square, we basically remove all the light structures and end up with an image of the "background" (see Fig. 6b). After inverting this image and computing a pixelwise multiplication of the result with Fig. 6a, we get Fig. 7a where the aneurisms really stand out. A simple thresholding of this image then provides an accurate detection (see Fig. 7b). Note that an alternative solution to this microaneurism detection problem is given in [16, 17].

4.3 Area Alternating Sequential Filters, Hierarchical Image Decomposition

Having a fast area opening/closing algorithm at our disposal allows us to use these transformations in more complex filters. In particular, since the $\{\gamma_\lambda^a\}_{\lambda\in\mathbb{N}}$ and the $\{\phi_\lambda^a\}_{\lambda\in\mathbb{N}}$ obviously constitute a size distribution and an anti-size distribution [7], they can be used in alternating sequential filters (ASF) [14, 11, 12].

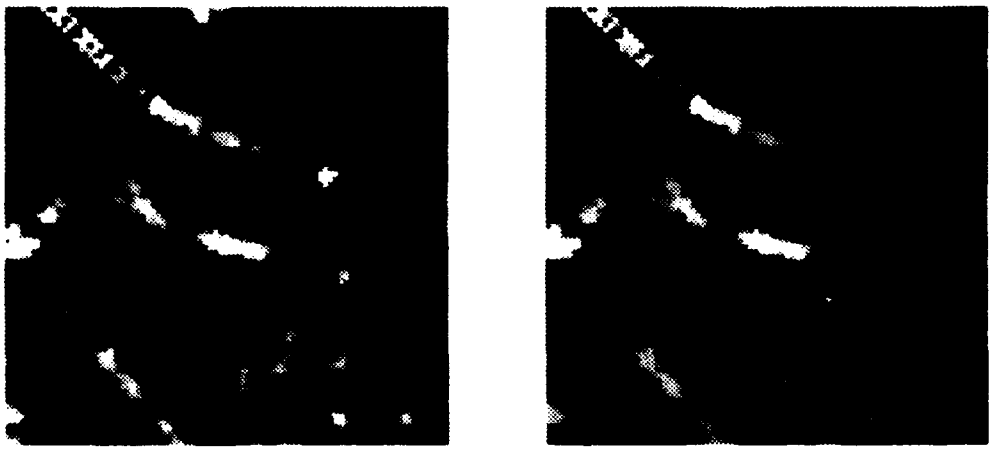
In most practical cases however, there is almost no difference between the following ASF

$$\phi_k^a \circ \gamma_k^a \circ \phi_{k-1}^a \circ \gamma_{k-1}^a \circ \cdots \circ \phi_1^a \circ \gamma_1^a$$

(a) (b)

Fig. 7. (a) pixelwise multiplication of inverted image 6b with Fig. 6a; (b) microaneurisms detected after straightforward thresholding.

and the simple open-close filter $\phi_k^a \circ \gamma_k^a$! (This statement would be wrong in the case of weird nested structures.) Besides, the latter is also extremely close to the close-open filter $\gamma_k^a \circ \phi_k^a$. It filters darks and lights equally well and is very good at removing impulse noise while preserving the shape of the underlying image structures, as illustrated by Fig. 8.



(a) (b)

Fig. 8. (a) A radar image with impulse noise and speckle; (b) its area open-close filter of size 9.

With increasing sizes of area ASF (or simply open-close), one progressively gets images with more and more flat "plateau" areas, originally corresponding to minima and maxima. As illustrated by Fig. 9, this process produces a series of images of decreasing complexity (or level of detail) and could therefore be

used in a hierarchical image decomposition scheme.

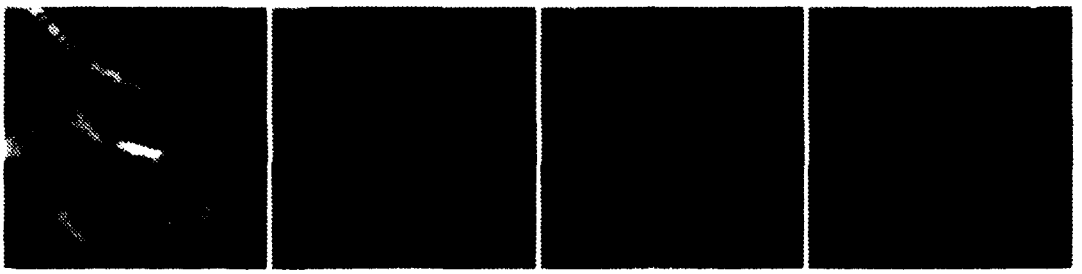


Fig. 9. Area open-close filters of increasing size of Fig. 8a.

5 Conclusions

In this paper, grey-scale morphological *area openings and closings* have been introduced, and their properties studied. It has been proved that the area opening of size λ is equivalent to a supremum of morphological openings with connected structuring elements of area $\geq \lambda$. We conjecture that, in fact, this is true with connected structuring elements of area *exactly* equal to λ. This latter result is true anyway in the discrete case and establishes the connectivity-preserving behavior of these openings and closings. It has been showed that these operators are ideal for many difficult image filtering tasks. Moreover, they can be of great interest in image segmentation and decomposition applications. A fast algorithm derived from the results of this paper has been outlined and will be detailed in future publications. Hopefully these new area openings and closings will be useful for solving a variety of image analysis problems.

References

1. Cheng, F., Venetsanopoulos, A. N. (1991). Fast, adaptive morphological decomposition for image compression, Proc. 25th Annual Conf. on Information Sciences and Systems, pp. 35–40.
2. Grimaud, M. (1992). A new measure of contrast: dynamics, Proc. SPIE Vol. 1769, Image Algebra and Morphological Processing III, San Diego CA.
3. Knuth, D.E. (1973). The Art of Computer Programming, Vol. 3 : Sorting and Searching, Addison Wesley.
4. Lantuéjoul, Ch., Maisonneuve, F. (1984) Geodesic methods in image analysis, Pattern Recognition 17, pp. 117–187.
5. Laÿ, B. (1987). Recursive Algorithms in Mathematical Morphology, Acta Stereologica, Vol. 6/III, Proc. 7th Int. Congress For Stereology, pp. 691–696.
6. Maragos, P., Schafer, R.W. (1987). Morphological filters—part II : their relations to median, order-statistics, and stack filters, IEEE Trans. on Acoustics, Speech, and Signal Processing 35 (8), pp. 1170–1184.

7. Matheron, G. (1975). Random Sets and Integral Geometry, J. Wiley & Sons, New York.
8. Meyer, F. (1979). Iterative image transformations for the automatic screening of cervical smears, J. Histochem. and Cytochem. 27, pp. 128–135.
9. Meyer, F. (1990). Algorithme ordonné de ligne de partage des eaux, Tech. Report CMM, School of Mines, Paris.
10. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.
11. Serra, J. (ed.) (1988). Image Analysis and Mathematical Morphology, Part II: Theoretical Advances, Academic Press, London.
12. Serra, J., Vincent, L. (1992). An overview of morphological filtering, Circuits, Systems, and Signal Processing 11 (1), pp. 47–108.
13. Soille, P., Serra, J., Rivest, J-F. (1992). Dimensional measurements and operators in mathematical morphology, Proc. SPIE Vol. 1658 Nonlinear Image Processing III, pp. 127–138.
14. Sternberg, S.R. (1986). Grayscale morphology, Computer Vision, Graphics, and Image Processing 35, pp. 333–355.
15. Vincent, L. (1990). Algorithmes Morphologiques à Base de Files d'Attente et de Lacets. Extension aux Graphes, PhD dissertation, School of Mines, Paris.
16. Vincent, L. (1992). Morphological grayscale reconstruction; definition, efficient algorithm, and applications in image analysis, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Champaign IL, pp. 633–635.
17. Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms, IEEE Trans. on Image Processing, 2, pp. 176–201.
18. Vincent, L. (1992). Morphological algorithms. In: Dougherty, E. (ed.), Mathematical Morphology in Image Processing, Marcel-Dekker, New York.
19. Wendt, P.D., Coyle, E.J., Gallagher, N.C. (1986). Stack filters, IEEE Trans. on Acoustics Speech, and Signal Processing 34 (4), pp. 898–911.

Manifold Shape: from Differential Geometry to Mathematical Morphology

Jos B.T.M. Roerdink

Department of Computing Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

Abstract. Much progress has been made in extending Euclidean mathematical morphology to more complex structures such as complete lattices or spaces with a non-commutative symmetry group. Such generalizations are important for practical situations such as translation and rotation invariant pattern recognition or shape description of patterns on spherical surfaces. Also in computer vision much use is made of spherical mappings to describe the world as seen by a human or machine observer. Stimulated by these developments the question is studied here of the shape description of patterns on arbitrary (smooth) surfaces based on mathematical morphology. The primary interest in this paper is to outline the mathematical structure of this description. Some concepts of differential geometry, in particular those of parallel transport and covariant differentiation, are used to replace the more restricted concept of invariance used so far in mathematical morphology. The corresponding morphological operators which leave the geometry on the surface invariant are then constructed.

Keywords: mathematical morphology, differential geometry, parallel transport, dilation, erosion, closing, opening, shape concepts, group invariance.

1 Introduction

Much progress has been made in extending Euclidean mathematical morphology as developed by Matheron and Serra [7, 12] to more complex structures such as complete lattices [13, 3, 11] or spaces with a non-commutative symmetry group [9, 10]. Such generalizations are important for practical situations like translation and rotation invariant pattern recognition or shape description of patterns on spherical surfaces (satellite data of the earth, microscopic images of virus particles, etc.). Also in computer vision and image understanding there is increasing use of group theoretical methods [4]. Stimulated by these developments the question is studied here of the shape description of patterns on arbitrary (smooth) surfaces based on mathematical morphology. The primary interest in this paper is to outline the mathematical structure of this description. It is clear that human observers are able to recognize patterns on a curved surface (say,

patterns on ceramics) as "similar". This notion is quantified by introducing some concepts of differential geometry, in particular those of parallel transport and covariant differentiation which can be used to replace the more restricted concept of invariance groups used so far in mathematical morphology. When using a geometric concept of shape in line with F. Klein's Erlanger Programm [6] such invariance concepts form an essential ingredient in shape descriptions. Next, the corresponding morphological operators which leave the geometry on the surface invariant will be constructed. In view of the fact that both differential geometry and mathematical morphology start from *local* operations it is perhaps not too surprising that a connection between the two can be established. The present paper presents a first step in this direction.

The organization of this paper is as follows. In Sect. 2 a number of prerequisites are stated from mathematical morphology, with particular emphasis on symmetry properties and their role in shape description. Then in Sect. 3 the required differential geometric concepts are briefly introduced. The study is mainly restricted to smooth (hyper)surfaces in n-dimensional Euclidean space, although most of the results carry over to more general Riemannian manifolds as well. Finally, in Sect. 4 these differential geometric concepts are used to construct morphological operators on smooth surfaces which leave the geometry of the surface invariant. The results presented here are of a preliminary nature. Both the mathematical treatment and the question of the usefulness of the approach outlined here require a more detailed study.

2 Invariance Concepts in Mathematical Morphology

In [10] a study was made of a *homogeneous space* (Γ, M); that is, a set M on which a transitive but not necessarily commutative group Γ of invertible transformations is defined. Here *transitive* means that for any pair of points in the set there is a transformation in the group which maps one point on the other. If this mapping is unique the transformation group is called *simply transitive* [15]. Each element $g \in \Gamma$ is a mapping $M \to M : x \mapsto g(x)$, satisfying

$$\text{(i)} \quad gh(x) = g(h(x)) \ , \quad \text{(ii)} \quad e(x) = x \ ,$$

where e is the unit element of Γ (i.e. the identity mapping $x \mapsto x$, $x \in M$), and gh denotes the product of two group elements g and h. The *inverse* of an element $g \in \Gamma$ will be denoted by g^{-1}. Usually we will also write gx instead of $g(x)$. The *stabilizer* of $x \in M$ is the subgroup $\Gamma_x := \{g \in \Gamma : gx = x\}$. The object space by which binary images on M are modelled is the Boolean lattice $\mathcal{P}(M)$ of all subsets of M, ordered by set inclusion. A brief sketch will be given of the construction of morphological operations on this homogeneous space with full invariance under the acting group Γ.

First recall the construction of dilations on $\mathcal{P}(M)$ without any invariance property, as given by Serra [13, Ch.2, Proposition 2.1] (a dilation/erosion is a mapping commuting with unions/intersections):

Proposition 1. *A mapping $\delta : \mathcal{P}(M) \to \mathcal{P}(M)$ is a dilation if and only if there exists a function $\gamma : M \to \mathcal{P}(M)$, called a "structuring function", such that*

$$\delta(X) = \bigcup_{x \in X} \gamma(x) \ . \tag{1}$$

This statement can be interpreted as follows. Attach to each point x of M a subset $\gamma(x)$ of M, that is, think of M as being completely "covered" by a collection of subsets of itself. Then the dilation $\delta(X)$ is the union of all the subsets which are attached to points of X. On the complete lattice $\mathcal{L} = \mathcal{P}(M)$, there exists for every dilation $\delta : \mathcal{L} \to \mathcal{L}$ a unique erosion $\varepsilon : \mathcal{L} \to \mathcal{L}$, called the *adjoint* of δ, such that (ε, δ) is an adjunction. Here a pair (ε, δ) of mappings on $\mathcal{L}$ is called an *adjunction* if for every $X, Y \in \mathcal{L}$ the following equivalence holds:

$$\delta(X) \leq Y \iff X \leq \varepsilon(Y) \ ;$$

see [3, 11]. It is easy to see that the erosion ε associated (by adjunction) with δ is given by

$$\varepsilon(X) = \{y \in M : \gamma(y) \subseteq X\} \ . \tag{2}$$

Next morphological operators which possess invariance properties under a transformation group are considered. Assume that M is a homogeneous space under a group Γ which acts transitively on M. In that case it is appropriate to take a *fixed* set A (the "structuring element") and attach to any $x \in M$ *all* the sets $gA := \{ga : a \in A\}$, where g runs over the complete collection of group elements which move a fixed point ω (called the "origin") to x. The set gA is sometimes referred to as the *(group) translate of A (by g)*.

Example. The translation group on the plane
Consider the plane $M = \mathbf{R}^2$, acted upon by the commutative group $\mathbf{T}$ of translations. This is the classical case [12, 7]. Here one uses translates $\tau_x(A) = \{x + a : a \in A\}$ of a single set A, where τ_x is the unique (Euclidean) translation which maps the origin to the point x.

Example. The translation-rotation group on the plane
Consider the plane $M = \mathbf{R}^2$, acted upon by the group Γ generated by translations and rotations of the plane (Euclidean motion group, group of rigid motions), a noncommutative group. Let the origin ω be the point $(0, 0)$. The stabilizer Σ is equal to the group $\mathbf{R}$ of rotations around the origin. The collection of all group elements which map ω to x is the set $\{\tau_x s : s \in \Sigma\}$, where τ_x is the unique translation which maps the origin to the point x (see the previous example). The basic objects in defining morphological operations with respect to this group are formed by all translated and rotated copies of a single set A. An application which occurs in the problem of motion planning for robots has been considered in great detail in [9].

Example. The rotation group on the sphere ([8])
Consider the unit 2-sphere $M = S^2$, acted upon by the three-dimensional rotation group $\Gamma = SO(3)$, also a noncommutative group. Let $\mathbf{v} = (x, v)$, $v \in S^1$,

be a unit tangent vector at $x \in S^2$. Choose the north pole $\mathcal{N}$ as the origin of the sphere, and define a base-vector b to be an arbitrary unit tangent vector at $\mathcal{N}$. Then the tangent vector v represents a unique rotation, i.e. the one which maps b to v. The stabilizer Σ is the set of rotations around the north pole $\mathcal{N}$. For a fixed set $A \subset S^2$ the set $\{g_x sA : x \in S^2, s \in \Sigma\}$, where for each $x \in S^2$, g_x is any particular group element (a representative) which maps ω to x, forms the basic collection from which morphological operations are constructed.

2.1 Morphological Operators

In the classical case, one uses Euclidean translations to define dilations and erosions.

$$\textit{Minkowski addition}: \quad X \oplus A = \{x + a : x \in X, a \in A\} = \bigcup_{a \in A} X_a \ , \tag{3}$$

$$\textit{Minkowski subtraction}: \quad X \ominus A = \bigcap_{a \in A} X_{-a} \ , \tag{4}$$

where

$$X_a = \tau_a(X) = \{x + a : x \in X\} \ .$$

Next consider the case of a homogeneous space (Γ, M). First the following definition is needed.

Definition 2. Let (Γ, M) be a homogeneous space with Σ the stabilizer of the origin ω in M. A subset X of M is called *Σ-invariant* if $X = \overline{X}$, where $\overline{X} = \bigcup_{s \in \Sigma} sX$. If X is not Σ-invariant, $\overline{X}$ is called the *Σ-invariant extension* of X.

Let $A \subset M$. Then the mapping

$$\delta(X) := \bigcup_{x \in X} \bigcup_{\{g \in \Gamma : g\omega = x\}} gA \tag{5}$$

is a dilation δ which is Γ-invariant; that is, $\delta(gX) = g\delta(X)$ for all $g \in \Gamma$, $X \in \mathcal{P}(M)$. Moreover, all Γ-invariant dilations are of this form [10]. Using Definition 2, (5) can be rewritten as

$$\delta(X) = \bigcup_{x \in X} \bigcup_{s \in \Sigma} g_x sA = \bigcup_{x \in X} g_x \overline{A} \ , \tag{6}$$

where, again for each x, g_x is any particular group element which maps ω to x. The adjoint erosion of (5) is formed by associating with a subset X the collection of points $y \in M$ such that $gA \subseteq X$ for *all* $g \in \Gamma$ which move the origin to y. For a representation of this erosion as an intersection of translated sets, see [10]. This shows that any Γ-invariant dilation on $\mathcal{P}(M)$ can be reduced to a dilation δ_A^Γ involving a *Σ-invariant* structuring element A; the same is true for erosions. *Openings* by a subset A of M can be defined by

$$\gamma_A^\Gamma(X) = \bigcup_{g \in \Gamma} \{gA : gA \subseteq X\} \ , \tag{7}$$

which is the union of all translates gA of A which are included in X. Such Γ-openings are generally not reducible to openings by a Σ-invariant structuring element, see [10].

2.2 The Role of Symmetry Groups in Shape Description

Usually, "shape" is defined as referring to those properties of geometrical figures which are invariant under the Euclidean similarity group [5]. Intuitively speaking, one first has to bring figures to a standard location, orientation and scale before being able to "compare" them. The following definition generalizes this.

Definition 3. Let M be a set, Γ a group acting on M. Two subsets X, Y of M are said to have *the same shape with respect to* Γ, or *the same* Γ*-shape*, if they are Γ-equivalent, meaning that there is a $g \in \Gamma$ such that $Y = gX$. If no such $g \in \Gamma$ exists, X and Y are said to have *different* Γ*-shape*.

In essence this definition goes back to F. Klein's "Erlanger Programm" (1872), which considers geometry to be the study of transformation groups and the properties invariant under these groups [6]. In Euclidean morphology, all translates of a set X by the Euclidean translation group T have the same T-shape. After adding rotations to obtain the Euclidean motion group M, rotated versions of X or its translates have the same M-shape as X.

This notion of shape is still too restricted in the case of sets on arbitrary surfaces M, for in general no group Γ exists which acts transitively on M. Therefore a more general definition of shape equivalence will be sought using a number of concepts from differential geometry. To motivate this whole enterprise, a simple but important example of a morphological operation on an arbitrary surface will first be given.

2.3 Motivating Example

Let M be a smooth surface in $\mathbb{R}^3$ supplied with the induced metric. That is, lengths are measured "along the surface". For any $x \in M$, let $D_r(x)$ be the disk of radius r centred at x. Then a dilation $\delta : \mathcal{P}(M) \to \mathcal{P}(M)$ can be defined as follows:

$$\delta(X) = \bigcup_{x \in X} D_r(x) \ . \tag{8}$$

That is, $\delta(X)$ is the union of all points of M with distance smaller than r to some point of X. Comparing with the cases studied above one could say that a disk of radius r is used here as the "structuring element". In the same way one can define erosions or openings by a disk of radius r. The problem described in the rest of the paper essentially boils down to the question of how this can be generalized when the structuring element is not a disk.

3 Elementary Concepts from Differential Geometry

A brief outline is given here of some background material on Riemannian manifolds. Since it is the aim to develop in the next section some new concepts in a way which is intuitively clear, the discussion here is mainly restricted to the case of smooth surfaces in Euclidean 3-space. Thorpe [16] is followed as far as terminology and notation is concerned. For results in a more abstract setting see Boothby [1] or Helgason [2].

A *surface of dimension* n, or *n-surface*, in $\mathbb{R}^{n+1}$ is a non-empty subset M of $\mathbb{R}^{n+1}$ of the form $M = f^{-1}(c)$ where $f : U \to \mathbb{R}$, U open in $\mathbb{R}^{n+1}$, is a smooth function with the property that $\nabla f(p) \neq 0$ for all $p \in M$, and $c \in \mathbb{R}$, where ∇ is the gradient operator. That is, M is a *level set* of f at *height* c. The gradient property implies that all points of M are *regular*. A *vector* at a point $p \in \mathbb{R}^{n+1}$ is a pair $\mathbf{v} = (p, v)$ where $v \in \mathbb{R}^{n+1}$. The set of all vectors tangent to M at p equals $[\nabla f(p)]^{\perp}$ and is called the *tangent space* of M at p, denoted by M_p in the following. A parameterized curve, or "curve" for short, in M is a smooth function $\alpha : I \to M$ where I is some open interval in $\mathbb{R}$. The space M_p consists of velocity vectors at p of all curves passing through p and is an n-dimensional vector subspace of the space of all vectors at p. The disjoint union of all tangent spaces,

$$T(M) = \bigcup_{p \in M} M_p \ ,$$

is called the *tangent bundle* of M.

A *geodesic* in M is a parameterized curve $\alpha : I \to M$ whose acceleration is everywhere orthogonal to M, that is, $\ddot{\alpha}(t) \in M^{\perp}_{\alpha(t)}$ for all $t \in I$. Geodesics have constant speed,

$$\frac{d}{dt} \|\dot{\alpha}(t)\|^2 = \frac{d}{dt} \langle \dot{\alpha}, \dot{\alpha} \rangle = 2 \langle \dot{\alpha}, \ddot{\alpha} \rangle = 0 \ ,$$

since $\dot{\alpha} \in M_{\alpha(t)}$ and $\ddot{\alpha}(t) \in M^{\perp}_{\alpha(t)}$ (angular brackets denote inner products). Given any $p \in M$ and any $\mathbf{v} \in M_p$ there exists a geodesic passing through p with velocity $\mathbf{v}$ at p. When the domain I of α is chosen as large as possible, the resulting geodesic is called a *maximal geodesic*. For each $p \in M, \mathbf{v} \in M_p$, there is a *unique* maximal geodesic α with $\alpha(0) = p$, $\dot{\alpha}(0) = \mathbf{v}$. For example, on the sphere S^2 the geodesics are great circles (or single points); on the cylinder they are straight lines, circles, or spirals (or points). An n-surface M is said to be *geodesically complete* if every maximal geodesic in M has domain $\mathbb{R}$. For example, the n-sphere is geodesically complete, the n-sphere with north pole deleted is not.

3.1 Parallel Transport

In Euclidean space one knows how to transport vectors from one point to another by using the operation of translation. On an n-surface a comparable operation can be defined, which is called *parallel transport* or *parallel translation*. The concepts are developed in a few steps.

A vector field $\mathbf{X}$ on $U \subset \mathbb{R}^{n+1}$ is a function $p \to \mathbf{X}(p) = (p, X(p))$ for some function $X : U \to \mathbb{R}^{n+1}$. A vector field $\mathbf{X}$ is smooth if the components of the function $X : U \to \mathbb{R}^{n+1}$ are all smooth, that is, have continuous partial derivatives of all orders. A vector field along the curve $\alpha : I \to M$ is a function $t \mapsto \mathbf{X}(t)$ where $\mathbf{X}(t) \in \mathbb{R}^{n+1}$; if $\mathbf{X}(t) \in M_{\alpha(t)}$ for all $t \in I$, the vector field is called *tangent to M along α*. If $\mathbf{X}$ is a tangent vector field along a curve $\alpha : I \to M$ then the derivative $\dot{\mathbf{X}}$ is generally not tangent to M. To obtain a vector field tangent to M one has to project $\dot{\mathbf{X}}$ orthogonally onto $M_{\alpha(t)}$. This operation is called *covariant differentiation* and the resulting vector field

$$\mathbf{X}'(t) = \dot{\mathbf{X}} - \langle \dot{\mathbf{X}}(t), \mathbf{N}(\alpha(t)) \rangle \mathbf{N}(\alpha(t)) \ ,$$

where

$$\mathbf{N}(p) = \frac{\nabla f(p)}{\|\nabla f(p)\|} \ , \quad p \in M \ ,$$

is a unit normal vector field, is called the covariant derivative of $\mathbf{X}$. The covariant derivative measures the rate of change of $\mathbf{X}$ as seen from the surface M. A curve $\alpha : I \to M$ is a geodesic if and only if the covariant acceleration $(\dot{\alpha})'$ is zero along α. A smooth vector field $\mathbf{X}$ tangent to M along α is called *constant* or *(Levi-Civita) parallel* if $\mathbf{X}' = 0$. If $\mathbf{X}$ and $\mathbf{Y}$ are parallel vector fields along α, then $\langle \mathbf{X}, \mathbf{Y} \rangle' = 0$, so $\langle \mathbf{X}, \mathbf{Y} \rangle$ is constant along α. In particular, $\mathbf{X}$ and $\mathbf{Y}$ have constant length; therefore, also the angle between $\mathbf{X}$ and $\mathbf{Y}$ is constant. For an example see Fig. 1. The velocity vector field along a parameterized curve in M

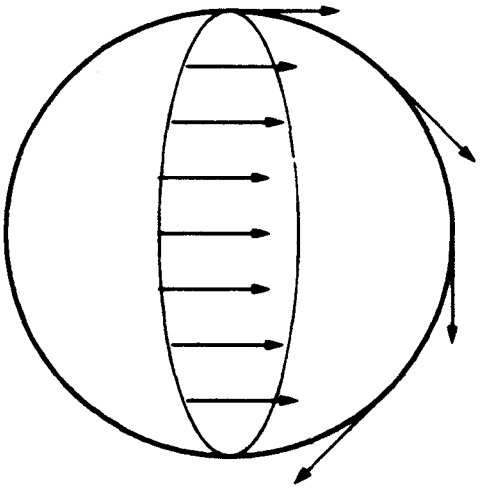


Fig. 1. Parallel vector fields along geodesics in the 2-sphere.

is parallel if and only if α is a geodesic. The following theorem is fundamental [16]:

Theorem 4. *Let M be an n-surface in $\mathbb{R}^{n+1}$, $\alpha : I \to M$ a parameterized curve in M, let $t_0 \in I$ and $\mathbf{v} \in M_{\alpha(t_0)}$. Then there exists a unique vector field $\mathbf{V}$, tangent to M along α, which is parallel and has $\mathbf{V}(t_0) = \mathbf{v}$.*

A simple characterization is possible for a 2-surface M: a vector field tangent to M along a geodesic α is parallel if and only if both $\|\mathbf{X}\|$ and the angle between $\mathbf{X}$ and $\dot{\alpha}$ are constant along α.

Parallelism can be used to transport tangent vectors from one point of an n-surface to another.

Definition 5. Let $p, q \in M$ and let $\alpha : [a, b] \to M$ be a parameterized curve from $\alpha(a) = p$ to $\alpha(b) = q$. For $\mathbf{v} \in M_p$ let $\mathbf{V}$ be the unique parallel vector field along α with $\mathbf{V}(a) = \mathbf{v}$. The map $P_\alpha : M_p \to M_q$ determined by

$$P_\alpha(\mathbf{v}) = \mathbf{V}(b)$$

is called *parallel transport* from p to q, and $P_\alpha(\mathbf{v})$ the *parallel translate* of $\mathbf{v}$ along α to q.

Parallel transport from p to q is path dependent: if α and β are two curves from p to q then, in general, $P_\alpha(\mathbf{v}) \neq P_\beta(\mathbf{v})$ (an exception occurs for surfaces of zero curvature, such as the Euclidean plane). More precisely, $P_\beta(\mathbf{v})$ differs from $P_\alpha(\mathbf{v})$ by a rotation around the normal to M at q. When a vector in M_p is transported along a closed curve beginning and ending in p, it will carry out a rotation in M_p. The set of such rotations of M_p generated by parallel translation along closed curves is called the *holonomy group* at p. Holonomy groups at different points of M are isomorphic.

The following result will be needed later [16]:

Theorem 6. *Let M be an n-surface in $\mathbb{R}^{n+1}$, $p, q \in M$ and α a piecewise smooth curve from p to q. Then parallel translation $P_\alpha : M_p \to M_q$ along α is a vector space isomorphism which preserves inner products:*

1. *P_α is a linear map;*
2. *P_α is 1-1 and onto;*
3. *$\langle P_\alpha(\mathbf{v}), P_\alpha(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{v}, \mathbf{w} \in M_p$.*

To study questions concerning lengths of curves in M, it is convenient to parameterize curves by arc length, that is, choose a reparameterization such that α has unit speed. A well-known result concerning geodesics then asserts that if α is a shortest unit-speed curve from p to q in M, then α is a geodesic (the reverse is in general not true, consider e.g. geodesics on a sphere).

3.2 The Exponential Map

Above we have seen how to transport vectors from one point of an n-surface to another. The next question is how to map points of M to vectors in the tangent bundle $T(M)$. This then will enable us below to transport subsets of M from one point to another.

Definition 7. For $\mathbf{v}$ in the tangent bundle $T(M)$, let $\alpha_{\mathbf{v}}$ denote the unique maximal geodesic in M with $\dot{\alpha}_{\mathbf{v}}(0) = \mathbf{v}$. Let $U = \{\mathbf{v} \in T(M) : 1 \in \text{ domain } \alpha_{\mathbf{v}}\}$. The map $\text{Exp} : U \to M$ defined by $\text{Exp}(\mathbf{v}) = \alpha_{\mathbf{v}}(1)$ is called the *exponential map* of M.

For $p \in M$, we will also write Exp_p to denote the mapping $M_p \to M : \mathbf{v} \mapsto \text{Exp}(\mathbf{v})$. Since geodesics have constant speed, $\text{Exp}_p(\mathbf{v})$ is the point on the unique geodesic determined by $\mathbf{v}$ whose distance from p along the geodesic is precisely $\|\mathbf{v}\|$; cf. Fig. 2. The following theorem summarizes the most important properties of the exponential map [16].

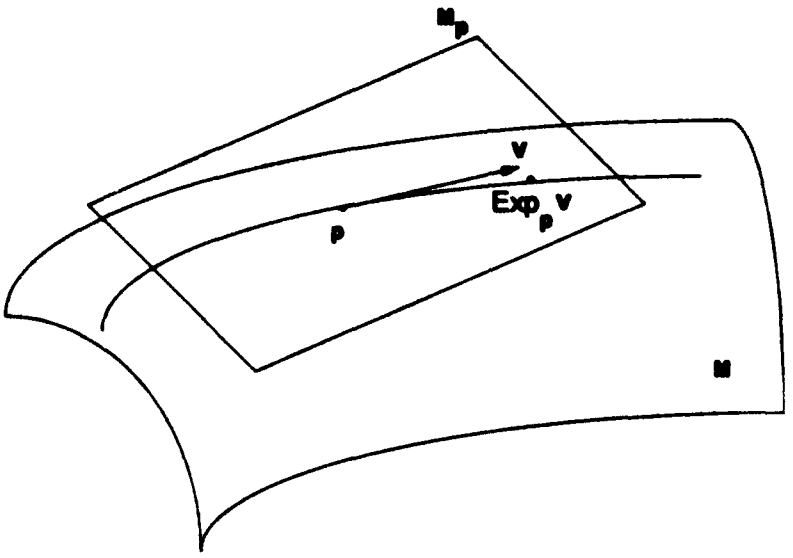


Fig. 2. The exponential map maps a tangent vector $\mathbf{v} \in M_p$ to the point lying at a distance $\|\mathbf{v}\|$ from p on the unique geodesic through p with initial velocity $\mathbf{v}$.

Theorem 8. *The exponential map* $\text{Exp} : U \to M$ *of an n-surface in* R^{n+1} *has the following properties:*

1. *The domain* U *of* Exp *is an open set in* $T(M)$.
2. *If* $\mathbf{v} \in U$ *then* $t\mathbf{v} \in U$ *for* $0 \le t \le 1$.
3. Exp *is a smooth map.*
4. *For each* $p \in M$ *and* $\mathbf{v} \in M_p$, *the maximal geodesic* $\alpha_{\mathbf{v}}$ *with* $\dot{\alpha}_{\mathbf{v}}(0) = \mathbf{v}$ *is given by the formula* $\alpha_{\mathbf{v}}(t) = \text{Exp}_p(t\mathbf{v})$.
5. *For* $\epsilon > 0$ *sufficiently small,* Exp_p *maps the* ϵ*-ball* $B_\epsilon = \{\mathbf{v} \in M_p : \|\mathbf{v}\| < \epsilon\}$ *diffeomorphically onto an open subset* U_ϵ *of* M *containing* p. *For* $q \in U_\epsilon$ *the curve* $\alpha_{\mathbf{v}}(t) = \text{Exp}_p(t\mathbf{v})$ $(0 \le t \le 1)$ *with* $\text{Exp}_p(\mathbf{v}) = q$ *is the* unique *geodesic joining* p *and* q; *it lies in* U_ϵ *and has length shorter than that of any other curve joining* p *and* q.

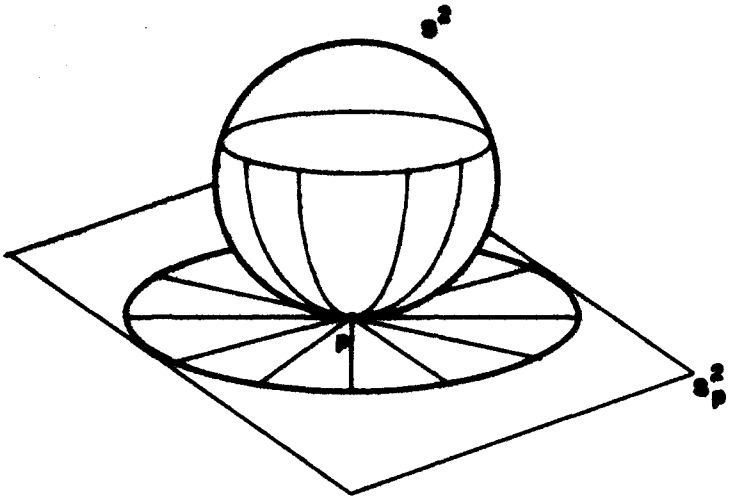


Fig. 3. The geodesics in S^2 through p are the images under the exponential map of the rays through 0 in S_p^2.

This theorem says that geodesics through $p \in M$ are images under Exp_p of the rays $\alpha(t) = t\mathbf{v}$ in M_p; see Fig. 3. In the case of the 2-sphere S^2, Exp_p maps the ball $\{\mathbf{v} \in S_p^2 : \|\mathbf{v}\| < \pi\}$ diffeomorphically onto $S^2 \setminus q$, where q is the antipodal point of p. In M_p, the geodesics through p are the orthogonal trajectories of hypersurfaces

$$\{\mathrm{Exp}_p(\mathbf{v}) : \mathbf{v} \in M_p, \|\mathbf{v}\| = \textit{constant}\} \quad .$$

If $p \in M$, the set $U_\delta(p)$ of points within distance δ of p is called a *spherical neighbourhood* of p or a *disk* of radius δ at p. A neighbourhood $U_\delta(p)$ such that there exists at most (at least) one geodesic segment contained in $U_\delta(p)$ joining any pair of points in $U_\delta(p)$ is called *simple* (*convex*). For sufficiently small δ, any neighbourhood $U_\delta(p)$ is simple and convex [2].

In general the domain and range of the exponential map is restricted. On so-called *geodesically complete* surfaces, such as compact surfaces, every maximal geodesic on M has domain $\mathbb{R}$, that is, can be infinitely extended. In that case the domain of the exponential map is all of $T(M)$. In certain cases the exponential map on a geodesically complete surface maps M_p diffeomorphically *onto* M for any $p \in M_p$, so that a 1-1 correspondence between points of M and points of M_p exists; that is, geodesics between any two points are unique. Examples of such spaces are simply connected geodesically complete surfaces of negative curvature; see Theorem 13.3 of Helgason [2].

Example. The one-sheeted hyperboloid defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \qquad (9)$$

has curvature

$$K = -\frac{1}{a^2b^2c^2}\left(\frac{x^2}{a^4}+\frac{y^2}{b^4}+\frac{z^2}{c^4}\right)^{-2},$$

which is everywhere negative [14]. For a sketch of this (ruled) surface, see Fig. 4.

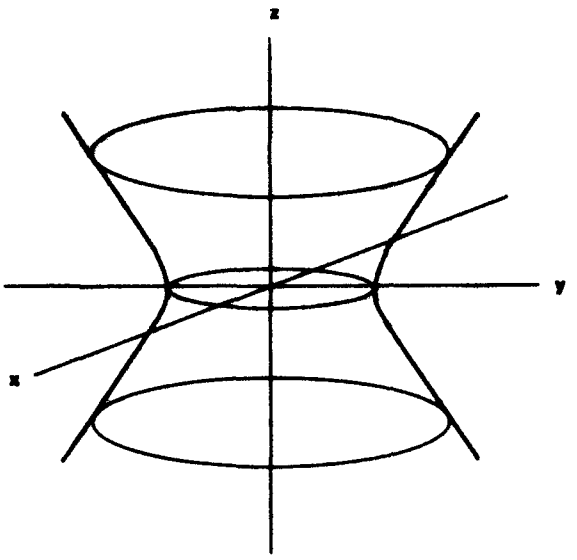


Fig. 4. Sketch of the one-sheeted hyperboloid defined by (9).

4 Mathematical Morphology on Smooth Surfaces

In this section a sketch is given how a morphological description of binary images on smooth surfaces can be developed. The general construction of dilations on complete lattices by Serra (see Prop. 1) holds also, of course, for the special case of the lattice $\mathcal{P}(M)$, where M is a smooth surface in $\mathbb{R}^3$ and $\mathcal{P}(M)$ denotes the set of all subsets of M. The problem is to define morphological operators satisfying some form of invariance. In Sect. 2 we have seen how to handle the case when a transitive group action on M exists. It will be shown that this theory carries over to a large extent to the case when M is an arbitrary surface by replacing group translations by parallel translations, which are based upon the concept of covariant differentiation. The resulting morphological transformations may thus be referred to as "covariant" operations.

The basic problem is how to "transport" subsets of M from one location to another while preserving as many geometric properties as possible. Let X be a neighbourhood of the point $p \in M$. To transport this set X from the point p to another point $q \in M$ we perform the following steps. First map X to the tangent

space M_p by using the inverse of the exponential map: the image under this map is denoted by $\tilde{X}$. Then use parallel translation from p to q along a curve with initial point p and endpoint q. This maps $\tilde{X}$ to a neighbourhood, say $\tilde{Y}$, of q. Finally map $\tilde{Y}$ back to M by the exponential map, thus obtaining a subset Y of q; see Fig. 5.

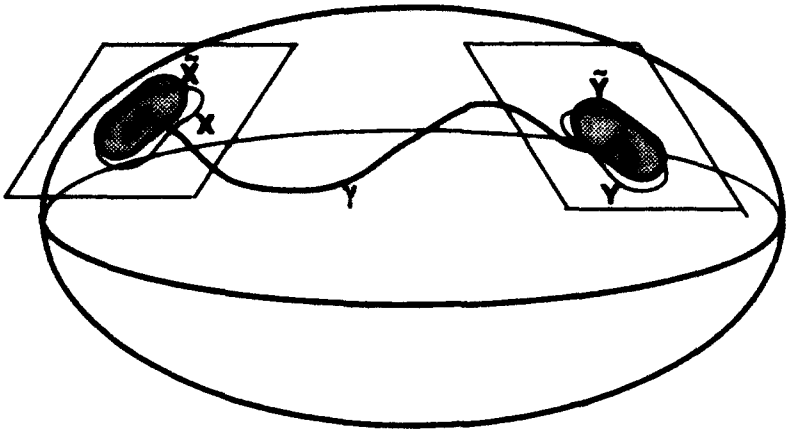


Fig. 5. Parallel transport of subsets of a surface.

To formalize this, let $\gamma = \gamma_{[p,q]}$ be a curve from p to q. Then an operator τ_γ can be defined by

$$Y = \tau_\gamma(X) = \mathrm{Exp}_q P_\gamma \mathrm{Exp}_p^{-1}(X) \ , \tag{10}$$

where P_γ is the parallel transport of tangent vectors from M_p to M_q along γ (see Definition 5) and $P_\gamma(\tilde{X})$ is simply the union of all translated vectors $P_\gamma \mathbf{v}$ when $\mathbf{v}$ runs over $\tilde{X} = \mathrm{Exp}_p^{-1}(X)$. By transporting the initial set X at a fixed point ω along all possible curves to other points of M we cover M by an infinite collection of diffeomorphic copies of X, which in addition preserve several metrical properties (lengths, angles of tangent vectors). It may be verified that the operation (10) reduces to Euclidean translation when M is a plane, and to rotation in the case of the sphere (in the latter case one has to take for X a subset of the sphere not containing the antipodal point of ω in order for $\mathrm{Exp}_\omega^{-1}(X)$ to be well defined).

The following points should now be made. First, the exponential map is in general only invertible (in fact, a diffeomorphism) for a sufficiently small neighbourhood of the origin in M_p, although on some manifolds the inverse exists for arbitrary neighbourhoods of the origin in M_p, so that there is a 1-1 correspondence between the neighbourhoods of a point $p \in M$ and the neighbourhoods of the point $\mathbf{0} \in M_p$; see the example at the end of Sect. 3. Therefore we will take as the basic "structuring element" a subset, not of M but of the tangent space at a given point ω of M. If $\tilde{A}$ is such a subset of M_ω, then an operator $\tilde{\tau}_\gamma$ (also

referred to as "parallel translation along γ") can be defined by

$$\tilde{\tau}_\gamma(\tilde{A}) = \mathrm{Exp}_p P_\gamma(\tilde{A}) \ , \tag{11}$$

where P_γ is the parallel transport of tangent vectors along a curve γ from ω to p.

Second, the image of the set X under parallel translation from p to q will in general depend on which path is taken. This, however, is a situation which we have already encountered when discussing mathematical morphology on spaces with a noncommutative group action; see Sect. 2. The solution found in that case works here as well: simply consider *all* possible paths from p to q.

Now it is possible to define dilations and erosions. Let $\tilde{A}$ (the "structuring element") be a subset of the tangent space M_ω, ω an arbitrary but fixed point of M. Then define a mapping $\delta : \mathcal{P}(M) \to \mathcal{P}(M)$ by

$$\delta_{\tilde{A}}(X) = \bigcup_{x \in X} \bigcup_{\gamma} \mathrm{Exp}_x P_{\gamma_{[\omega,x]}}(\tilde{A}) \ , \tag{12}$$

where the second union runs over all curves $\gamma = \gamma_{[\omega,x]}$ from ω to x. This can be rewritten as follows. Choose for every $x \in M$ a particular curve ("representative") from ω to x. Let $\tilde{\tau}_x$ denote parallel translation along this particular curve. Then, if Σ is the holonomy group at ω, which for the surfaces in $\mathbb{R}^3$ considered here is simply the group of rotations around the normal at ω, (12) can be written as

$$\delta_{\tilde{A}}(X) = \bigcup_{x \in X} \bigcup_{s \in \Sigma} \tilde{\tau}_x(s\tilde{A}) \ . \tag{13}$$

It is obvious that this mapping is a dilation, either by direct proof or through the invocation of Prop. 1. Since parallel translation commutes with unions—both Exp_p and the vector space isomorphism P_γ do—we also can write (13) in the form

$$\delta_{\tilde{A}}(X) = \bigcup_{x \in X} \tilde{\tau}_x(\overline{A}) \ , \tag{14}$$

where

$$\overline{A} := \bigcup_{s \in \Sigma} s\tilde{A} \tag{15}$$

may be called the *Σ-invariant extension* of $\tilde{A}$. For example, if $\tilde{A}$ is a line segment of length r starting at ω then $\overline{A}$ is a disk of radius r centred at ω. The similarity of these expressions with the results in Sect. 2 is clear. Erosions can be defined in a similar way. If $\overline{A}$ is a Σ-invariant structuring element then the mapping

$$\varepsilon_{\overline{A}}(X) = \{x \in M : \tilde{\tau}_x(\overline{A}) \subseteq X\} \ , \tag{16}$$

is an erosion which extracts all the points x of M such that the parallel translate of $\overline{A}$ from ω to x fits in X.

Openings can also be easily defined, where one does not have to restrict oneself to Σ-invariant structuring elements. For any neighbourhood $\tilde{A}$ in M_ω, let

$$\alpha_{\tilde{A}}(X) = \bigcup_{x \in M} \{\tilde{\tau}_x(\tilde{A}) : \tilde{\tau}_x(\tilde{A}) \subseteq X\} \tag{17}$$

be the union of all parallel translates of $\tilde{A}$ along curves starting at ω which are included in X. It is obvious that this is an opening. Closings can be defined similarly.

Example. This example was already discussed in Sect. 2. Take for $\tilde{A}$ a disk of radius r centred at the origin in M_ω. Then the parallel translate $\tilde{\tau}_x(\tilde{A})$ is a spherical neighbourhood at $x \in M$. The dilation $\delta_{\tilde{A}}(X)$ by $\tilde{A}$ is the union of all points of distance smaller than r to some point of X, and the opening by $\tilde{A}$ extracts from X all spherical neighbourhoods of radius r which fit into X.

Example. Take for $\tilde{A}$ a straight line segment of length L through the origin in M_ω. Then the parallel translates of $\tilde{A}$ are geodesic segments and the opening by $\tilde{A}$ extracts from X all geodesic segments of length L which fit into X.

5 Discussion

In this paper the study of shape description of patterns on arbitrary (smooth) surfaces based on mathematical morphology has been initiated. The main aim has been to give an outline of the mathematical structure of this description based on concepts of differential geometry, in particular those of parallel transport and covariant differentiation which can be used to replace the more restricted concept of invariance groups used so far in mathematical morphology. Various morphological operators have been constructed on a surface M which are defined in terms of neighbourhoods of M which are obtained by parallel translation of a single set $A \subseteq M$ (the "structuring element"). If M is Euclidean space or a sphere then these morphological operations reduce to the known ones which are invariant under the appropriate group (translations, rotations). What has not been discussed here is a precise formulation—in algebraic terms—of the invariance properties satisfied by the operators introduced here for arbitrary surfaces. This is an open problem which requires a more detailed study.

References

1. Boothby, W.M. (1975). An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press, New York.
2. Helgason, S. (1962). Differential Geometry and Symmetric Spaces, Academic Press, New York.
3. Heijmans, H.J.A.M., Ronse, C. (1989). The algebraic basis of mathematical morphology. Part I: dilations and erosions, Computer Vision, Graphics and Image Processing 50, pp. 245–295.
4. Kanatani, K. (1990). Group-Theoretical Methods in Image Understanding, Springer-Verlag, New York.
5. Kendall, D. (1984). Shape manifolds, procrustean metrics, and complex projective spaces, Bull. London Math. Soc. 16, pp. 81–121.
6. Klein, F. (1872). Vergleichende Betrachtungen über neuere geometrische Forschungen, Gesammelte mathematische Abhandlungen, Vol. I, pp. 460–497.

7. Matheron, G. (1975). Random Sets and Integral Geometry, J. Wiley & Sons, New York, NY.
8. Roerdink, J.B.T.M. (1990). Mathematical morphology on the sphere, Proc. SPIE Conf. Visual Communications and Image Processing '90, Lausanne, pp. 263–271.
9. Roerdink, J.B.T.M. (1990). On the construction of translation and rotation invariant morphological operators, Report AM-R9025, Centre for Mathematics and Computer Science, Amsterdam. To appear in: Mathematical Morphology: Theory and Hardware, Haralick, R.M. (ed.), Oxford Univ. Press.
10. Roerdink, J.B.T.M. (1992). Mathematical morphology with non-commutative symmetry groups, Chapter 7 in Mathematical Morphology in Image Processing, Dougherty, E.R. (ed.), pp. 205–254, Marcel Dekker, New York.
11. Ronse, C., Heijmans, H.J.A.M. (1991). The algebraic basis of mathematical morphology, Part II: openings and closings, Computer Vision, Graphics and Image Processing: Image Understanding 54, pp. 74–97.
12. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.
13. Serra, J. (ed.) (1988). Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances, Academic Press, London.
14. Spivak, M. (1979). A Comprehensive Introduction to Differential Geometry, Vol. 3 (2nd ed.), Publish or Perish Inc., Berkeley, CA.
15. Suzuki, M. (1982). Group Theory, Springer-Verlag, Berlin.
16. Thorpe, J.A. (1979). Elementary Topics in Differential Geometry, Springer-Verlag, New York.

On Negative Shape

Pijush K. Ghosh

National Centre for Software Technology, Gulmohar Cross Road No.9, Juhu,
Bombay 400049, India

Abstract. The notion of *negative shape* is that it is an artifice suggested in algebra on various occasions when geometric problems are translated into the language of algebra. Though the notion appears to be very useful, its significance and potential have hardly been explored till this day. In this paper two such cases are considered, namely, an algebraic formulation of Minkowski addition of two geometric objects and an analytic formulation of the area/volume of an object, and it is shown how such formulations indicate the negative shape notion. By means of a *mixed area/volume* concept, it is shown that negative shape notions derived from the two formulations are exactly identical. The usefulness of the negative shape concept is also briefly indicated.

Keywords: shape description, mathematical morphology, slope diagram representation, negative shape, Minkowski addition and decomposition, dilation, group, boundary addition, shape algebra, mixed area, signed area.

1 Introduction

1.1 The Basic Idea

A simple geometric interpretation of negative shape is that it is like a hole without any positive region surrounding the hole. In contrast to this, the shape of any ordinary object in our natural world may be considered *positive*. For any ordinary shape there is a restriction that every hole in the object must be completely surrounded by a positive region. This restriction on shape is clearly imposed by our experience of the natural world. By introducing negative shape we effectively extend the conventional shape domain to a domain where a shape is allowed to be an unrestricted combination of positive regions and holes.

1.2 Some Relevant Questions

Any extension of this sort poses three questions to be answered in precise terms: (a) What is the usefulness, that is, what is the extra advantage of such an

extension? (b) Can we redefine all the necessary operations on this extended set so that they do not contradict the existing definitions of these operations on the restricted set? For example, assume we start with the set of all real numbers and the four arithmetic operations (addition, subtraction, etc.) defined on them. Now if we extend the real number set to the set of all complex numbers, it is necessary to redefine the arithmetic operations on complex numbers in such a way that they agree with the existing definitions when applied to two real numbers. (c) How does such an extension evolve? Is it possible to arrive at the same notion by other means?

In this paper an attempt is made, as far as is possible within the scope of a single paper, to answer these questions.

1.3 The Organization of the Paper

The paper is organized in the following form:

- We begin with a short discussion on the evolution of *negative numbers* in the number system, since there is a close analogy between negative numbers in the number domain and negative shapes in the shape domain. The answers to some of the aforementioned questions for negative shapes can be indirectly obtained if we answer them for negative numbers.
- In the next section (Sect. 3), following the analogy of negative numbers which were evolved in an attempt to solve an equation of the form $x + a = b$ within the set of *natural numbers* $\{0, 1, 2, \ldots, \}$, it is shown that the concept of negative shape evolves naturally if one tries to solve the equation $X \oplus A = B$, where A and B denote two sets of points in the real Euclidean d-dimensional space E^d and $\oplus$ denotes *Minkowski addition* operation (also known as *dilation* in mathematical morphology). Minkowski addition is essentially the vector addition of two set of points,

$$S = A \oplus B = \{a + b \mid a \in A,\ b \in B\} \ ,$$

  where "+" denotes the normal vector addition of two points, and A and B are called the summands of the sum S. Note that the geometric shapes A and B are represented as sets of points, which is one of the most general representation schemes for geometric shapes.
- In Sect. 4 some of the immediate advantages of introducing the notion of negative shape are indicated.
- Minkowski addition is not the only way to arrive at the negative shape notion. In Sect. 5 it is shown that a similar notion also follows from an analytic formulation of the length/area/volume of a geometric shape. Indeed the notion already existed in the mathematical literature from the beginning of the nineteenth century. This is not surprising, because whenever algebraic/analytic tools are employed in solving geometric problems, not only is the task of proving a result in geometry shifted to that of proving a corresponding result in algebra, but something deeper than that happens. In the process the properties of algebraic entities, particularly the properties of

numbers, are also imposed on the geometric concepts, and their geometric interpretations may lead to the discovery of new and unsuspected geometrical results.
- One question may arise at this point. Are the two notions of negative shape – one derived from the concept of Minkowski addition and the other from the length/area/volume consideration – identical? Although this question should be dealt with rigorously, in this paper the question is answered in a simpler way by the indirect means of *mixed area/volume* concept. The topic of mixed area/volume is an appropriate one since it combines the notions of Minkowski addition and the area/volume of geometric shape within a single concept. The discussion in this section also elucidates some of the basic properties and usefulness of negative shape.
- Section 7 is the concluding section where some related problems are posed that are of immediate significance.

2 On Negative Numbers

For the sake of brevity any discussion of the nature, the origin, and the theory of *natural numbers*, as well as of the four fundamental arithmetic operations (addition, subtraction, multiplication, and division) relating them, will be omitted. It may be noted that they originate from our immediate physical experience of the natural world. The reader may compare this state with the present state of geometric shapes.

2.1 Difficulties in Accepting Negative Numbers: A Historical Note

The negative numbers (for that matter, any other numbers, say, rational, algebraic, transcendental numbers) are not very natural. The need for negative numbers was felt quite early, but their acceptance as numbers did not happen easily. Most probably the Hindu mathematicians first introduced negative numbers (around the middle of the seventh century), but with much reservation. Though the Arabs were familiar with negative numbers through the work of the Hindus, they rejected such a concept. Even after a thousand years, most of the European mathematicians of the sixteenth and seventeenth centuries did not accept negative numbers as numbers, or if they did, would not accept them as roots of equations. Such mathematicians included even Pascal and Descartes. Take one typical argument against negative numbers. Antoine Arnauld (1612-94) doubted that $-1 : 1 = 1 : -1$ because, he argued, -1 is smaller than 1; hence, how could a smaller is to a greater be equal to a greater is to a smaller? Even in the eighteenth century, opposition to negative numbers was frequently expressed. As late as 1831 mathematicians like De Morgan insisted that it was absurd to consider numbers less than zero, and it was always possible to avoid them altogether. He illustrated this by means of a simple problem: "At present a father is 56 and his son is 29 years old. When will the father be twice as old as the son?" If we solve, $56 + x = 2(29 + x)$, we obtain $x = -2$. This result,

according to De Morgan, is absurd. He concluded that the original problem was phrased wrongly and thus led to the unacceptable negative answer. It is not intended here to extend the list of such objections, and the reader is referred to Kline [8] for more such details. In short the concept of negative numbers was not well understood until very modern times. (The same is equally true for other kinds of numbers.)

2.2 Genesis of the Negative Number Concept and its Utilities

The basic problem is that, unlike natural numbers, negative numbers lack an immediate physical meaning. From where did they evolve then? Why were they accepted finally? Note that the arithmetic operation addition (and also multiplication) can be carried out for any pair of natural numbers, but this is not the case for its inverse operation subtraction (and division). In order that the difference $b-a$ be defined within natural numbers, it is necessary that b must be greater than a. This is a very troublesome restriction. When we have to solve a problem in which the given quantities depend on the particular case of the problem, then instead of obtaining a solution which can be expressed by a general formula, we have to consider several cases to carry out its complete treatment. This untidiness can be avoided by the introduction of zero and negative numbers.

There were basically three reasons that compelled mathematicians to finally accept negative numbers. The first reason is, of course, the necessity to eliminate the troublesome restriction mentioned above. It is a mathematical stratagem. The second reason is the possibility of utilizing negative numbers to solve concrete problems. For example, to deal with oriented quantities, the best way to distinguish those oriented in one direction from those oriented in the other is to use negative numbers; this is the case with assets and debts, temperatures above or below zero, or dates before or after Christ. The third reason, which is the most decisive one, is the accumulated labours of many mathematicians over the centuries to examine carefully all the operations that could be defined on natural numbers and then to extend/redefine them to both positive and negative numbers. Clearly, this was the most difficult part. The reader will be aware, for example, of the problem that was caused by extending the square root operation to negative numbers.

Once the set of all negative numbers was accepted as a true extension of the set of natural numbers N to form the set of all integers Z, a great mathematical unification was achieved. The mathematicians realized that one could dispense with the *subtraction* operation altogether. In modern terms, the algebraic system $(Z, +)$ is now the best known *Abelian group* where the addition $+$ is the internal composition law. On the other hand, the algebraic system $(N, +)$, that is, the set of all natural numbers under addition, is just a subsystem of $(Z, +)$ and an *Abelian monoid*. In the next section we shall see the implications of these facts.

3 Minkowski Addition and Negative Shape

3.1 Arithmetic Addition vs. Minkowski Addition

Let us turn our attention to the conventional geometric shapes under Minkowski addition operation. We denote this system by $(\mathcal{G}\oplus)$ where $\mathcal{G}$ denotes the set of all geometric shapes (that is, the set of all subsets of E^d; for all practical purposes $d \leq 3$). Note the close resemblance between the two systems $(\mathcal{G}, \oplus)$ and $(N, +)$:

	System $(\mathcal{G}, \oplus)$	System $(N, +)$
1. *Closure:*	If A, B are in $\mathcal{G}$, then $A \oplus B$ is also in $\mathcal{G}$.	If a, b are in N, then $a + b$ is also in N.
2. *Associative:*	For any $A, B, C \in \mathcal{G}$, $A \oplus (B \oplus C) = (A \oplus B) \oplus C$.	For any $a, b, c \in N$, $a + (b + c) = (a + b) + c$.
3. *Identity:*	If $\{o\}$ denotes the origin point of our coordinate system, then $\{o\} \in \mathcal{G}$ and it is the identity element, since for any $A \in \mathcal{G}$, $A \oplus \{o\} = A$.	The number $0 \in N$ is the identity element, since for any $a \in N$, $a + 0 = a$.
4. *Commutative:*	For any $A, B \in \mathcal{G}$, $A \oplus B = B \oplus A$.	For any $a, b \in N$, $a + b = b + a$.
5. *No inverse:*	For all $A \in \mathcal{G}$, there does not exist an element $A^{-1} \in \mathcal{G}$ such that $A \oplus A^{-1} = \{o\}$.	For all $a \in N$, there does not exist an element $a^{-1} \in N$ such that $a + a^{-1} = 0$.

In summary, both the systems $(\mathcal{G}, \oplus)$ and $(N, +)$ are Abelian (commutative) monoids – but not groups.

Since by appending the set of negative numbers (which are the additive inverses of the numbers in N) to the set N we obtain the group $(Z, +)$, it is clear that we have to extend the set $\mathcal{G}$ to include negative shapes (which should be the inverses of the shapes in $\mathcal{G}$) in order to form a group structure with geometric shapes. If this could be done, then it is possible to add and subtract geometric shapes in exactly the way we add and subtract integer numbers. Therefore, the natural question at this point is: How can we define inverse shapes (negative shapes) under Minkowski addition operation?

3.2 In Search of Negative/Inverse Shape

Before proceeding further, let me mention a convention adopted here. Note that, according to the conventional definition of Minkowski addition, the sum $S = A \oplus B$ depends on the choice of origin and the locations of A and B in that coordinate system. This is inconvenient, since we are primarily interested only in the "shapes" of the objects – and not upon their positions relative to an arbitrarily chosen coordinate system. It is easy to show that the shape of the sum S remains invariant under changes of origin and parallel displacement of A and B; under such circumstances S undergoes only a parallel displacement. It is, therefore, more natural to assume that all the translates, say $[A]_T$, of an

object A are equivalent, and, to consider Minkowski addition as operating on translational classes $[A]_T$ of objects instead of on the objects themselves, that is, $[S]_T = [A]_T \oplus [B]_T$.

According to this convention, therefore, the identity element $\{o\}$ is not only the origin point of the coordinate system, but any singleton point set in the space.

Coming back to the question of inverse shapes, we ask the question: Is there any operation already known which can be regarded as the inverse of Minkowski addition, as there is a subtraction operation in arithmetic? The answer is "partly yes"; *Minkowski decomposition* operation $\ominus$ (also called *erosion* in mathematical morphology) is the inverse of Minkowski addition in a restricted sense. Given two sets of points S and B in E^d, Minkowski decomposition $S \ominus B$ is defined as

$$C = S \ominus B = \bigcap_{-b \in \check{B}} S_{-b} \ ,$$

where the set $\check{B} = \{-b \mid b \in B\}$ is called the *symmetrical set* of B with respect to the origin point, and S_{-b} denotes the translate of the set S by a vector $-b$, that is, $S_{-b} = S \oplus \{-b\}$.

We use the word "restricted" because, in general, $(S \ominus B) \oplus B$ is not equal to S, but $(S \ominus B) \oplus B \subseteq S$. The equality holds if and only if B is a summand of S, i.e., $S = A \oplus B$. Moreover, if $S = A_1 \oplus B = A_2 \oplus B = \ldots = A_n \oplus B$, then $(S \ominus B)$ yields the biggest set, say A_m, of all these A_i's; to be more precise, $A_m = A_1 \cup A_2 \ldots \cup A_n$.

However, in the domain of compact convex sets in E^d, if B is a summand of S, then Minkowski decomposition behaves exactly like the inverse of Minkowski addition. In such a situation, if $A_1 \oplus B = A_2 \oplus B$ then $A_1 = A_2$, that is, the other summand is unique, and $S \ominus B$ yields that unique summand. For this reason, we begin our search for negative shapes from the compact convex domain.

3.3 Minkowski Decomposition of Convex Polygons and Emergence of the Notion of Negative Shape

Let A and B be two convex polygons in the plane. Any planar polygon can be represented by a cyclic sequence of its edges $e_1, e_2, \ldots, e_n$, and each edge can be represented by its length and the direction of its outer normal. The Minkowski sum $A \oplus B$ can be easily obtained by carrying out the following algorithm:

1. List the edges of A such that the corresponding outer normals are arranged in some sorted angular order; do the same for the edges of B.
2. Merge the two list of edges by maintaining the sorted angular order and concatenate the edges accordingly. By "concatenation" is meant joining the end-point of one edge to the start-point of the next edge.

It has been shown in [3, 4] that the above procedure can be very conveniently carried out by a method called the *slope diagram method*. The slope diagram of a polygon is a representation of the polygon on a unit circle in the following

way: The outer normal direction at each edge can be represented by the corresponding point on a unit circle, while the outer normal directions at every vertex is represented by the corresponding arc on the unit circle. (By "corresponding point" is meant that point on the unit circle where the outer normal direction is the same as the outer normal direction of the edge.) We may term them *edge point* and *vertex arc* respectively. The length of each edge is associated with its corresponding edge point like a label (Fig. 1a). The slope diagram method of computation of $A \oplus B$ is then nothing but the merging of the slope diagrams of A and B into one, and realising the sum polygon from that merged slope diagram. By "realization" is meant the concatenation of the edges in the sequence in which they appear in the merged slope diagram.

Note that if an edge of A is parallel to an edge of B, that is, if both the edges have the same outer normal direction, their lengths are added automatically by the concatenation process. *Thus Minkowski addition of two convex polygons in E^2 essentially involves sorting and addition of real numbers.* In Fig. 1b we demonstrate this process.

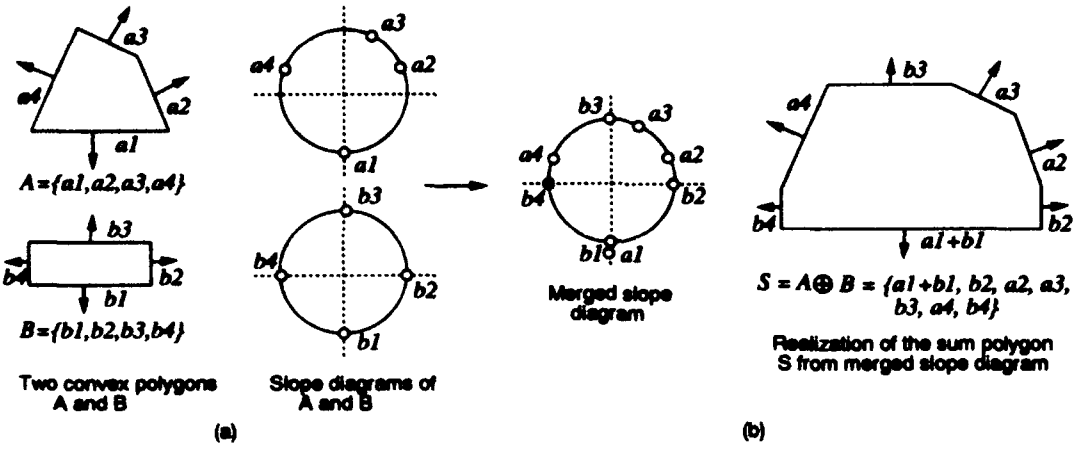


Fig. 1. Minkowski addition of convex polygons by means of slope diagrams.

Consider the task of determining $S \ominus B$, where $S = A \oplus B$ and A, B are two convex polygons. We expect to obtain $S \ominus B = A$. This expected result can be easily obtained if we carry out the following algorithm: (a) merge the slope diagrams of S and B into one, and (b) at the time of realization of the polygon from the merged slope diagram, whenever an edge of B has the same outer normal direction as that of S, "subtract" the length of that edge of B from that of S. Clearly, the subtraction of a directed edge from another is nothing but reversing the direction of the former and then adding/concatenating with the latter. Thus the computation of $S \ominus B$ turns out to be exactly like the computation of $A \oplus B$, except that at the final stage the length of every edge of B has to be subtracted from the corresponding edge of the other operand, instead of being added. For the sake of easy reference in future we shall denote

this computation procedure as P.

As in the integer arithmetic we regard subtraction operation as the addition operation with a negative number, here too we may view the decomposition $S \ominus B$ as the Minkowski addition of S and the additive inverse of B, say B^{-1} (sometimes, to maintain the analogy with negative number, we also write it as $-B$). That means, $S \ominus B = S \oplus B^{-1}$.

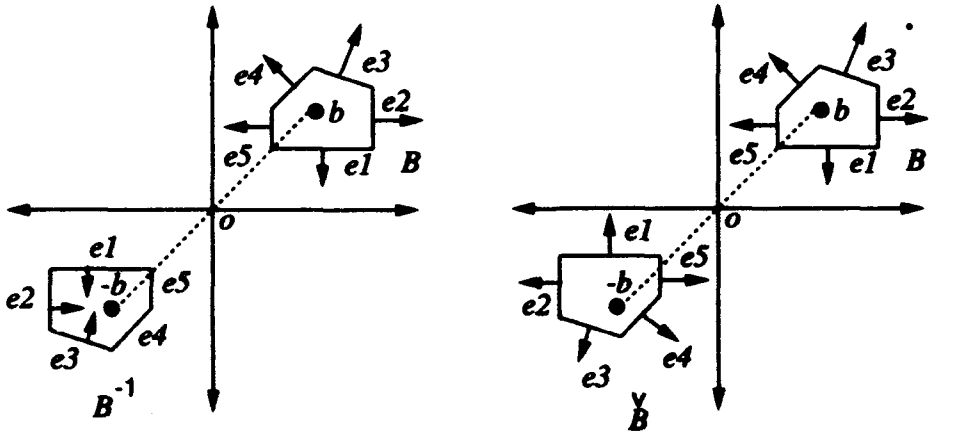
Thus we have arrived at a *procedural definition* of the negative shape B^{-1}: It is such a geometric object that in the case of its Minkowski addition with an ordinary (positive) convex polygon, its edges have to be subtracted instead of being added.

Interestingly, from this definition of negative shape, it is possible to obtain a geometric interpretation too.

3.4 A Geometric Interpretation of Negative Shape B^{-1}

Let S be a singleton point set $\{o\}$. Then $\{o\} \ominus B = \{o\} \oplus B^{-1} = B^{-1}$. Therefore, by following the procedural definition of B^{-1} one finds that the shape of B^{-1} will appear exactly like the shape of $\check{B}$ (the symmetrical set of B). On the other hand, B^{-1} cannot be $\check{B}$, since to be the inverse of B, $B \oplus B^{-1}$ should be equal to $\{o\}$, but $B \oplus \check{B} \neq \{o\}$. The distinction between B^{-1} and $\check{B}$ can be obtained by reversing the direction of the outer normal at each of the faces of $\check{B}$. This is equivalent to reversing the *sense* of the outer normals of B. In Fig. 2 an example of B^{-1} polygon is presented where the corresponding $\check{B}$ is also shown for comparison.

The word "sense" of a normal, which is implicitly present in the name "outer normal", may be elucidated further. Two distinct concepts are associated with



(a) Polygon B and corresponding B^{-1} polygon

(b) Polygon B and corresponding $\check{B}$ polygon

Fig. 2. Geometric representation of negative object B^{-1}; $\check{B}$ is also shown for comparison.

the notion of normal:

- *direction* of the normal which is commonly specified by a unit vector u;
- *sense* of the normal which specifies whether the normal is *diverging outward* from the object or *converging inward*.

Consider an ordinary convex polygonal object. The outer normals at any two adjacent edges appear to diverge outwards from points inside the polygon. That is why the name "outer/outward" normal is used in practice. We may then think of its reverse situation, that is, of an object whose normals from any two adjacent edges converge inwards to points inside the object. Such normals should be called "inner/inward" normals. However, to conform with the present terminology, we shall say that an inward normal is an outer normal but having opposite sense. In other words, if an ordinary outer normal is considered to have the positive sense, then an inward normal has the negative sense.

B^{-1} is an object whose every outer normal has the negative sense. Geometrically it appears like a hole without any positive region surrounding that hole. Of course such an object cannot exist physically, and, for the time being, it can be considered as a purely mathematical object, such as $\sqrt{-1}$.

For convex objects, the differences among the objects B, $\check{B}$, and B^{-1} become clear. The sense of every outer normal of both B and $\check{B}$ is the same, and it is positive, while the directions of the outer normals at the corresponding edges of the two are exactly opposite. Because of the positive sense of the outer normals, we consider both B, and $\check{B}$ as positive objects. On the contrary, the directions of the outer normals at the corresponding edges of B and B^{-1} are exactly the same, but the senses are opposite. Because of the negative sense of the outer normals, B^{-1} is considered as a negative object.

We shall see later (in Sect. 4) that the notion of the sense of the outer normal plays a crucial role in distinguishing nonconvex objects from convex objects.

What will be the slope diagram representation of the B^{-1} polygon? Everything should remain same, except that we have now to distinguish between the positive and the negative sense of a normal. We adopt the convention that if the sense of an outer normal is negative, it will be shown by thick black points or arcs. In contrast to that, an outer normal having positive sense will be shown by thin lines.

3.5 Self-crossing Polygon in the Convex Domain: a Combination of Positive and Negative Shapes

So far we have considered $S \ominus B$ where $S = A \oplus B$. We shall now try to determine $S \ominus B$ where $S \neq A \oplus B$. In this case there is no guarantee that for every edge of B^{-1} there will correspond an edge of S whose length is equal to or greater than that of B^{-1}; for example, a vertex of S, or an edge of S shorter in magnitude, may correspond to an edge of B^{-1}. By "corresponding edges" is meant the edges of B^{-1} and S whose outer normal directions are the same, though they may be of opposite senses. Therefore, by applying the same procedure **P** we shall obtain a *self-crossing polygon*, as shown in Fig. 3.

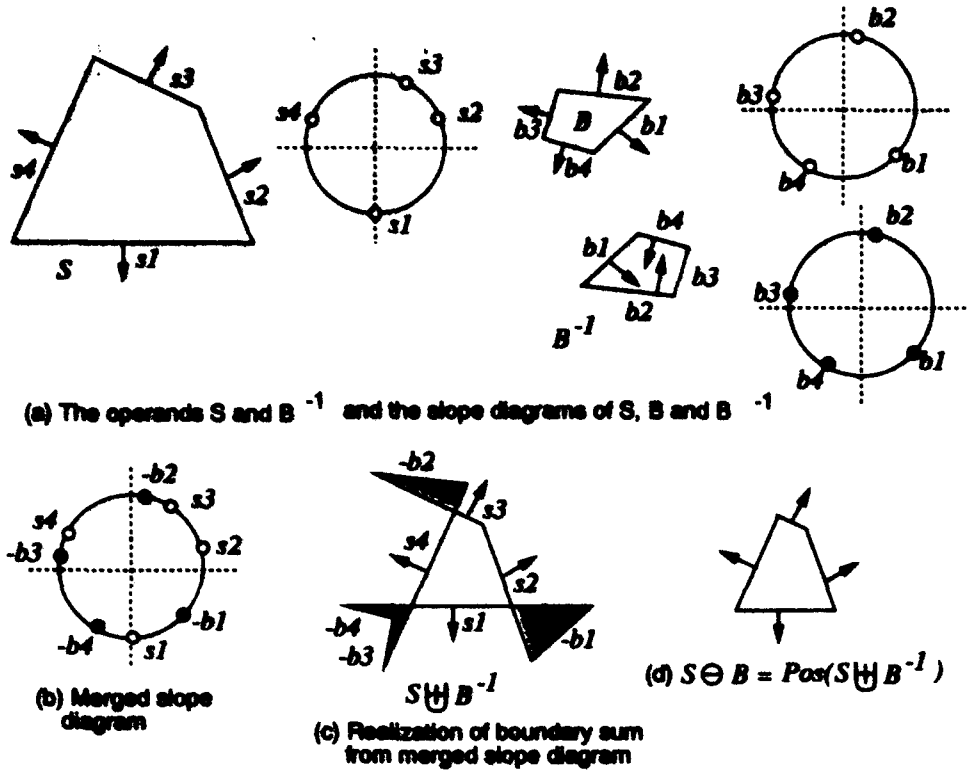


Fig. 3. Determination of $S \ominus B$ when both S and B are convex, but B is not a summand of S.

Note that the resulting self-crossing polygon normally contains a positive and a few isolated negative portions (negative portions are shown shaded in Fig. 3c). An interesting fact is that, *the positive portion (i.e., the physically realizable portion) is equal to* $S \ominus B$. A proof of this result for polygons is given in [2, 5]. (The proof implicitly assumes that $S \ominus B \neq \emptyset$. That means the self-crossing polygon generated by **P** has at least some positive portion. Therefore, in the following discussion it is assumed that S is either bigger or equal to B to ensure such a condition.)

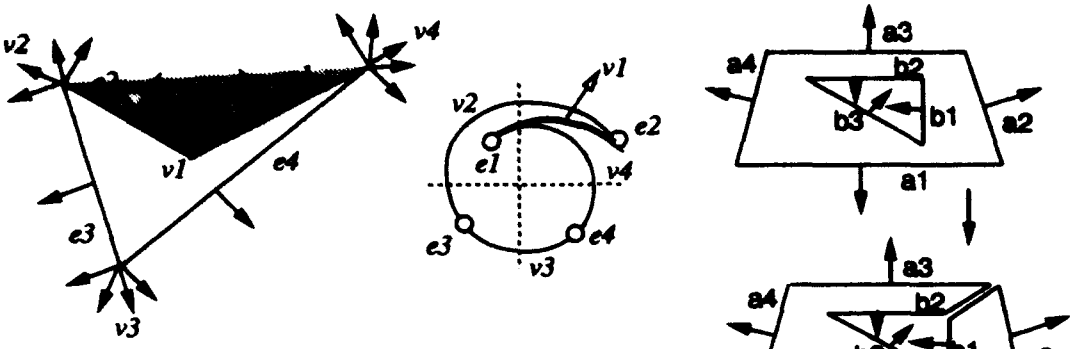
In fact, a slight change in the notational system allows the above result to be expressed in a more compact form. The procedure **P** can be viewed as a binary operation, say the *boundary addition* operation (denoted by "$\uplus$"), and the determination of the positive portion of a generalized object O (by "generalized object" it is meant that O may have both positive and negative portions) as a unary operation, say $Pos(O)$. Therefore, we can write $S \ominus B = Pos(S \uplus B^{-1})$. Following the same notion we can also write, $A \oplus B = Pos(A \uplus B)$, since for convex polygons, $A \uplus B$ does not contain any negative portion, so that $A \uplus B = Pos(A \uplus B)$.

This result quickly explains why $(S \ominus B) \oplus B$ is, in general, not equal to S, but only a subset of S. The reason is, in computing $S \ominus B$, we ignore the negative regions occurred in $S \uplus B^{-1}$. Clearly, the equality is achieved only when

$S \uplus B^{-1} = Pos(S \uplus B^{-1})$. We have observed that such a situation happens in the convex domain when $S = A \oplus B$. On the other hand, it is always true that $(S \uplus B^{-1}) \uplus B = S$.

3.6 Simple Polygon: a Fusion of Positive and Negative Shapes

By a "simple polygon" is meant a simply connected nonconvex polygon. Some of its edges and vertices are nonconvex, while the rest are convex. For example, in Fig. 4a the vertex v_1, and the edges e_1, e_2 of the polygon are nonconvex, while v_2, v_3, v_4, or edges e_3, e_4 are convex. A vertex is called a nonconvex vertex if the internal angle at the vertex is more than 180 degrees; otherwise it is convex. The edges of the polygon that are incident to a nonconvex vertex are called nonconvex edges.



(a) A simple polygon and its slope diagram

(b) Transforming a doubly-connected polygon to a simple polygon

Fig. 4. A simple polygon may be viewed as a fusion of positive and negative shapes.

Note one basic difference between convex and nonconvex portions of a polygon: *the nature of the outer normal at a nonconvex portion is different from that at a convex portion.* As we have already noted, for a positive object, the outer normals at any two adjacent convex portions appear to *diverge* outwards from a point inside the object. In contrast to that, the outer normals at any two adjacent nonconvex portions *converge* to a point outside the object. This difference can be nicely captured by the notion of negative shape and the slope diagram representation.

Consider the complementary region of the nonconvex portion of a simple polygon (part of this complementary region is shown shaded in Fig. 4a). This complementary region appears exactly like a hole or a negative region. The only difference is that, unlike a conventional hole, this hole is not surrounded

by positive regions from all directions. This can be observed more explicitly in Fig. 4b where a doubly-connected polygon with a hole inside is transformed into a simple polygon by cutting an infinitesimally thin slit. For most of the practical purposes including Minkowski operations (but not for topological purposes), these two polygons could be thought of as equivalent. Here notice how a hole turns into a nonconvex portion of a simple polygon. Thus, intuitively we may view a nonconvex object as a fusion of positive and negative shapes.

By means of the slope diagram representation this situation can be depicted very clearly. In the case of convex portions of a polygon, as we have seen, proper topological connections are automatically established if the consecutive edge points and vertex arcs are appropriately marked on the unit circle (Fig. 1a). On the other hand, for the nonconvex portions we have to observe a forward and backward motion along the unit circle in order to maintain the topological connectivity of the edges (Fig. 4a). Besides, the vertex arc corresponding to the nonconvex vertex must be depicted by thick black lines, since the sense of the outer normals is negative there.

To see the consistency of this view of a simple polygon, consider the following fact: by applying the same procedure **P** which is developed for positive and negative convex polygons, we can also determine Minkowski addition and decomposition of simple polygons. In [3, 5] this algorithm is described in detail. Here, for the sake of completeness, the method is briefly indicated. In the slope diagram of a simple polygon, the path along the unit circle corresponding to a nonconvex portion is traversed three times – twice in the positive sense and once in the negative sense (Fig. 4a or Fig. 5b). Therefore, according to the procedure **P**, if there is any edge point of the other summand lying within this portion, it must be considered three times in the appropriate manner. By the "appropriate manner" it is meant that, if the edge point is a positive one then in the negative portion it has to be subtracted, while in the positive portion it has to be added, and so on. In Fig. 5 this method is shown.

3.7 In the Three-dimensional World

Though the discussion has been centered on polygons in E^2, the key notions are completely general and can be easily extended to three- and higher-dimensional spaces. For example, we distinguish between a positive and a negative polygon by the senses of the boundary normals, i.e., whether the normals are outer or inner. The same rule applies to higher-dimensional objects as well. Another key notion is the slope diagram representation. The slope diagram of a polygon in E^2 is represented on a unit circle. Therefore, in the case of a three-dimensional object we have to use a *unit sphere*. In [5] these extensions have been worked out in detail. More interestingly, it was shown there that Minkowski addition (decomposition) of convex polytopes in E^d eventually reduces to Minkowski addition (decomposition) of convex polygons in E^2. That means, it finally reduces to the sorting and addition (subtraction) of real numbers. The implication is that the same procedure **P** can be used to compute Minkowski addition and decomposition of both polygonal and polyhedral objects, and of any higher-dimensional objects as well.

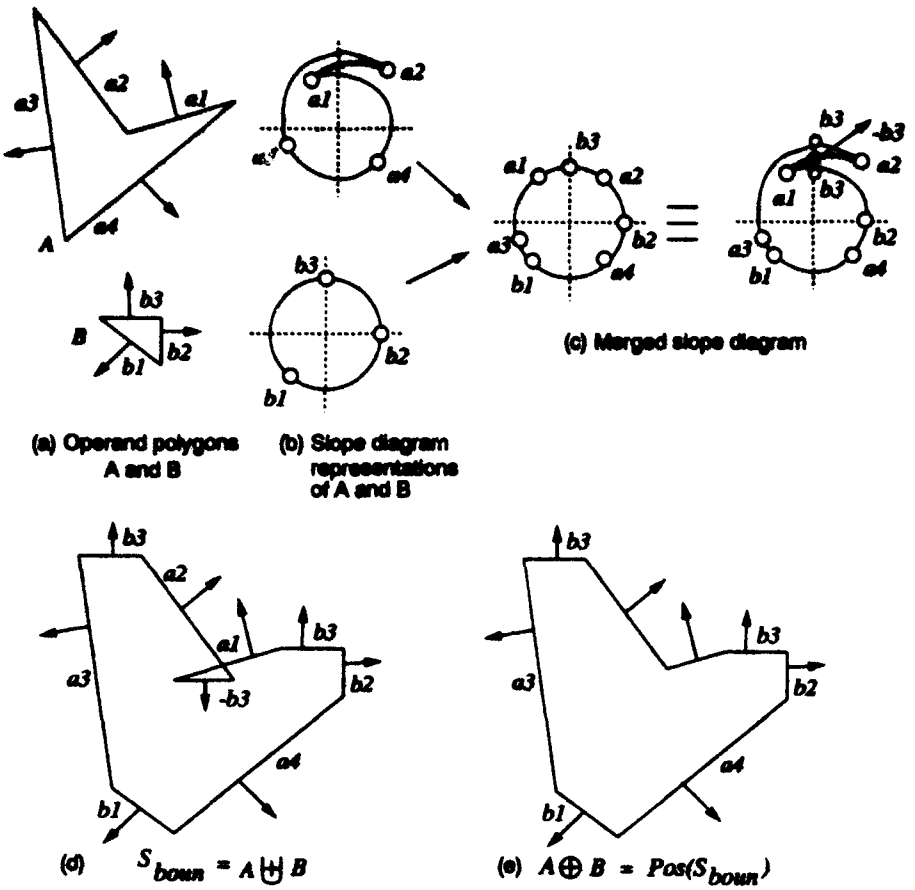


Fig. 5. Minkowski addition with a simple polygon.

4 Some Immediate Advantages of Introducing the Notion of Negative Shape

The development of the concept of negative shape, at this stage, is still in its infancy, and its total significance is yet to be fully comprehended. However, we have already observed some of its immediate advantages, particularly the scope for generalizations/unifications of various geometric concepts. In this section the advantages are briefly summarized.

1. Algebra of shapes. With the introduction of the notion of negative shape, we are now in a position to add and subtract geometric shapes exactly the way we can add and subtract integer numbers. It is now possible, within the *generalized shape domain*, to solve every equation of the type $X \uplus B = A$ as $X = A \uplus B^{-1}$. Such manipulations of shapes are now possible since we have extended the conventional shape domain to include negative shapes. (Only when the question

of "physical realization" of a generalized shape arises, do we exclude its negative portions by means of the unary function $Pos(X)$.)

2. Unification of the Minkowski operations. We have already observed that the unnecessary distinction between Minkowski addition and decomposition disappears within the generalized shape domain. Minkowski addition is essentially the boundary addition $\uplus$ of two positive objects, while Minkowski decomposition is the boundary addition of a positive and a negative object. More precisely,

$$A \oplus B = Pos(A \uplus B) \ ,$$
$$A \ominus B = Pos(A \uplus B^{-1}) \ .$$

3. A generalized concept of shapes and a new categorization of convex and nonconvex objects. The notion of *sense* of a normal apart from its *direction* generalizes the concept of shapes to a great extent. The physically realizable objects in our natural world, as we have seen, can have only positive (outer) boundary normals. But we can conceptualize objects in a mathematical world whose boundary normals may be both positive or negative (inner).

With respect to the sense of a normal, a convex object can be seen as a *pure* shape – either completely positive or completely negative. A nonconvex object, on the other hand, is like a *fusion* of positive and negative objects. The same is true with self-crossing objects.

A question of the following type may arise here: Is it possible to construct a simple (nonconvex) polygon by means of convex polygons – positive and negative? The answer appears to be "yes". In Fig. 6 it is shown how a negative simple polygon may be constructed by the boundary addition of two convex polygons. The example figure also demonstrates that *a simple polygon may also be viewed as a special case of a self-crossing polygon.* (*Warning:* In Fig. 6b we violate one of our assumptions that, in carrying out the operation $A \uplus B^{-1}$, A must be greater than or equal to B. In certain circumstances such violations may give rise to wrong results.)

Further characterization of geometric objects is possible if we take into account not only the senses of the boundary normals, but also their directions. For a convex object, the *sense* of every normal is of the *same type*, and the normals are *sorted* in terms of their *directions* (Fig. 7a). (Since in higher dimensions this sorting order is difficult to visualize, these points will be illustrated by means of two-dimensional examples.) In contrast to that, we may think of a geometric object where the sense of every normal is of the same type (say, positive), but their directions are not sorted. Such an object appears to be a self-crossing object which includes positive portions inside positive portions (Fig. 7b). Alternatively we may also think of an object where the directions are sorted, but the senses of the normals are not of the same type. Such an object, in general, appears to be a self-crossing object, but the positive and negative portions are outside each other (Fig. 7c). As was already noted in Fig. 6, as a special case of such a self-crossing object, we may obtain a simply-connected object.

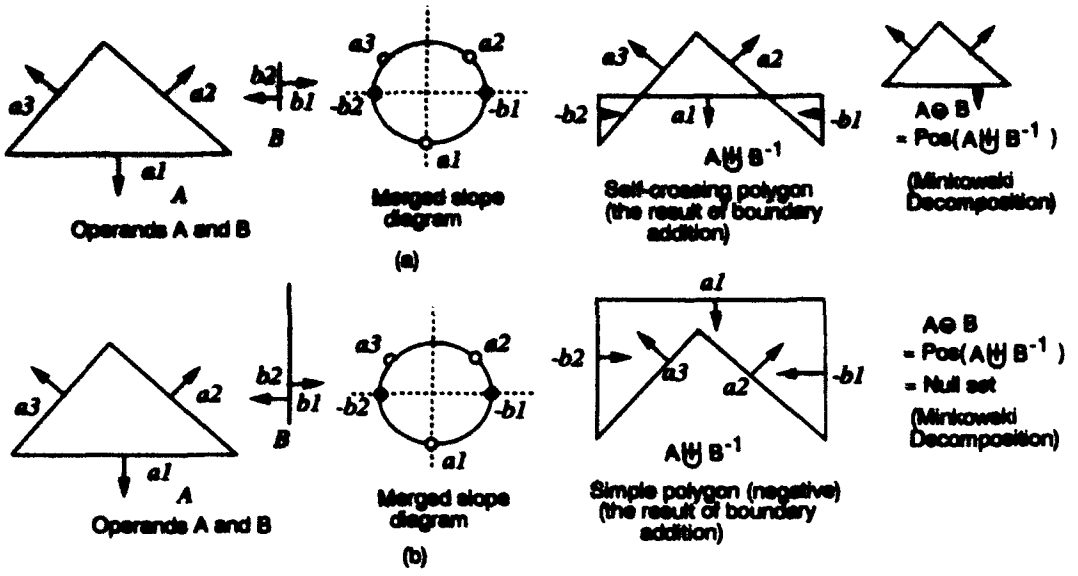


Fig. 6. Construction of a simple polygon from convex ones by means of boundary addition operation. The first operands A in both (a) and (b) are the same, but B in (b) is bigger than that in (a). Note that the merged slope diagrams in (a) and (b) are exactly alike.

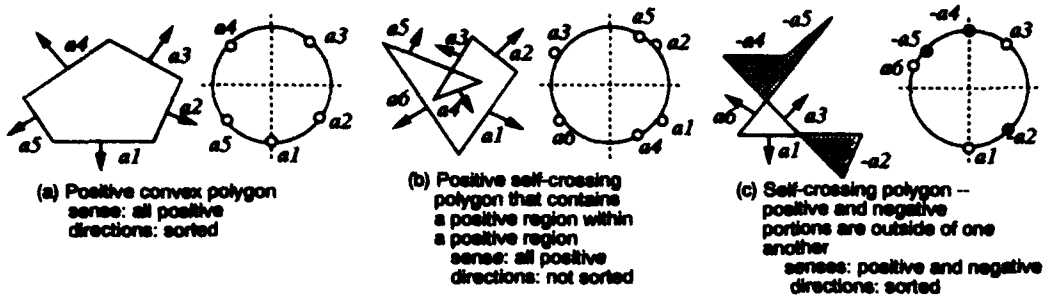


Fig. 7. Constructions of various kinds of polygons by varying senses and directions of boundary normals.

4. A new geometric framework. Using the notions of positive and negative shapes, it is possible to recast many geometric results, particularly those concerning Minkowski operations, in a new and advantageous way. Since it is beyond the scope of this paper to go into those details, we sketch here our intuitive line of approach by means of an example.

Consider the following theorem: If A and B are convex objects, then $A \oplus B$ is also convex. (Here by "convex object" is meant an ordinary (positive) convex object.)

Since A and B are both convex, every boundary normal occurring in $A \uplus B$ has the positive sense. That means $A \uplus B$ is either a positive convex object, or a self-crossing object having positive portions within positive portions (see Fig. 7a and

Fig. 7b). But the latter is not possible, since the directions of the normals in both A and B are sorted and the boundary addition operation merges the faces of A and B by maintaining that sorted order. Therefore, $A \uplus B = Pos(A \uplus B) = A \oplus B$, and $A \oplus B$ is a convex object.

As an exercise the reader may examine the validity of the following theorem taking the above line of approach: If S is a convex object, then $S \ominus B$ is either convex or null.

5 Arriving at the Notion of Negative Shape Through a Different Route

In this section it is shown that, from an analytic formulation of the length-/area/volume of objects, there emerges a notion close to our concept of negative shape.

5.1 Analytic Formulation Introduces a Definitive Sign

We begin by formulating the length $\mathcal{L}$ of a straight line, the area $\mathcal{A}$ of a triangle, and the volume $\mathcal{V}$ of a tetrahedron (Fig. 8):

$$\mathcal{L}_{1,2} = \frac{1}{1} \begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix}, \quad \mathcal{A}_{1,2,3} = \frac{1}{1.2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, \quad \mathcal{V}_{1,2,3,4} = \frac{1}{1.2.3} \begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}.$$

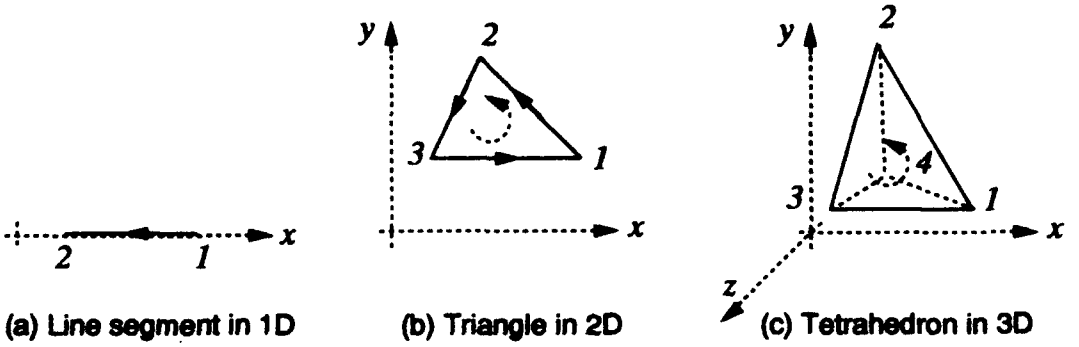


Fig. 8. A line segment, triangle, and tetrahedron in a right-handed coordinate system.

The common practice is to take the *absolute value* of the determinants as the length, area, or volume of the objects. But note that any of these analytic formulae furnishes, in addition to the magnitude of the determinants, a *definitive sign* which is ordinarily ignored.

Let us inquire here as to the geometric significance of this sign. Two points may be immediately noted:

- The sign depends upon the order in which the vertices are taken, that is, $\mathcal{L}_{1,2} = -\mathcal{L}_{2,1}$, $\mathcal{A}_{1,2,3} = -\mathcal{A}_{2,1,3}$, $\mathcal{V}_{1,2,3,4} = -\mathcal{V}_{2,1,3,4}$, etc.
- In the right-handed coordinate system (as in Fig. 8), the formula for length/area/volume has a positive or negative sign according as the order of the vertices turns out to be counterclockwise or the reverse. In the left-handed coordinate system, on the other hand, the sign will be positive when the order of the vertices is clockwise, etc.

The general practice is to represent the positive orientation of the vertices by means of directed edges (Fig. 9). However, one may also use directed boundary normals to achieve the same purpose. The direction of the normal of a directed edge may be assigned as follows: If one moves along the edge facing its indicated direction then, in the right-handed (left-handed) coordinate system, the direction of the right (left) hand will be the direction of the normal (Fig. 9). In summary, *the length/area/volume becomes positive if the order of the vertices are taken in such a way that the normals are directed outwards. Conversely, if the normals are directed inwards, the length/area/volume becomes negative.*

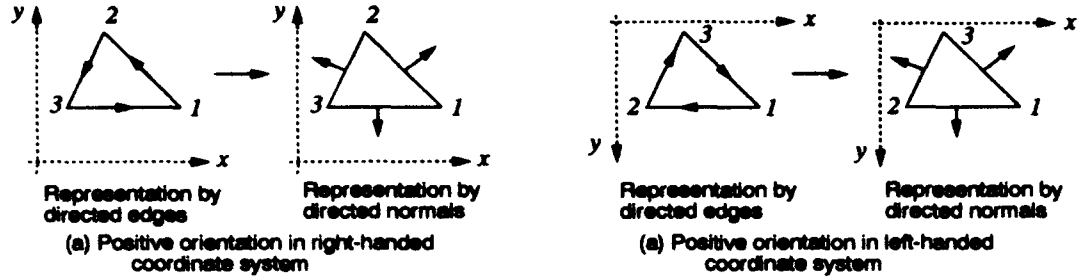


Fig. 9. Representations of the positive orientation of the vertices of a triangle by means of directed edges as well as by directed normals.

If an object, whose length/area/volume turns out to be negative, is viewed as a negative shape, it becomes akin to our concept of negative shape reached earlier through Minkowski operations. (However there is a subtle difference. The length/area/volume of an object A is the same as that of $T(A)$, where $T(A)$ is an isometric (congruent) transformation of A. Therefore we cannot decide, from the sign consideration of the length/area/volume, which of the instances of $T(A)$ would correspond to the shape of A^{-1}. Minkowski operations, on the other hand, indicate that the shape of A^{-1} ought to be that of $\check{A}$.)

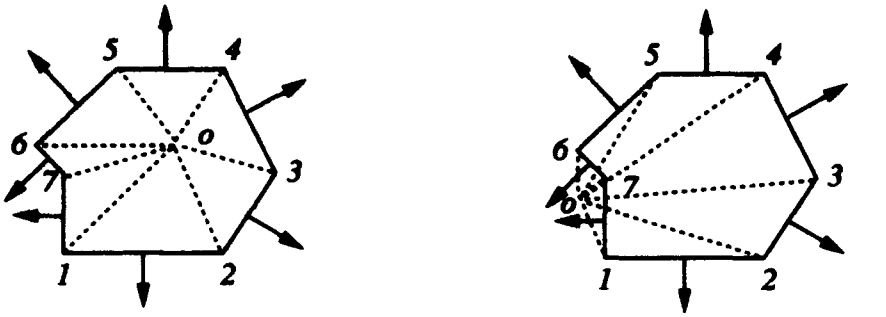
5.2 Signed Length/Area/Volume Provides Greater Generalization

It is well-known that a greater generalization of length/area/volume becomes possible if we accept that the length/area/volume of a negatively oriented object is negative. For example, consider the simple polygon in Fig. 10. If o denotes any

point in the plane (say, the origin), then its area will be given by

$$\mathcal{A}_{1,2,\ldots,n} = \mathcal{A}_{0,1,2} + \mathcal{A}_{0,2,3} + \cdots + \mathcal{A}_{0,n-1,n} + \mathcal{A}_{0,n,1} \ .$$

Now depending on the position of the point o, some of the component triangles may appear to be negative. For example, in Fig. 10b the triangles $\Delta_{0,6,7}$ and $\Delta_{0,7,1}$ are negatively oriented, while every triangle in Fig. 10a is positive. If the areas of those two triangles are considered negative, we shall obtain the same value of $\mathcal{A}_{1,2,\ldots,n}$ in both the cases.



(a) Point o is chosen in such a way that all the component triangles are positive

(b) Point o is such that some of the component triangles are negative

Fig. 10. Determination of the area of a simple polygon as the sum of the signed-areas of its component triangles.

In fact it is possible to achieve further generalization. As we know, it is not possible, in general, to talk about the area of a self-crossing polygon. But now, by means of the signed-area concept, we may assign a value for the area of such a polygon. For example, the area of the polygon in Fig. 11a would be given by, $\mathcal{A}_{1,2,3,4} = \mathcal{A}_{0,2,3} + \mathcal{A}_{0,4,1}$, where we choose the crossing point as the point o. It is clear that the first triangle must have negative area and the second positive area; hence the area of the self-crossing quadrilateral will be equal to the absolute value of the area of the triangle $\Delta_{0,4,1}$ minus that of $\Delta_{0,2,3}$. The areas of the polygons in Fig. 11b or Fig. 11c can be determined in a similar way. For example, in Fig. 11b the inner positive part within the positive part will be counted twice in the area calculation, while the outer positive part will be counted once. For more details refer to [7].

In a similar fashion we can assign the volume of an arbitrary polyhedron by resolving it into component tetrahedra, and adding and subtracting their volumes depending on whether they are positive or negative. (Here may be mentioned the remarkable fact that, in determining the volume of a polyhedron in the above manner, Möbius observed that *there are polyhedra to which one cannot in any way assign a volume*, whereas one can, as we have seen, define area for any plane polygon no matter in how complicated a manner it intersects itself. This observation finally led to the discovery of the *Möbius strip*.)

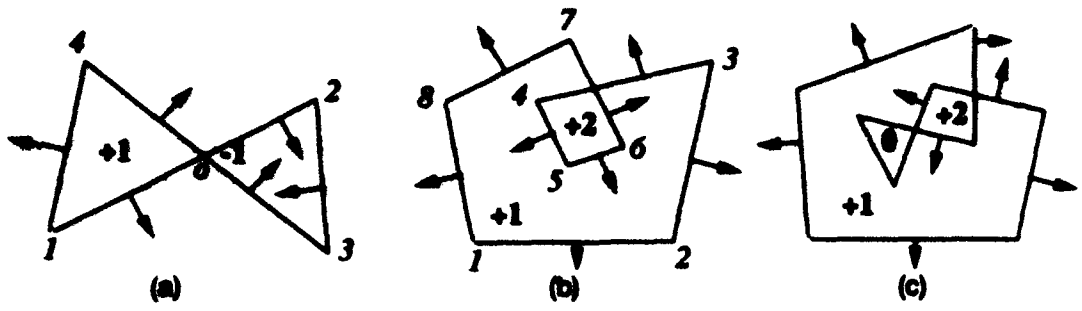


Fig. 11. Determination of the areas of self-crossing polygons.

6 Mixed Area

The primary purpose of this section is to answer the question raised in the introduction: Are the two notions of negative shape – one obtained from the concept of Minkowski operations and the other from the analytic formulation of length/area/volume – identical? In the process should be understood the importance of *mixed area/volume* in defining a more generalised concept of area/volume of objects containing both positive and negative shapes.

6.1 A Generalised Definition of Area/Volume

The concept of mixed area/volume has been brought into the mathematical literature in the following way [1, 9]. If A and B are two convex polygons in the plane, then the area of the polygon $\lambda A \oplus \mu B$ (where λ, μ are two positive numbers and $\lambda A = \{\lambda a \mid a \in A\}$) is given by

$$\mathcal{A}(\lambda A \oplus \mu B) = \lambda^2 \mathcal{A}(A) + 2\lambda\mu(\textit{mixed area of } A \,\&\, B) + \mu^2 \mathcal{A}(B) \ .$$

Similarly, if A and B are two convex polyhedra, then the volume of the polyhedra $\lambda A \oplus \mu B$ is given by

$$\begin{aligned}\mathcal{V}(\lambda A \oplus \mu B) = \lambda^3 \mathcal{V}(A) + 3\lambda^2\mu(\textit{mixed volume of } A, A, B) + \\ 3\lambda\mu^2(\textit{mixed volume of } A, B, B) + \mu^3 \mathcal{V}(B) \ .\end{aligned}$$

Interestingly, it is possible to generalize, from the above two equations, the concept of area/volume of a generalized object. Such a generalization can be achieved by suitably extending the notion of *m-content* to generalized objects. Thus for any object A, let A^m denote this new notion of m-content which is a real number, positive or negative. That means, $|A^1|$ is the 1-content or the length of A, $|A^2|$ is the 2-content or the area of A, $|A^3|$ is the 3-content or the volume of A, etc. For reasons that will be obvious later, we take A^0 to be equal to 1. Similarly for the extended notion of *mixed* $(m+n)$*-content* of A^m and B^n, let us adopt the notation $A^m \circ B^n$.

We further state:

1. $A^m \circ B^n = B^n \circ A^m$.
2. $A^k \circ A^{m-k} = A^m$.
3. $(\lambda A)^m \circ (\mu B)^n = \lambda^m \mu^n (A^m \circ B^n)$. [To accommodate negative shape, we allow λ, μ to be any real numbers — positive or negative. One may treat a negative shape $B^{-1} = -B = (-1)B$.]
4.

$$(\lambda A \uplus \mu B)^m = \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} (\lambda A)^k \circ (\mu B)^{m-k}$$
$$= \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} \lambda^k \mu^{m-k} (A^k \circ B^{m-k}) \ .$$

(Notice its resemblance with the binomial expansion.)
More generally, if $A_1, A_2, \ldots, A_l$ are l number of generalized objects, then

$$(\lambda_1 A_1 \uplus \lambda_2 A_2 \ldots \uplus \lambda_l A_l)^m =$$
$$\sum_{\substack{i_1+\cdots+i_l=m \\ 0 \le i_j \le m}} \frac{m!}{i_1! i_2! \ldots i_l!} \lambda_1{}^{i_1} \lambda_2{}^{i_2} \ldots \lambda_l{}^{i_l} (A_1{}^{i_1} \circ A_2{}^{i_2} \circ \ldots \circ A_l{}^{i_l}) \ . \quad (1)$$

One crucial step in this generalization is to use the boundary addition operation $\uplus$ instead of Minkowaki addition operation $\oplus$. The reader must have noticed that Eqn.(1) agrees with the equations of area and volume as given for convex polygons and polyhedra, since for convex objects, $\lambda A \oplus \mu B = Pos(\lambda A \uplus \mu B) = \lambda A \uplus \mu B$.

6.2 An Answer to the Question

We are now in a position to show, though in an informal way, that the two notions of negative shape are indeed identical. For the sake of simplicity discussion will be restricted to the two-dimensional domain, and reasoned in the following manner. First, we determine the area of the polygon $\lambda A \uplus \mu B$ using Eqn.(1). Second, from the shape of $\lambda A \uplus \mu B$ we identify its positive and negative portions, and determine the area of $\lambda A \uplus \mu B$ by appropriately adding and subtracting the signed areas of those portions (as explained in Sect. 5). We can show that the values of the area calculated in these two different ways will be the same. We shall take up two simple examples to demonstrate this fact.

In the discussion that follows we make use of the following fact: It is always possible to suppose, by means of introducing edges of zero length, that any two convex polygons have pairwise parallel and similarly directed edges. In other words, according to our slope diagram terminology, we can always suppose that any two convex polygons have exactly the same slope diagrams, though some of the edge points in the diagrams may have zero length.

Let o be an interior point of a convex polygon A. Let $h_i{}^A$ denote the perpendicular distance from o to the ith edge of A, and a_i denote the length of

that ith edge. Then the area of A is given by, $\mathcal{A}(A) = \frac{1}{2}\sum_{i=1}^{n} a_i h_i{}^A$, where n is the number of edges in A. Similarly, the area of another convex polygon B is $\mathcal{A}(B) = \frac{1}{2}\sum_{i=1}^{n} b_i h_i{}^B$. The mixed area of A and B, denoted by $\mathcal{M}(A, B)$, is then given by [1, 9]

$$\mathcal{M}(A, B) = \frac{1}{2}\sum_{i=1}^{n} a_i h_i{}^B = \frac{1}{2}\sum_{i=1}^{n} b_i h_i{}^A .$$

For $A = B$, the mixed area $\mathcal{M}(A, B)$ coincides with the usual area of A. (For this reason, sometimes the usual area of A is denoted by $\mathcal{A}(A, A)$, instead of $\mathcal{A}(A)$, and the mixed area by $\mathcal{A}(A, B)$, instead of $\mathcal{M}(A, B)$. Notice the resemblance of this notation to Eqn.(1), that is, $|\mathcal{A}(A, A)| = |A^2|$, $|\mathcal{A}(A, B)| = |A^1 \circ B^1|$, etc.).

Example 1. Summand polygons are all positive
Let the summands A, B, C be three positive line segments having lengths a, b, c respectively, and making angles 0, θ_1, θ_2 with some arbitrarily chosen axis (Fig. 12a). We shall determine the area of their boundary sum $S = (A \uplus B \uplus C)$ in two different ways. Clearly, S will be a centre-symmetric positive hexagon as shown in Fig. 12b. The area of every line segment is zero, that is, $A^2 = B^2 = C^2 = 0$. The mixed area of A and B , that is, $A^1 \circ B^1 = \frac{1}{2}bh' = \frac{1}{2}ab\sin\theta_1$ (see Fig. 12c). Similarly, $C^1 \circ A^1 = \frac{1}{2}ca\sin\theta_2$, and $B^1 \circ C^1 = \frac{1}{2}bc\sin(\theta_2 - \theta_1)$. Therefore, according to Eqn.(1),

$$\begin{aligned} S^2 = (A \uplus B \uplus C)^2 &= A^2 + B^2 + C^2 + 2(A^1 \circ B^1) + 2(B^1 \circ C^1) + 2(C^1 \circ A^1) \\ &= ab\sin\theta_1 + bc\sin(\theta_2 - \theta_1) + ca\sin\theta_2 . \end{aligned}$$

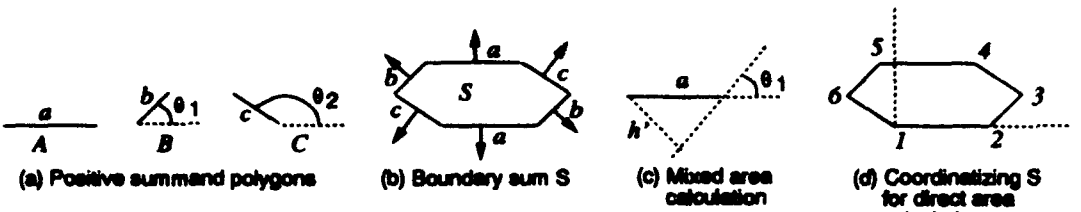


Fig. 12. Demonstration that the area of a sum polygon determined by two different methods yields the same result: all summands are positive.

To calculate the area of S by the method described in Sect. 5, we place S within a conveniently chosen coordinate system as shown in Fig. 12d. It is easy to see that the coordinates of the vertices 1, 2, 3, 4, 5, 6 will be $(0, 0)$, $(a, 0)$, $(a + b\cos\theta_1, b\sin\theta_1)$, $(a + b\cos\theta_1 + c\cos\theta_2, b\sin\theta_1 + c\sin\theta_2)$, $(b\cos\theta_1 + c\cos\theta_2, b\sin\theta_1 + c\sin\theta_2)$, $(c\cos\theta_2, c\sin\theta_2)$ respectively. Therefore, the area of S will be given by

$$\mathcal{A}(S) = \mathcal{A}_{1,2,3} + \cdots + \mathcal{A}_{1,5,6} = ab\sin\theta_1 + bc\sin(\theta_2 - \theta_1) + ca\sin\theta_2 .$$

Thus we find, $|S^2| = |\mathcal{A}(S)|$.

Example 2. One positive and one negative summand
Let the summands A, B be respectively a triangle and a line segment (Fig. 13a). We shall determine the area of the boundary sum $S = A \uplus B^{-1}$ (Fig. 13b). The mixed area $A^1 \circ B^1 = \frac{1}{2}bh' = \frac{1}{2}a_1 b \sin\theta$ (refer to Fig. 13c). Then, according to Eqn.(1), the area of S will be given by

$$S^2 = (A \uplus B^{-1})^2 = A^2 + B^2 - 2(A^1 \circ B^1) = \mathcal{A}(A) - a_1 b \sin\theta \ ,$$

where $\mathcal{A}(A)$ denotes the area of the triangle A.

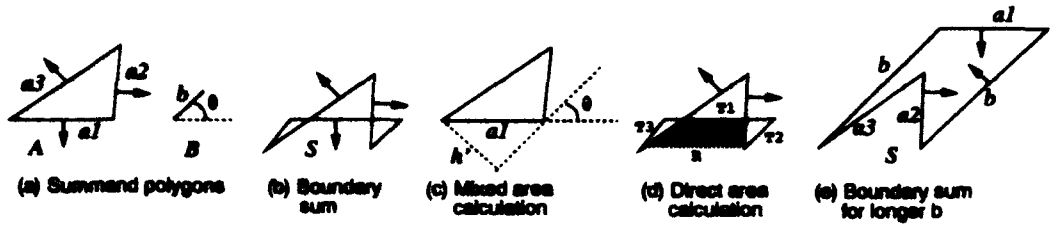


Fig. 13. Determinations of the area of a sum polygon when one summand is positive and one is negative.

To determine the area of S in the other method, we find that S has one positive portion T_1 and two negative portions T_2 and T_3 (Fig. 13d). That means

$$\mathcal{A}(S) = |\mathcal{A}(T_1)| - |\mathcal{A}(T_2)| - |\mathcal{A}(T_3)| \ .$$

If the region shown shaded in S is denoted by R, then we notice that

$$|\mathcal{A}(T_1)| + |\mathcal{A}(R)| = |\mathcal{A}(A)|; \ |\mathcal{A}(T_2)| + |\mathcal{A}(t_3)| + |\mathcal{A}(R)| = a_1 b \sin\theta \ .$$

Subtracting the second equation from the first we get, $\mathcal{A}(S) = \mathcal{A}(A) - a_1 b \sin\theta$. Thus we again find that, $|S^2| = |\mathcal{A}(S)|$.

(Interestingly, if the length b of the summand B is made longer, the sum S will become a negative simple polygon (see Fig. 13e and refer to the discussion in Sect. 4). It is easy to see that the area of this simple polygon is also $\mathcal{A}(A) - a_1 b \sin\theta$.)

7 Concluding Remarks

In this paper the concept of negative shape has been introduced in an intuitive way. But in order to make negative shape a "true" extension of the concept of geometric shape, a more formal approach is needed. Further, its introduction leads to many fundamental questions that remain unanswered. For example:

- The concept of "outer/positive" normal has been rigorously introduced into geometry by means of the *supporting function* $H(A, u)$ of a set $A \subset E^d$ [6]. It is defined, for all $u \in E^d$, as

$$H(A, u) = \sup\{< a, u > \mid a \in A\} .$$

  What will happen if we use "inf", instead of "sup", in the definition? Is it possible, in that way, to define the concept of "inner/negative" normal more formally?
- Some of the most common operations on the conventional geometric shapes are the set operations (that is, union, intersection, difference, complement, etc.), or, various geometric transformations. How do we redefine those operations so that they could be extended, without contradicting any of our existing notions, to generalized shapes?
- Can the shape domain be extended even further, just as the integer number domain was finally extended to the complex number domain? Will that serve any useful purpose? For example, a real line $a_1x_1 + a_2x_2 + a_3 = 0$ always contains real points, such as the point $(-a_3/a_1, 0)$, if $a_1 \neq 0$, but it also contains the "imaginary" point $(-(ia_2 + a_3)/a_1, i)$. Or, a real circle, ${x_1}^2 + {x_2}^2 + 1 = 0$ has a real centre at $(0, 0)$, but contains no real points. What are the geometric interpretations of such phenomena?

Clearly, at present we are leaving our exploration in a state of incompleteness.

References

1. Benson, R.V. (1966). Euclidean Geometry and Convexity, McGraw-Hill, New York.
2. Ghosh, P.K. (1990). A solution of polygon containment, spatial planning, and other related problems using Minkowski operations, Comput. Vision Graphics Image Process. 49, pp. 1–35.
3. Ghosh, P.K. (1991). An algebra of polygons through the notion of negative shapes, CVGIP: Image Understanding 54, No.1, pp. 119–144.
4. Ghosh, P.K. (1991). Vision, geometry, and Minkowski operators, Contemporary Mathematics: Vision Geometry, Vol. 119, Am. Math. Soc., pp. 63–83.
5. Ghosh, P.K. (1993). A unified computational framework for Minkowski operations, Computers and Graphics, Vol. 17 (4).
6. Grünbaum, B. (1967). Convex Polytopes, Interscience, London.
7. Klein, F. (1939). Elementary Mathematics from an advanced standpoint: Geometry, Macmillan, New York.
8. Kline, M. (1972). Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York.
9. Lyusternik, L.A. (1963). Convex Figures and Polyhedra, Dover Publications, New York.

An Overview of the Theory and Applications of Wavelets *

Björn Jawerth[1] *and Wim Sweldens*[2]

[1] University of South Carolina, Department of Mathematics, Columbia, SC 29208, USA

[2] Katholieke Universiteit Leuven, Department of Computer Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium, and University of South Carolina, Department of Mathematics

Abstract. In this paper an attempt is made to give an overview of some existing wavelet techniques. The continuous wavelet transform and several wavelet-based multiresolution techniques leading to the fast wavelet transform algorithm are briefly discussed. Different families of wavelets and their construction are discussed and compared. The essentials of two major applications are outlined: data compression and compression of linear operators.

Keywords: wavelets, Littlewood-Paley techniques, Calderón-Zygmund theory, continuous wavelet transform, multiresolution analysis, discrete wavelet transform, splines, orthogonal wavelets, biorthogonal wavelets, fast wavelet transform, multidimensional wavelets, data compression, Burger's equation.

1 Introduction

Wavelets and wavelet techniques have recently generated much interest, both in applied areas as well as in more theoretical ones. The class of wavelet techniques is not really precisely defined, and what is placed in this class keeps changing. Here an overview of some of the important wavelet techniques and existing wavelet functions is presented. We shall also briefly discuss, whenever appropriate, their advantages and disadvantages. This short overview is unavoidably incomplete and does not cover many important and interesting developments. Many of the results we do not mention are more significant than the ones we include, and we apologize to the people whose work is not discussed. For example, we hardly mention the significant volume of work done more in the direction of approximation theory, and the efforts in the field of fractal functions and the more applied areas are left out almost entirely.

Although wavelets are a relatively recent phenomenon, there are already several books on the subject, for example [10, 11, 20, 24, 32, 47, 56, 63].

* The first author is partially supported by AFOSR Grant 89-0455, DARPA Grant AFOSR 89-0455 and ONR Grant N00014-90-J-1343. The second author is Research Assistant of the National Fund of Scientific Research Belgium, and is partially supported by ONR Grant N00014-90-J-1343.

2 Notation and Definitions

Much of the notation will be presented as we go along. Here we just note that the inner product of two functions $f, g \in L^2(\mathbb{R})$ is defined as

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(x)\,\overline{g(x)}\,dx \ .$$

The Fourier transform of a function $f \in L^2(\mathbb{R})$ is defined as

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\,e^{-i\omega x}\,dx \ .$$

We shall also use the following formula

$$\sum_l \langle f(x), g(x-l) \rangle\, e^{-i\omega l} = \sum_k \hat{f}(\omega + k2\pi)\,\overline{\hat{g}(\omega + k2\pi)} \ ,$$

which in the case of $f = g$ becomes the Poisson summation formula. If no bounds are indicated under a summation sign, $\in \mathbb{Z}$ is understood.

A countable system $\{f_n\}$ of a Hilbert space is a *Riesz basis* if every element f of the space can be written uniquely as $f = \sum_n c_n f_n$, and positive constants A and B exist such that

$$A \|f\|^2 \leq \sum_n |c_n|^2 \leq B \|f\|^2 \ .$$

3 A Short History of Wavelets

Wavelet theory involves representing general functions in terms of simpler, fixed building blocks at different scales and positions. This has been found to be a useful approach in several different areas. For example, we have subband coding techniques, quadrature mirror filters, pyramid schemes, etc., in signal and image processing, while in mathematical physics similar ideas are studied as part of the Theory of Coherent States. In abstract mathematics it has been known for quite some time that techniques based on Fourier series and Fourier transforms are not quite adequate for many problems and so-called Littlewood-Paley techniques are often effective substitutes. These techniques were initially developed in the Thirties to understand, for example, summability properties of Fourier series and boundary behaviour of analytic functions. However, in the Fifties and Sixties they developed into powerful tools for understanding other things such as solutions of partial differential equations and integral equations. It was realized that they fit into Calderón–Zygmund theory, an area of harmonic analysis which is still very heavily researched. One of the standard approaches, not only in Calderón-Zygmund theory but in analysis in general, is to break up a complicated phenomenon into many simple pieces and study each of the pieces separately. In the Seventies, sums of simple functions, called atomic decompositions [19], were widely used, especially in Hardy space theory. One method used

to establish that a general function f has such a decomposition is to start with the "Calderón formula": for a general function f,

$$f(x) = \int_0^{+\infty} \int_{-\infty}^{+\infty} (\psi_t * f)(y)\, \tilde{\psi}_t(x-y)\, dy\, \frac{dt}{t} .$$

The $*$ denotes convolution. Here $\psi_t(x) = t^{-1}\psi(x/t)$, and similarly for $\tilde{\psi}_t(x)$, for appropriate fixed functions ψ and $\tilde{\psi}$. In fact, as we shall see below, this representation is an example of a continuous wavelet decomposition. In the context of trying to further understand Hardy spaces, as well as other spaces used to measure the size and smoothness of functions, and showing very deep, but also very abstract, functional analytic properties, the first orthogonal wavelets were constructed by Strömberg [67]. A discrete version of the Calderón formula had also been used for similar purposes in [41] and long before this there were results by Haar [37], Franklin [28], Ciesielski [13], Peetre [61], and others.

Independently from these developments in harmonic analysis, Grossmann, Morlet et al. studied the wavelet transform in its continuous form [34, 35, 36]. The theory of "frames" [25] provided a suitable general framework for these investigations.

In the early to mid Eighties there were several groups, perhaps most notably the one associated with Meyer and his collaborators, that independently realized, with some excitement, that some of the tools that had been so effective in Calderón-Zygmund theory, in particular the Littlewood-Paley representations, had discrete analogues and could be used both to give a unified view of many of the results in harmonic analysis and also, at least potentially, could be effective substitutes for Fourier series in numerical applications. (The first named author of this paper came to this understanding through joint work with Frazier [29, 30, 31].) As the emphasis shifted more towards the representations themselves, and the building blocks involved, the name also shifted: Meyer and Morlet suggested the word "wavelet" for the building blocks, and what earlier had been referred to as Littlewood-Paley theory now started to be called wavelet theory.

Lemarié and Meyer [48], independently of Strömberg, constructed new orthogonal wavelet expansions. With the notion of multiresolution analysis, introduced by Mallat and Meyer and which we shall discuss below, a systematic framework for understanding these orthogonal expansions was provided [53, [illegible], 55]. Soon Daubechies [23] gave a construction of wavelets, non-zero only on a finite set and with arbitrarily high, but fixed, regularity. This takes us up to a fairly recent time in the history of wavelet theory. Several people have made substantial contributions to the field over the past few years. Some of their work and the appropriate references will be discussed in the body of the paper.

4 The Continuous Wavelet Transform

As this overview is brief, more detailed treatments of the continuous wavelet transform can be found in [10, 34, 38, 43]. As mentioned above, a wavelet expansion consists of translations and dilations of one fixed function, the wavelet

$\psi \in L^2(\mathbb{R})$. In the continuous wavelet transform the translation and dilation parameter vary continuously. This means that we use the functions

$$\psi_{b,a}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) \quad \text{with} \quad a, b \in \mathbb{R}, a \neq 0 \ .$$

These functions are scaled so that their $L^2(\mathbb{R})$ norms are independent of a. The *continuous wavelet transform* of a function $f \in L^2(\mathbb{R})$ is now defined as

$$\mathcal{W}(a,b) = \langle f, \psi_{b,a} \rangle \ . \tag{1}$$

Using the Parseval identity we can also write this as

$$2\pi \mathcal{W}(a,b) = \langle \hat{f}, \hat{\psi}_{b,a} \rangle \ . \tag{2}$$

and

$$\hat{\psi}_{b,a}(\omega) = \frac{a}{\sqrt{|a|}} e^{-ib\omega} \hat{\psi}(a\omega) \ . \tag{3}$$

Note that the continuous wavelet transform takes a one-dimensional function into a two-dimensional one. The representation of a function by its continuous wavelet transform is redundant and the inverse transform is possibly not unique. Furthermore, not every function $\mathcal{W}(a,b)$ is the continuous wavelet transformation of a function f.

We assume that the wavelet ψ and its Fourier transform $\hat{\psi}$ are window functions with centres $\bar{x}$ and $\bar{\omega}$ and radii Δ_x and Δ_ω. The latter quantities are defined as

$$\bar{x} = \frac{1}{\|\psi\|_{L_2}^2} \int_{-\infty}^{+\infty} x \, |\psi(x)|^2 \, dx \ ,$$

and

$$\Delta_x = \frac{1}{\|\psi\|_{L_2}} \sqrt{\int_{-\infty}^{+\infty} (x-\bar{x})^2 \, |\psi(x)|^2 \, dx} \ ,$$

and similarly for $\bar{\omega}$ and Δ_ω. Although the variable x typically represents either time or space, we shall refer to it as time. From (1) and (3) we see that the continuous wavelet transform at (a,b) essentially contains information from the time interval $[b + a\bar{x} - a\Delta_x, b + a\bar{x} + a\Delta_x]$ and the frequency interval $[(\bar{\omega} - \Delta_\omega)/a, (\bar{\omega} + \Delta_\omega)/a]$. These two intervals determine a *time-frequency window*. Its width, height and position are governed by a and b. Its area is constant and given by $4\Delta_x\Delta_\omega$. Due to the *Heisenberg uncertainty* the area has to be greater then 2. The time-frequency windows are therefore also called *Heisenberg boxes*.

Suppose that the wavelet ψ satisfies the *admissibility condition*

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} \, d\omega < \infty \ ,$$

then the continuous wavelet transform $\mathcal{W}(a,b)$ has an inverse given by the relation

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}(a,b) \, \psi_{b,a}(x) \, \frac{da \, db}{a^2} \ . \tag{4}$$

From the admissibility condition we see that $\hat{\psi}(0)$ has to be 0, and, hence, ψ is oscillatory. This together with the decay property gave ψ the name *wavelet* or "small wave" (French: *ondelette*). Other, more efficient inverse transforms exist that only use $\mathcal{W}(a,b)$ for positive values of a in the reconstruction, or even only use $\mathcal{W}(a,b)$ at discrete values of a.

This transform is used to analyze signals, e.g. in geophysics. The transform is often graphically represented as two two-dimensional images with colour or grey value corresponding to the modulus and phase of $\mathcal{W}(a,b)$. The continuous wavelet transformation is also used in *singularity detection* and characterization [29, 50]. A typical result in this direction is that if a function f is Lipschitz continuous of order $0 < \alpha < 1$, so that $|f(x+h) - f(x)| = \mathcal{O}(h^\alpha)$, then the continuous wavelet transformation has an asymptotic behaviour like

$$\mathcal{W}(a,b) \;=\; \mathcal{O}(a^{\alpha+1/2}) \quad \text{for} \quad a \to 0 \;.$$

In fact, the converse is true as well. The advantage of this characterization with respect to the Fourier transform is that it does not only provide information on the kind of singularity, but also on its location in time. There is also a corresponding characterization of Lipschitz continuous functions of higher order $\alpha \geq 1$; the wavelet must then have a number of vanishing moments greater than α, that is

$$\int_{-\infty}^{+\infty} \psi(x)\, x^p\, dx \;=\; 0 \quad \text{for} \quad 0 \leq p \leq \alpha \;.$$

Example: A classical example of a wavelet is the *Mexican hat*

$$\psi(x) \;=\; (1 - 2x^2)e^{-x^2} \;.$$

This is the second derivative of a Gaussian and it has thus two vanishing moments.

5 Multiresolution Analysis

5.1 The Scaling Function

There are at least two ways to introduce wavelets: one is through the continuous wavelet transform as in the previous section; another is through multiresolution analysis. Here we shall start by introducing the concept of multiresolution analysis and then point out the similarities with the continuous wavelet transform.

A *multiresolution analysis* of $L^2(\mathbb{R})$ is defined by means of a sequence of closed subspaces V_j, with $j \in \mathbb{Z}$, that has the following properties [23, 53]:

1. $V_j \subset V_{j+1}$,
2. $v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$,
3. $v(x) \in V_0 \Leftrightarrow v(x+1) \in V_0$,
4. $\bigcup_{j=-\infty}^{+\infty} V_j$ is dense in $L^2(\mathbb{R})$ and $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$,

5. A *scaling function* $\phi \in V_0$ with a non vanishing integral exists such that the collection $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is a Riesz basis of V_0.

Let us make a couple of simple observations related to this definition. Since $\phi \in V_0 \subset V_1$, a sequence $(h_k) \in l^2(\mathbb{Z})$, referred to as the *scaling parameters*, exists such that the scaling function satisfies a *refinement equation*

$$\phi(x) = 2 \sum_k h_k \, \phi(2x-k) \ . \tag{5}$$

The collection of functions $\{\phi_{j,l} \mid l \in \mathbb{Z}\}$ with $\phi_{j,l}(x) = \sqrt{2^j}\, \phi(2^j x - l)$, is clearly a Riesz basis of V_j.

We also note that a multiresolution analysis allows us to approximate a given function f and obtain an approximation f_j in each of the spaces V_j. Since the union $\cup_{j=-\infty}^{+\infty} V_j$ is dense in $L^2(\mathbb{R})$, we are guaranteed that there are such approximations converging to the original function, $f = \lim_{j \to +\infty} f_j$.

By integrating both sides of (5) and using the fact that the integral of ϕ does not vanish, we see that

$$\sum_k h_k \ = \ 1 \ . \tag{6}$$

The properties of a scaling function are closely related to its scaling parameters. In fact, the scaling function is, under very general conditions, uniquely defined by its refinement equation and the normalization

$$\int_{-\infty}^{+\infty} \phi(x)\, dx \ = \ 1 \ .$$

In many cases, no explicit expression for ϕ is available. However, there are quick algorithms that use the refinement equation to evaluate the scaling function ϕ at dyadic points ($x = 2^{-j}k$, $j, k \in \mathbb{Z}$) (see, for example, [65]). In most applications, we never need the scaling function itself; instead we may often work directly with the scaling parameters h_k.

To be able to use the collection $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ to approximate even the simplest functions (such as constants), it is natural to assume that the scaling function and its integer translates form a *partition of the unity*, or, in other words,

$$\forall x \in \mathbb{R} \ : \ \sum_k \phi(x-k) \ = \ 1 \ .$$

This is also used to prove that a certain ϕ generates a multiresolution analysis. By Poisson's summation formula, the partition of unity relation is a consequence of

$$\hat{\phi}(2\pi k) = \delta_k \quad \text{for} \quad k \in \mathbb{Z} \ . \tag{7}$$

By (5), the Fourier transform of the scaling function must satisfy

$$\hat{\phi}(\omega) \ = \ H(w/2)\, \hat{\phi}(\omega/2) \ , \tag{8}$$

where H is a 2π-periodic function defined as

$$H(\omega) = \sum_k h_k e^{-ik\omega} .$$

Using (7) and (8), we see that we obtain a partition of the unity if

$$H(\pi) = 0 \quad \text{or} \quad \sum_k (-1)^k h_k = 0 .$$

We also see that (6) can be written as

$$H(0) = 1 .$$

Since $\hat{\phi}(0) = 1$, we can apply (8) recursively. This yields, at least formally,

$$\hat{\phi}(\omega) \sim \prod_{j=1}^{\infty} H(2^{-j}\omega) .$$

The convergence of this product is examined in [14, 23]. The product formula for $\hat{\phi}$ is nice to have in many situations. For example, it can be used to construct $\phi(x)$ from its scaling parameters.

Examples of scaling functions:

- A well known family of scaling functions are the cardinal B–splines. The cardinal B–spline of order 1 is the box function $N_1(x) = \chi_{[0,1]}(x)$. For $m > 1$ the cardinal B–spline N_m is defined as a convolution,

  $$N_m = N_{m-1} * N_1 .$$

  These splines satisfy

  $$N_m(x) = 2^{m-1} \sum_k \binom{m}{k} N_m(2x - k) ,$$

  and

  $$\hat{N}_m(\omega) = \left(\frac{1 - e^{-i\omega}}{i\omega} \right)^m .$$

- Another classical example is the Shannon sampling function,

  $$\phi(x) = \frac{\sin(\pi x)}{\pi x} \quad \text{with} \quad \hat{\phi}(\omega) = \chi_{[-\pi,\pi]}(\omega) .$$

  We may take

  $$H(\omega) = \chi_{[-\pi/2,\pi/2]}(\omega) \quad \text{for} \quad \omega \in [-\pi, \pi] ,$$

  and, consequently,

  $$h_{2k} = 1/2\,\delta_k \quad \text{and} \quad h_{2k+1} = \frac{(-1)^k}{(2k+1)\pi} \quad \text{for} \quad k \in \mathbb{Z} .$$

5.2 The Wavelet Function

We will use W_j to denote a space complementing V_j in V_{j+1}, i.e. a space that satisfies

$$V_{j+1} = V_j \oplus W_j \ ,$$

and, consequently,

$$\bigoplus_j W_j \ = \ L^2(\mathbb{R}) \ .$$

The symbol $\oplus$ stands for direct sum. A function ψ is a *wavelet* if the collection of functions $\{\psi(x-l) \mid l \in \mathbb{Z}\}$ is a Riesz basis of W_0. Since the wavelet ψ is also an element of V_1, *wavelet parameters* g_k exist such that

$$\psi(x) = 2 \sum_k g_k \, \phi(2x-k) \ . \tag{9}$$

Also here the wavelet has to satisfy

$$\int_{-\infty}^{+\infty} \psi(x) dx \ = \ 0 \ . \tag{10}$$

The collection of wavelet functions $\{\psi_{j,l} \mid l,j \in \mathbb{Z}\}$ now is a Riesz basis of $L^2(\mathbb{R})$.

The Fourier transform of the wavelet is given by

$$\hat{\psi}(\omega) \ = \ G(\omega/2) \, \hat{\psi}(\omega/2) \ , \tag{11}$$

with G a 2π-periodic function defined as

$$G(\omega) \ = \ \sum_k g_k \, e^{-ik\omega} \ .$$

From (9) and (10) we have

$$\sum_k g_k \ = \ 0 \quad \text{or} \quad G(0) \ = \ 0 \ .$$

Each space V_j and W_j has an $L^2(\mathbb{R})$ complement denoted by V_j^c and W_j^c, respectively. We have:

$$V_j^c = \bigoplus_{i=j}^{\infty} W_i \quad \text{and} \quad W_j^c = \bigoplus_{i \neq j} W_i \ .$$

We define P_j and Q_j as the projection operators onto V_j and W_j parallel to V_j^c or W_j^c, respectively. A function f can now be written as

$$f(x) \ = \ \sum_j Q_j f(x) \ = \ \sum_{j,l} \mu_{j,l} \, \psi_{j,l}(x) \ .$$

This can be seen as a discrete version of (4). The mapping from the function f to the coefficients $\mu_{j,l}$ is usually referred to as the *discrete wavelet transform*. How the coefficients $\mu_{j,l}$ are found will become clear in the following sections.

6 Orthogonal Wavelets

A particularly interesting class of wavelets consists off the *orthogonal wavelets*. We start their construction by introducing an orthonormal scaling function. This is a function ϕ such that

$$\langle \phi(x), \phi(x-l) \rangle \;=\; \delta_l \qquad\qquad l \in \mathbb{Z} \;. \tag{12}$$

As a result the collection of functions $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is an orthonormal basis of V_0 and the collection of functions $\{\phi_{j,l} \mid l \in \mathbb{Z}\}$ is an orthonormal basis of V_j. Using Poisson's formula, (12) follows from

$$\forall \omega \in \mathbb{R} \;:\; \sum_k |\hat{\phi}(\omega + k2\pi)|^2 \;=\; 1 \;. \tag{13}$$

Using (8), or the refinement equation, we can write this as a property of the scaling parameters

$$\forall \omega \in \mathbb{R} \;:\; |H(\omega)|^2 + |H(\omega+\pi)|^2 \;=\; 1 \;, \tag{14}$$

or

$$\sum_k h_k \, h_{k-2l} = \delta_l/2 \quad \text{for} \quad l \in \mathbb{Z} \;.$$

The last two equations are equivalent but they provide only a necessary condition for the orthogonality of the scaling function and its translates. This relationship is investigated in [45] .

We define the wavelet spaces W_j here as the *orthogonal complement* of V_j in V_{j+1}. An *orthogonal wavelet* is now a function ψ such that the collection of functions $\{\psi(x-l) \mid l \in \mathbb{Z}\}$ is an orthonormal basis of W_0 and consequently the collection of functions $\{\psi_{j,l} \mid j, l \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$. This is the case if:

$$\langle \psi(x), \psi(x-l) \rangle \;=\; \delta_l \quad \text{and} \quad \langle \psi(x), \phi(x-l) \rangle \;=\; 0 \qquad\qquad l \in \mathbb{Z} \;.$$

These conditions are a consequence from

$$\forall \omega \in \mathbb{R} \;:\; \sum_k |\hat{\psi}(\omega + k2\pi)|^2 \;=\; 1 \;,$$

and

$$\forall \omega \in \mathbb{R} \;:\; \sum_k \hat{\psi}(\omega + k2\pi)\, \overline{\hat{\phi}(\omega + k2\pi)} \;=\; 0 \;.$$

Again a necessary condition is given by

$$\forall \omega \in \mathbb{R} \;:\; |G(\omega)|^2 + |G(\omega+\pi)|^2 \;=\; 1 \;,$$

and

$$\forall \omega \in \mathbb{R} \;:\; G(\omega)\,\overline{H(\omega)} + G(\omega+\pi)\,\overline{H(\omega+\pi)} = 0 \;.$$

From this last equation we see that a possible choice for the function $G(\omega)$ is

$$G(\omega) = -e^{-i\omega}\,\overline{H(\omega+\pi)}\ .$$

This means that we can derive an orthogonal wavelet from an orthogonal scaling function by choosing

$$g_k = (-1)^k \overline{h_{1-k}}\ . \tag{15}$$

Now, using the condition (14) and the fact that $H(0) = G(\pi) = 1$ and $G(0) = H(\pi) = 0$, we see that $H(\omega)$ essentially represents a low-pass filter $[0, \pi/2]$ and $G(\omega)$ represents a band-pass filter $[\pi/2, \pi]$. Then, from (8) and (11) we conclude that the main part of the energy of $\hat{\phi}(\omega)$ and $\hat{\psi}(\omega)$ is concentrated in the intervals $[0, \pi]$ and $[\pi, 2\pi]$, respectively. This means that the wavelet expansion essentially splits the frequency space into dyadic blocks $[2^j\pi, 2^{j+1}\pi]$ with $j \in \mathbb{Z}$. The time-frequency localization is one of the most important characteristics of the wavelet transform.

In [46] an orthonormalization procedure to find orthonormal wavelets is proposed. It states that if a function g and its integer translates are a Riesz basis of V_0, then an orthonormal basis of V_0 is given by ϕ and its integer translates with

$$\hat{\phi}(\omega) = \frac{\hat{g}(\omega)}{\sqrt{\sum_k |\hat{g}(\omega + k2\pi)|^2}}\ . \tag{16}$$

The fact that we started from a Riesz basis guarantees that the denominator is non-zero. We see that ϕ now indeed satisfies (13). The projection operators P_j and Q_j are now orthogonal projections and can be written as

$$P_j f(x) = \sum_l \langle f, \phi_{j,l} \rangle\, \phi_{j,l}(x) \quad \text{and} \quad Q_j f(x) = \sum_l \langle f, \psi_{j,l} \rangle\, \psi_{j,l}(x)\ .$$

They are the best $L^2(\mathbb{R})$ approximations of a function in V_j or W_j. For every function $f \in L^2(\mathbb{R})$ we have now an orthogonal expansion

$$f(x) = \sum_{j,l} \mu_{j,l}\, \psi_{j,l}(x) \quad \text{with} \quad \mu_{j,l} = \langle f, \psi_{j,l} \rangle\ .$$

Again, this can be viewed as a discrete version of the continuous wavelet transform.

Examples of orthogonal wavelets:

- Two simple examples of orthogonal scaling functions are the box function and the Shannon sampling function. The orthogonality conditions are trivial to verify here either in time or frequency space. The corresponding wavelet for the box function is the Haar wavelet

$$\psi(x) = \chi_{[0,1/2]}(x) - \chi_{[1/2,1]}(x)\ ,$$

and the Shannon wavelet is

$$\psi(x) = \frac{\sin(2\pi x) - \sin(\pi x)}{\pi x} .$$

These two, however, are not very useful in practice since the first has very low regularity and the second has very slow decay.

- A more interesting example is the Meyer wavelet [32, 56] . This function is C^∞ and has faster-than-polynomial decay. Its Fourier transform is compactly supported and it has an infinite number of vanishing moments which makes it particularly suited for singularity detection. It is also symmetric around $x = 1/2$.
- The Battle-Lemarié wavelet is constructed by orthogonalizing B–spline functions using (16) and has exponential decay [6, 46]. The wavelet of order m is a piecewise polynomial of degree $m-1$ that belongs to C^{m-2}.
- Probably the most commonly used wavelets are the Daubechies wavelets [23, 24]. They are a family of wavelets indexed by $N \in \mathbb{N}$, where N is the number of vanishing wavelet moments. They have compact support $[-N+1, N]$, which makes them particularly useful in engineering applications. A disadvantage is that they cannot be symmetric or antisymmetric and thus cannot have generalized linear phase. This can lead to distortion in filtering. Their regularity increases linearly with N and is approximately equal to $0.3N$.

7 Biorthogonal Wavelets

The orthogonality property puts a strong limitation on the construction of wavelets. As was mentioned in the previous section compactly supported, symmetric, orthogonal wavelets do not exist. Hence, the generalization to *biorthogonal wavelets* has been introduced. In this case, dual scaling functions and wavelets $\tilde{\phi}_{j,l}$ and $\tilde{\psi}_{j,l}$ exist such that

$$\langle \phi_{j,l}, \tilde{\phi}_{j,l'} \rangle = \delta_{l-l'} \qquad l, l', j \in \mathbb{Z} , \tag{17}$$

and

$$\langle \psi_{j,l}, \tilde{\psi}_{j',l'} \rangle = \delta_{j-j'}\delta_{l-l'} \qquad l, l', j, j' \in \mathbb{Z} . \tag{18}$$

Notice that again the biorthogonality for the scaling functions is only needed on each level separately and for the wavelets on all levels at the same time. The dual functions also generate a multiresolution analysis $\tilde{V}_j$ and $\tilde{W}_j$. This is not necessarily the same as the one generated by the primary functions. The dual functions also satisfy:

$$\tilde{\phi}(x) = 2\sum_k \tilde{h}_k \tilde{\phi}(2x-k) \quad \text{and} \quad \tilde{\psi}(x) = 2\sum_k \tilde{g}_k \tilde{\phi}(2x-k) . \tag{19}$$

The functions $\tilde{H}$ and $\tilde{G}$ are defined similarly to H and G. The biorthogonality conditions can now be written as

$$\forall \omega \in \mathbb{R} : \begin{cases} \tilde{H}(\omega)\overline{H(\omega)} + \tilde{H}(\omega+\pi)\overline{H(\omega+\pi)} = 1 \\ \tilde{G}(\omega)\overline{G(\omega)} + \tilde{G}(\omega+\pi)\overline{G(\omega+\pi)} = 1 \\ \tilde{G}(\omega)\overline{H(\omega)} + \tilde{G}(\omega+\pi)\overline{H(\omega+\pi)} = 0 \\ \tilde{H}(\omega)\overline{G(\omega)} + \tilde{H}(\omega+\pi)\overline{G(\omega+\pi)} = 0 \end{cases},$$

or

$$\forall \omega \in \mathbb{R} : \begin{bmatrix} \tilde{H}(\omega) & \tilde{H}(\omega+\pi) \\ \tilde{G}(\omega) & \tilde{G}(\omega+\pi) \end{bmatrix} \overline{\begin{bmatrix} H(\omega) & G(\omega) \\ H(\omega+\pi) & G(\omega+\pi) \end{bmatrix}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

or

$$\forall \omega \in \mathbb{R} : \begin{cases} \overline{H(\omega)}\tilde{H}(\omega) + \overline{G(\omega)}\tilde{G}(\omega) = 1 \\ \overline{H(\omega)}\tilde{H}(\omega+\pi) + \overline{G(\omega)}\tilde{G}(\omega+\pi) = 0 \end{cases}. \tag{20}$$

The projection operators take the form

$$P_j f(x) = \sum_l \langle f, \tilde{\phi}_{j,l} \rangle \phi_{j,l}(x) \quad \text{and} \quad Q_j f(x) = \sum_l \langle f, \tilde{\psi}_{j,l} \rangle \psi_{j,l}(x) .$$

From (17), (18) and (19) we see that

$$\tilde{h}_{k-2l} = \langle \tilde{\phi}(x-l), \phi(2x-k) \rangle \quad \text{and} \quad \tilde{g}_{k-2l} = \langle \tilde{\psi}(x-l), \phi(2x-k) \rangle ,$$

such that

$$\phi(2x-k) = \sum_l \tilde{h}_{k-2l}\, \phi(x-l) + \sum_l \tilde{g}_{k-2l}\, \psi(x-l) . \tag{21}$$

and since primary and dual functions are interchangeable,

$$\tilde{\phi}(2x-k) = \sum_l h_{k-2l}\, \tilde{\phi}(x-l) + \sum_l g_{k-2l}\, \tilde{\psi}(x-l) . \tag{22}$$

A class of wavelets somewhere between orthogonal and biorthogonal wavelets are the so–called *semi-orthogonal wavelets* or *prewavelets* [42, 58, 59, 62]. In this case

$$\langle \psi_{j,l}(x), \psi_{j',l'}(x) \rangle = 0 \quad \text{for} \quad j \neq j' .$$

This means that the wavelet spaces W_j are still mutually orthogonal, but the wavelets of a certain level j are not orthogonal. This has the advantage that the projection operators P_j and Q_j are still orthogonal, and the expansion

$$f(x) = \sum_j Q_j f(x)$$

is an orthogonal expansion. Since here not only $W_j \perp \tilde{W}_{j'}$ for $j \neq j'$, but also $W_j \perp W_{j'}$ for $j \neq j'$, we have that $W_j = \tilde{W}_j$ and $V_j = \tilde{V}_j$ for $j \in \mathbb{Z}$ and thus that primary and dual functions generate the same multiresolution analysis. This means that under certain conditions a dual scaling function can be found by letting

$$\hat{\tilde{\phi}}(\omega) = \frac{\hat{\phi}(\omega)}{\sum_k |\hat{\phi}(\omega + k2\pi)|^2} .$$

Examples:

- Typical examples of biorthogonal wavelets are the ones developed by Cohen, Daubechies and Feauveau [14, 16]. The scaling functions are the cardinal B-splines and the wavelets are spline functions too. All functions including the dual ones have compact support and linear phase. Moreover, all scaling and wavelet parameters are rationals which property is of use in hardware implementations. A disadvantage is that the dual functions have very low regularity.
- Examples of semi-orthogonal wavelets are the ones constructed by Chui and Wang [12]. The scaling functions are cardinal B–splines of order m and the wavelet functions are splines with compact support $[0, 2m-1]$. All primary and dual functions still have generalized linear phase and all scaling and wavelet parameters are rationals. A powerful feature here is that analytic expressions for the wavelet, scaling function and dual functions are available. A disadvantage is that the dual functions do not have compact support; they have exponential decay. The same wavelets, but in a different setting, were derived by Aldroubi, Eden and Unser [70].

8 Wavelets and Polynomials

The moments of the scaling function and wavelet are defined as:

$$\mathcal{M}_p = \langle x^p, \phi(x) \rangle \quad \text{and} \quad \mathcal{N}_p = \langle x^p, \psi(x) \rangle \quad \text{with} \quad p \geq 0 \ .$$

These inner products only make sense if ϕ and ψ have sufficient decay. The scaling function has $\mathcal{M}_0 = 1$. Recursion formulae to calculate these moments are derived in [69]. The number of vanishing wavelet moments is denoted by N where N is at least 1: $\mathcal{N}_p = 0$ for $0 \leq p < N$, and $\mathcal{N}_N \neq 0$. This can also be written as

$$\hat{\psi}^{(p)}(0) = 0 \quad \text{or} \quad G^{(p)}(0) = 0 \quad \text{for} \quad 0 \leq p < N \ .$$

Similar definitions and equations hold for the dual functions, involving $\tilde{\mathcal{M}}_p$, $\tilde{\mathcal{N}}_p$, $\tilde{N}$, $\tilde{G}(\omega)$. We have already seen that the number of vanishing wavelet moments is important for the characterization of singularities. It also defines the convergence rate of the wavelet approximation for smooth functions [27, 65, 66], since if $f \in C^{\tilde{N}}$, then

$$\| P_j f(x) - f(x) \| = \mathcal{O}(h^{\tilde{N}}) \quad \text{with} \quad h = 2^{-j} \ .$$

An asymptotic error expansion in powers of h which can be used in numerical extrapolation is derived in [68]. There it is also shown that the wavelet approximation of a smooth function interpolates the functions in twice the number of points as compared to the number of basis functions. Another way to look at $\tilde{N}$ is:

Proposition 1. *For every $j \in \mathbb{Z}$, any polynomial with degree smaller than $\tilde{N}$ can be written as a linear combination of the functions $\phi_{j,l}$ with $l \in \mathbb{Z}$.*

9 The Fast Wavelet Transform Algorithm

Since V_j is equal to $V_{j-1} \oplus W_{j-1}$, a function $v_j \in V_j$ can be written uniquely as the sum of a function $v_{j-1} \in V_{j-1}$ and a function $w_{j-1} \in W_{j-1}$:

$$v_j(x) = \sum_k \nu_{j,k}\, \phi_{j,k}(x) \;=\; v_{j-1}(x) + w_{j-1}(x)$$
$$= \sum_l \nu_{j-1,l}\, \phi_{j-1,l}(x) + \sum_l \mu_{j-1,l}\, \psi_{j-1,l}(x) \;.$$

There is a one-to-one relationship between the coefficients of these functions.

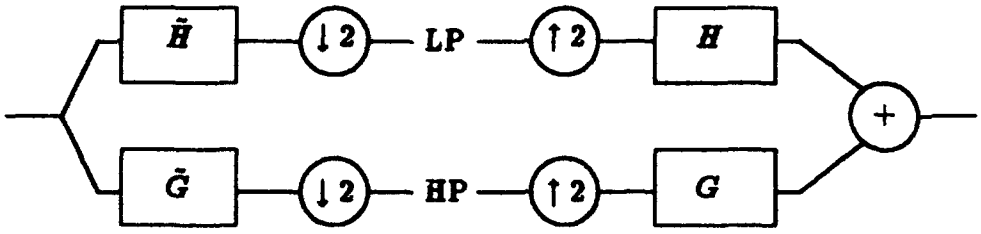


Fig. 1. The subband coding scheme.

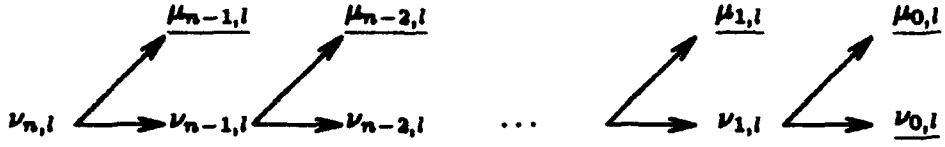


Fig. 2. The decomposition scheme.

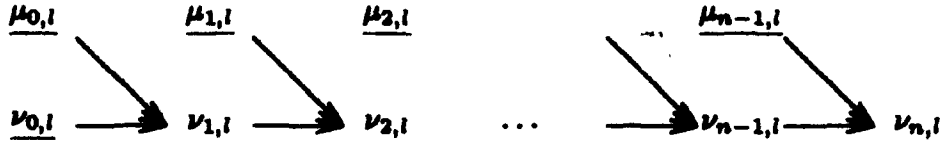


Fig. 3. The reconstruction scheme.

The decomposition formulae can be found using (19),

$$\nu_{j-1,l} = \langle v_j, \phi_{j-1,l} \rangle = \sqrt{2} \langle v_j, \sum_k \tilde{h}_{k-2l} \phi_{j,k} \rangle$$

$$= \sqrt{2} \sum_k \tilde{h}_{k-2l} \nu_{j,k} \ , \tag{23}$$

and similarly,

$$\mu_{j-1,l} = \sqrt{2} \sum_k \tilde{g}_{k-2l} \nu_{j,k} \ . \tag{24}$$

The reconstruction step involves calculating the $\nu_{j,k}$ from the $\nu_{j-1,l}$ and the $\mu_{j-1,l}$. Using (21) we have

$$\nu_{j,k} = \sqrt{2} \sum_l h_{k-2l} \nu_{j-1,l} + \sqrt{2} \sum_l g_{k-2l} \mu_{j-1,l} \ . \tag{25}$$

When applied recursively, these formulae define a transformation, the *fast wavelet transform* [53, 54]. In signal processing this technique is known as subband coding, see Fig. 1. The decomposition step consists of applying a low-pass ($\tilde{H}$) and a band-pass ($\tilde{G}$) filter followed by downsampling ($\downarrow 2$) (i.e. retaining only the even index samples). The reconstruction consists of upsampling ($\uparrow 2$) (i.e. adding a zero between every two samples) followed by filtering and addition. One can show that the conditions (20) correspond to the exact reconstruction of a subband coding scheme.

A multiresolution analysis on the interval $[0,1]$ can be constructed by periodising the basis functions and defining:

$$\phi^*_{j,l}(x) = \chi_{[0,1]}(x) \sum_m \phi_{j,l}(x+m) \quad \text{for} \quad 0 \le l < 2^j \quad \text{and} \quad j \ge 0 \ . \tag{26}$$

Similar definitions hold for $\psi^*_{j,l}$, $\tilde{\phi}^*_{j,l}$ and $\tilde{\psi}^*_{j,l}$.

In the description of the algorithm we assume that the h_k, $\tilde{h}_k$ are non-zero for $-L \le k \le L$, and that the $\tilde{g}_k$, g_k are non-zero for $-M \le k \le M$ with L and M odd ($L = 2L'+1, M = 2M'+1$). We start with 2^n coefficients $\nu_{n,l}$ of a function of V_n and can thus apply n steps of the algorithm. How these coefficients $\nu_{n,l}$ can be calculated efficiently from function evaluations of f is described in [69].

for $j \leftarrow n-1\,(-1)\,0$

  for $l \leftarrow 0\,(1)\,2^j - 1$

$$\nu_{j,l} \leftarrow \sqrt{2} \sum_{k=-L}^{L} \tilde{h}_k \nu_{j+1,(k+2l) \bmod 2^{j+1}}$$

$$\mu_{j,l} \leftarrow \sqrt{2} \sum_{k=-M}^{M} \tilde{g}_k \nu_{j+1,(k+2l) \bmod 2^{j+1}}$$

  end for

end for

The reconstruction algorithm can be deduced from formula (25):

for $j \leftarrow 1\,(1)\,n$

  for $k \leftarrow 0\,(1)\,2^j - 1$

    if k even then

$$\nu_{j,k} \leftarrow \sqrt{2} \sum_{l=-L'}^{L'} h_{2l}\, \nu_{j-1,(k/2-l) \bmod 2^{j-1}} + \sqrt{2} \sum_{l=-M'}^{M'} g_{2l}\, \mu_{j-1,(k/2-l) \bmod 2^{j-1}}$$

    else (k odd)

$$\nu_{j,k} \leftarrow \sqrt{2} \sum_{l=-L'-1}^{L'} h_{2l+1}\, \nu_{j-1,((k-1)/2-l) \bmod 2^{j-1}} + \sqrt{2} \sum_{l=-M'-1}^{M'} g_{2l+1}\, \mu_{j-1,((k-1)/2-l) \bmod 2^{j-1}}$$

    endif

  end for

end for

For L or M even, the reconstruction algorithm has to be modified slightly.

10 Multidimensional Wavelets

Up till now we have focused on the one-dimensional situation. However, there are also wavelets in higher dimensions. A simple way to obtain these is to use tensor products. To fix ideas, let us consider the case of the plane. Let

$$\Phi(x,y) = \phi(x)\,\phi(y) = \phi \otimes \phi(x,y) \ ,$$

and define

$$V_0 = \{f : f(x,y) = \sum_{k_1,k_2} \lambda_{k_1,k_2}\, \Phi(x-k_1, y-k_2), \lambda \in l^2(\mathbb{Z}^2)\} \ .$$

Of course, if $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is an orthonormal set, then $\{\Phi(x-k_1, y-k_2)\}$ form an orthonormal basis for V_0. By dyadic scaling we obtain a multiresolution analysis of $L^2(\mathbb{R}^2)$. The complement W_0 of V_0 in V_1 is similarly generated by the translates of the three functions

$$\Psi^{(1)} = \phi \otimes \psi, \quad \Psi^{(2)} = \psi \otimes \phi, \quad \text{and} \quad \Psi^{(3)} = \psi \otimes \psi \ . \tag{27}$$

There is another, perhaps even more straightforward, wavelet decomposition in higher dimensions. By carrying out a one-dimensional wavelet decomposition for each variable separately, we obtain

$$f(x,y) = \sum_{i,l} \sum_{j,k} \langle f, \psi_{i,l} \otimes \psi_{j,k} \rangle\, \psi_{i,l} \otimes \psi_{j,k}(x,y) \ . \tag{28}$$

Note that the functions $\psi_{i,\cdot} \otimes \psi_{j,k}$ involve two scales, 2^{-i} and 2^{-j}, and each of these functions are (essentially) supported on a rectangle. The decomposition (28) is the *rectangular wavelet decomposition* of f while the functions (27) are the basis functions of the *square wavelet decomposition.*

There are also several other extensions to higher dimensions, e.g. non-separable basis functions [15, 62, 64], other lattices corresponding to different symmetries [18] and Clifford valued wavelets. However we leave these topics for now.

11 Wavelets on Closed Sets

So far we have been discussing wavelet theory on the real line (and its higher dimensional analogues). For many applications the functions involved are only defined on a finite set, such as an interval or a square, and to apply wavelets then requires some modifications.

To be specific, let us discuss the case of the unit interval $[0,1]$. Given a function f on $[0,1]$, the most obvious approach is to set $f(x) = 0$ outside $[0,1]$, and then use wavelet theory on the line. However, for a general function f this "padding with 0s" introduces discontinuities at the endpoints 0 and 1; consider for example the simple function $f(x) = 1$, $x \in [0,1]$. Now, as we have said earlier, wavelets are effective for detecting singularities, so artificial ones are likely to introduce significant errors.

Another approach, which is often better, is to consider the function to be periodic with period 1, $f(x+1) = f(x)$. This was done in the description of the algorithms of the fast wavelet transform. Expressed in another way, we assume that the function is defined on the torus and identify the torus with $[0,1]$. Wavelet theory on the torus parallels that on the line. In fact, note that if f has period 1, then the the wavelet coefficients on a given scale satisfy $\langle f, \psi_{j,k} \rangle = \langle f, \psi_{j,k+2^j} \rangle$, $k \in \mathbb{Z}, j \geq 0$. This simple observation readily allows us to rewrite wavelet expansions on the line as analogous ones on the torus, with wavelets defined on $[0,1]$. This "wrap around" procedure is satisfactory in many situations (and certainly takes care of functions like $f(x) = 1$, $x \in [0,1]$, for example). However, unless the behaviour of the function f at 0 matches that at 1, then the periodic version of f will have singularities there. A simple function like $f(x) = x$, $x \in [0,1]$, gives a good illustration of this.

What really is needed then are wavelets intrinsically defined on $[0,1]$. Such wavelets were recently given by Meyer [57], and we shall very briefly sketch his construction next. We start from the Daubechies wavelets and a scaling function with $2N$ non-zero coefficients:

$$\phi(x) = 2 \sum_{k=0}^{2N-1} h_k \, \phi(2x-k) \; . \tag{29}$$

It easy to see that $\mathrm{supp}(\phi) = [0, 2N-1]$, and, as a consequence, there are $2^j + 2N - 2$ functions whose supports intersect with $(0,1)$ if $j \geq 0$. For sufficiently

small scales, $j \geq j_0$, say, a function ϕ_{jk} can only intersect at most one of the endpoints 0 or 1. We now let $V_j^{[0,1]}$ denote the restriction of functions in V_j:

$$V_j^{[0,1]} = \{f : f(x) = g(x), x \in [0,1], \text{for some function } g \in V_j\} \ .$$

The $V_j^{[0,1]}$, $j \geq j_0$, form a multiresolution analysis of $L^2([0,1])$. It is also obvious that the functions in $\{\phi(x-l)|_{[0,1]} : -2N+2 \leq l \leq 2^j - 1\}$ span $V_j^{[0,1]}$. Here $g(x)|_{[0,1]}$ denotes the restriction of $g(x)$ to $[0,1]$. Not quite as obvious, but still easy, is the fact that the functions in this collection are linearly independent and, hence, form a basis for $V_j^{[0,1]}$. In order to obtain an orthonormal basis, we may argue as follows. As long as $\text{supp}(\phi_{jk}) \subset [0,1]$, restricting it to $[0,1]$ does not affect it. The orthogonality is only violated for the functions whose support intersects an endpoint and can be re-established with a Gram-Schmidt procedure.

Now, if we let $W_j^{[0,1]}$ denote the restriction of functions in W_j to $[0,1]$, then we of course have that $V_{j+1}^{[0,1]} = V_j^{[0,1]} + W_j^{[0,1]}$. So, the basis elements in $V_j^{[0,1]}$ together with the restriction of the wavelets ψ_{jk} to $[0,1]$ span $V_{j+1}^{[0,1]}$. However there are $2^j + 2N - 2$ wavelets that intersect $(0,1)$, and since $\dim V_{j+1}^{[0,1]} - \dim V_j^{[0,1]} = 2^j$ we in fact have too many functions. The restrictions of the wavelets in W_j whose support is a subset of $[0,1]$, are still mutually orthogonal and they are also orthogonal to $V_j^{[0,1]}$. Among the functions which intersect an endpoint, we use (21) to find the redundant ones and remove them. After that we just apply a Gram-Schmidt argument again, and we have an orthonormal basis for $W_j^{[0,1]}$.

Meyer's elegant construction has a couple of disadvantages. Among the functions ϕ_{jk} that intersect an endpoint there are some that are almost zero there. Hence, some functions are almost linearly dependent, and, as a consequence, the condition number of the matrix, corresponding to the change of basis to the orthonormal one, becomes quite large. Furthermore, we have $\dim V_j^{[0,1]} \neq \dim W_j^{[0,1]}$ which means that there is an inherit imbalance between these spaces, which is not present in the case of the whole real line.

As was noted earlier (proposition 1) all polynomials of degree $\leq N-1$ are in V_j. Hence, the restriction of such polynomials to $[0,1]$ are in $V_j^{[0,1]}$. Since this fact is directly linked to many of the approximation properties of wavelets, any construction of a multiresolution analysis on $[0,1]$ should preserve this. The construction in [17] uses this as a starting point and is slightly different than the one by Meyer. This construction starts again with the scaling function ϕ from the Daubechies construction with $2N$ non-zero scaling parameters, and assumes a sufficiently fine scale so that the endpoints are independent as before. Now, at each boundary, N specific linear combinations of the $2N-1$ functions whose support intersects an endpoint are taken such that each polynomial of degree smaller than N can still be written as a linear combination of these $2N$ new functions at the boundary and the $2^j - 2N$ interior functions whose support is a subset of $(0,1)$. Again an orthonormal basis is obtained by orthogonalizing

the boundary functions. It is easy to see that the spaces V_j^* generated by these functions are nested and define a multiresolution analysis.

To get to the corresponding wavelets is straightforward as well. We let W_j^* be the orthogonal complement of V_j^* in V_{j+1}^*. There are $2^j - 2N$ interior wavelets. The remaining $2N$ functions required for an orthonormal basis of W_j can easily be found, for example by using (21) again. The dimension of a space at scale 2^{-j} is now 2^j. This last construction also carries over to more general situations [39]; for example we can use biorthogonal wavelets and also much more general closed sets than $[0, 1]$.

There are also other constructions of wavelets on $[0, 1]$. In fact, for historical perspective it is interesting to notice that Franklin's original construction [28] was given for $[0, 1]$. Another interesting one, in the case of semi-orthogonal spline-wavelets, has been given by Chui and Quak [9] (see the original paper for details).

12 Applications

12.1 Data Compression

One of the applications of wavelet theory is data or image compression. There are two basic kinds of compression schemes: lossless and lossy. In the case of lossless compression one is interested in reconstructing a message or image exactly, without any loss of information. We shall consider lossy compression. In this case, we are ready to accept an error as long as the quality after compression is acceptable. To be specific, let us assume that we are given a digitized image. With lossy compression schemes we potentially can achieve much higher compression ratios than with lossless compression. The compression ratio is defined as the number of bits the initial image takes to store on the computer divided by the number of bits required to store the compressed image. The interest in compression in general has grown as the amount of information we pass around has increased. This is easy to understand when we consider the fact that to store a moderately large image, say a 512×512 pixels, 24-bit colour image, takes about 0.75 MBytes. This is only for still images; in the case of video, the situation becomes even worse. Then we need this kind of storage for each frame and we have something like 30 frames per second. There are several other reasons than just the storage requirement for the interest in compression techniques. However, instead of going into this, let us now look at the connection with wavelet theory.

Fig. 4. Image transform coding

First, let us define, somewhat mathematically, what we mean by an image. Let us for simplicity discuss an $L \times L$ grey-scale image with 256 grey-scales (i.e. 8 bit). This can be considered to be a piecewise constant function f defined on a square

$$f(x,y) = p_{ij}, \quad \text{for} \quad i \le x < i+1 \quad \text{and} \quad j \le y < j+1 \quad \text{and} \quad 0 \le i,j < L \ ,$$

where $0 \le p_{ij} \le 255$ are integers. Now, one of the standard procedures for lossy compression is through transform coding, see Fig. 4. The most common transform used in this context is the "Discrete Cosine Transform" which uses a Fourier transform of the image f. However, we are more interested in the case when the transform is the wavelet transform.

There are in fact several ways to use the wavelet transform for compression purposes [52, 51]. One way is to consider compression to be an approximation problem [26]. More specifically, let us fix an orthogonal wavelet ψ. Given an integer $M \ge 1$ we try to find the "best" approximation of f by using a representation

$$f_M(x) = \sum_{kl} b_{jk}\psi_{jk} \quad \text{with } M \text{ non-zero coefficients } b_{jk}. \tag{30}$$

The basic reason why this potentially might be useful is that each wavelet picks up information about the image f essentially at a given location and at a given scale. Where the image has more interesting features, we can spend more coefficients, and where the image is nice and smooth we can use fewer and still get good quality of approximation. In other words, the wavelet transform allows us to focus on the most relevant parts of f. Now, to give this mathematical meaning we need to agree on an error measure. Ideally, for image compression we should use a norm that corresponds as closely as possible to the human eye. However, let us make it simple and discuss the case of L^2.

So we are interested in finding an optimal approximation minimizing the error $\|f - f_M\|_{L^2}$. Because of the orthogonality of the wavelets this equals

$$\Big(\sum_{jk} |\langle f, \psi_{jk}\rangle - b_{jk}|^2\Big)^{1/2} \ . \tag{31}$$

A moment's thought, reveals that the best way to pick M non-zero coefficients b_{jk}, making the error as small as possible, is by simply picking the M coefficients with the largest absolute value, and setting $b_{j,k} = \langle f, \psi_{jk}\rangle$ for these numbers. This then yields the optimal approximation f_M^{opt}.

Another fundamental question is which images can be approximated well by using the procedure just sketched? Let us take this to mean that the error satisfies

$$\|f - f_M^{opt}\| = \mathcal{O}(M^{-\beta}) \ , \tag{32}$$

for some $\beta > 0$. The larger β, the fewer coefficients are generally needed to obtain a certain error. The exponent β can be found quite easily; a simple argument

shows that in fact

$$\left(\sum_{M\geq 1}(M^{\beta}\|f-f_M^{opt}\|_{L^2})^p\frac{1}{M}\right)^{1/p}\approx(\sum_{jk}|\langle f,\psi_{jk}\rangle|^p)^{1/p}\ , \qquad (33)$$

with $1/p = 1/2 + \beta$. The maximal β for which (32) is valid can be estimated by finding the smallest p for which the right-hand side of (33) is finite. The expression on the right is one of many equivalent norms on the Besov space $\dot{B}_p^{2\beta,p}$; recall that Besov spaces are smoothness spaces generalizing the Lipschitz continuous functions (which is the case $p = q = \infty$). However, the β in the left-hand side of (33) is not exactly the same as in (32). For practical purposes, the difference is of no consequence.

12.2 Numerical Analysis

As mentioned earlier, one interest in wavelets has historically come from the fact that they are effective tools for studying problems in partial differential equations and operator theory. More specifically, they are useful for understanding properties of Calderón-Zygmund operators.

Let us first make a general observation about the representation of a linear operator T and wavelets. Suppose that f has the representation

$$f(x)=\sum_{jk}\langle f,\psi_{jk}\rangle\psi_{jk}(x)\ .$$

Then

$$Tf(x)=\sum_{jk}\langle f,\psi_{jk}\rangle T\psi_{jk}(x)\ ,$$

and, using the wavelet representation of the function, this equals

$$\sum_{jk}\langle f,\psi_{jk}\rangle\sum_{il}\langle T\psi_{jk},\psi_{il}\rangle\psi_{il}(x)=\sum_{il}\sum_{jk}\langle T\psi_{jk},\psi_{il}\rangle\ \langle f,\psi_{jk}\rangle\psi_{il}(x)\ .$$

In other words, the action of the operator T on the function f is directly translated into the action of the "matrix" $A_T = \{\langle T\psi_{jk},\psi_{il}\rangle\}_{il,jk}$ on the sequence $\{\langle f,\psi_{jk}\rangle\}_{jk}$. This representation of T as the matrix A_T is often referred to as the "standard representation" of T [7]. There is also a "nonstandard representation"; for virtually all linear operators there is a function (or, more generally, a distribution) $K = K_T$ such that

$$Tf(x)=\int K(x,y)f(y)\,dy\ .$$

The "nonstandard representation" of T is now simply the (two-dimensional) wavelet coefficients of the kernel K, using the square decomposition $\{\langle K,\Psi^{(j)}_{k_1,k_2}\rangle\}$ (again, we have more than one wavelet function in two dimensions).

Let us then briefly discuss the connection with Calderón-Zygmund operators. Consider now a typical example. Let H be the *Hilbert transform*,

$$Hf(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(s)}{x-s}\, ds \ .$$

The basic idea is now that the wavelets ψ_{jk} are approximate eigenfunctions for this, as well as for many other related (Calderón-Zygmund) operators. We note that if ψ_{jk} were exact eigenfunctions, then we would have $H\psi_{jk}(x) = \lambda_{jk}\psi_{jk}(x)$, for some number λ_{jk} and the standard representation would be a diagonal "matrix":

$$A_H = \{ \langle H\psi_{il}, \psi_{jk} \rangle \} = \{\lambda_{il} \langle \psi_{il}, \psi_{jk} \rangle \} = \{\lambda_{il}\delta_{il,jk}\} \ .$$

This is unfortunately not quite the case. However, it turns out that A_T is in fact an almost-diagonal operator, in the appropriate, technical sense, with the off-diagonal elements quickly becoming small. To get some idea why this is the case, note that for large $|x|$, we have, at least heuristically,

$$H\psi_{jk}(x) \approx \frac{1}{x} \int \psi_{jk}(y)\, dy \ .$$

A priori, the decay of the right-hand side would thus be $\mathcal{O}(1/x)$, which of course is far from the rapid decay of a wavelet ψ_{jk} (recall that some wavelets are even zero outside a finite set). Recall, however, that ψ_{jk} has at least one vanishing moment so the decay is in fact much faster than just $\mathcal{O}(1/x)$, and the shape of $H\psi_{jk}(x)$ closely resembles that of $\psi_{jk}(x)$.

So, for a large class of operators the matrix representation, either the standard or the nonstandard, has a rather precise structure with many small elements. In this representation, we then expect to be able to compress the operator by simply omitting small elements. In fact, note that this is essentially the same situation, especially in the case of the nonstandard representation, as in the case of image compression, the "image" now being the kernel $K(x,y)$. Hence, if we could do basic operations such as inversion, and multiplication, with compressed matrices, rather than with the discritized versions of T, then we may significantly speed up the numerical treatment. This program of using the wavelet representations for the efficient numerical treatment of operators was initiated in [7]. We also refer to [1, 2] for related material and many more details.

In a different direction, because of the close similarities between, on the one hand, the scaling function and its translates and dilates, and finite elements, on the other, it seems natural to try wavelets where traditionally finite element methods are used, for example for solving boundary value problems [40]. There are interesting results showing that this might be fruitful; for example, it has been shown [8, 22, 60] that for many problems the condition number of the $N \times N$ stiffness matrix remains bounded as the dimension N goes to infinity. This is in contrast with the situation for regular finite elements where the condition number in general tends to infinity.

One of the first problems we have to address when discussing boundary problems on domains is how to take care of the boundary values and the fact that the problem is closely associated with a finite set rather than with the entire Euclidean plane. This is similar to the problem we discussed with wavelets on closed sets, and, indeed, the techniques discussed there can be often used to handle these two problems [3, 4].

Wavelets have also been used in the solution of evolution equations [5, 33, 44, 49]. A typical test problem here is Burger's equation: $\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}$. The time discretization is obtained here using standard schemes such as Crank-Nicholson or Adams-Mouton. Wavelets are used in the space discretization. Adaptivity can be used both in time and space [5].

One of the nice features of wavelets and finite elements is that they allow us to treat a large class of operators or partial differential equations in a unified way, allowing for example general pde solvers to be designed. In specific instances, though, it is sometimes possible to find particular wavelets, adapted to the operator or problem at hand. For exampie, Dahlke and Weinrich constructed wavelets adapted specifically to each operator in a general class of pseudo-differential operators [21].

References

1. Alpert, B., Beylkin, G., Coifman, R., Rokhlin, V. (1990). Wavelets for fast resolution of second-kind integral equations, Technical Report Yaleu/dcs/rr-837, Department of Computer Science, Yale University.
2. Alpert, B. K. (1992). Wavelets and other bases for fast numerical linear algebra. In: Chui, C. K. (ed.), Wavelets: A Tutorial in Theory and Applications, Academic Press, pp. 181–216.
3. Andersson, L., Hall, N., Jawerth, B., Peters, G. (1993). Wavelets on closed subsets of the real line, In: Schumacher, L. L., Webb, G. (eds.), Topics in the Theory and Applications of Wavelets, Academic Press, to appear.
4. Auscher, P. (1992). Wavelets with boundary conditions on the interval. In: Chui, C. K. (ed.), Wavelets: A Tutorial in Theory and Applications, Academic Press, pp. 217–236.
5. Bacry, E., Mallat, S., Papanicolaou, G. (1991). A wavelet based space-time adaptive numerical method for partial differential equations, Technical Report 591, Courant Institute of Mathematical Sciences.
6. Battle, G. (1987). A block spin construction of ondelettes, Comm. Math. Phys. 110, pp. 601–615.
7. Beylkin, G., Coifman, R., Rokhlin, V. (1991). Fast wavelet transforms and numerical algorithms I, Comm. Pure and Applied Math. 44, pp. 141–183.
8. Beylkin, G., Coifman, R., Rokhlin, V. (1992). Wavelets in numerical analysis, In: Ruskai, M. B., *et al.* (eds.), Wavelets and Their Applications, Jones and Bartlett, pp. 181–210.
9. Chui, C., Quak, E. (1992). Wavelets on a bounded interval. In: Braess, D., Schumaker, L.L. (eds.), Numerical Methods of Approximation Theory, Birkhäuser Verlag, Basel, pp. 1–24.
10. Chui, C. K. (1992). An Introduction to Wavelets, Academic Press.

11. Chui, C. K. (ed.) (1992). Wavelets: A Tutorial in Theory and Applications, Academic Press.
12. Chui, C. K., Wang, J. Z. On compactly supported spline wavelets and a duality principle, Trans. Am. Math. Soc., to appear.
13. Ciesielski, Z. (1973). Constructive function theory and spline systems, Studia Math. 52, pp. 277–302.
14. Cohen, A. (1992). Biorthogonal wavelets. In: Chui, C. K. (ed.), Wavelets: A Tutorial in Theory and Applications, Academic Press, pp. 123–152.
15. Cohen, A. Daubechies, I. (1991). Non-separable bidimensional wavelet bases, Preprint AT&T Bell Laboratories, New Jersey.
16. Cohen, A., Daubechies, I., Feauveau, J. (1992). Bi-orthogonal bases of compactly supported wavelets, Comm. Pure and Appl. Math., 45, pp. 485–560.
17. Cohen, A., Daubechies, I., Jawerth, B., Vial, P. Multiresolution analysis, wavelets and fast algorithms on an interval, C. R. Académie des Sciences Paris., to appear.
18. Cohen, A., Schlenker, J.-M. (1991). Compactly supported bidimensional wavelet bases with hexagonal symmetry, AT&T Bell Laboratories, New Jersey, preprint.
19. Coifman, R. R. (1974). A real variable characterization of H^p. Studia Math, 51.
20. Combes, J.M., Grossmann, A., Tchamitchian, Ph. (eds.) (1989). Wavelets: Time-Frequency Methods and Phase Space, Inverse problems and theoretical imaging, Springer-Verlag.
21. Dahlke, S., Weinrich, I. (1991). Wavelet-Galerkin-methods: An adapted biorthogonal wavelet basis, Technical Report A-91-25, Freie Univ. Berlin.
22. Dahmen, W., Kunoth, A. (1992). Multilevel preconditioning, Numer. Math., 63 (2), pp. 315–344.
23. Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, Comm. Pure and Applied Math. 41, pp. 909–996.
24. Daubechies, I. (1992). Ten Lectures on Wavelets, Number 61 in CBMS–NSF Series in Applied Mathematics, SIAM Publications, Philadelphia.
25. Daubechies, I., Grossmann, A., Meyer, Y. (1986). Painless nonorthogonal expansions, J. Math. Phys. 27(5), pp. 1271–1283.
26. DeVore, R. A., Jawerth, B., Lucier, B. J. (1992). Image compression through wavelet transform coding, IEEE Trans. on Inf. Theory 38(2), pp. 719–746.
27. Fix, G., Strang, G. (1969). Fourier analysis of the finite element method in Ritz-Galerkin theory, Stud. Appl. Math 48, pp. 265–273.
28. Franklin, P. (1928). A set of continuous orthogonal functions, Math. Ann 100, pp. 522–529.
29. Frazier, M., Jawerth, B. (1985). Decomposition of Besov spaces, Indiana Univ. Math. J. 34(4), pp. 777–799.
30. Frazier, M., Jawerth, B. (1988). The ϕ-transform and applications to distribution spaces. In: Cwikel, M., et al. (eds.), Function Spaces and Applications, Lecture Notes in Mathematics 1302, Springer-Verlag, Berlin, pp. 223–246.
31. Frazier, M., Jawerth, B. (1990). A discrete transform and decompositions of distribution spaces, J. Func. Anal 93, pp. 34–170.
32. Frazier, M., Jawerth, B., Weiss, G. (1991). Littlewood-Paley theory and the study of function spaces, Regional Conference Series in Mathematics 79, American Mathematical Society, Providence.
33. Glowinski, R., Lawton, W., Ravechol, M., Tenenbaum, E. (1989). Wavelet solution of linear and nonlinear elliptic parabolic and hyperbolic problems in one space dimension, Technical Report AD 890527.1.1, Aware Inc.

34. Grossman, A., Morlet, J. (1984). Decompostion of Hardy functions into square integrable wavelets of constant shape, SIAM J. Math. Anal. 15(4), pp. 723–736.
35. Grossman, A., Morlet, J. (1985). Decomposition of functions into wavelets of constant shape, and related transforms. In: Streit, L. (ed.), Mathematics and Physics, Lectures on Recent Results, World Scientific, Singapore.
36. Grossman, A., Morlet, J., Paul, T. (1985). Transforms associated to square integrable group representations I. General results, J. Math. Phys. 26(10), pp. 2473–2479.
37. Haar, A. (1910). Zur Theorie der orthogonalen Funktionen-Systeme, Math. Ann. 69, pp. 331–371.
38. Heil, C. E., Walnut, D. F. (1989). Continuous and discrete wavelet transforms, SIAM Review 31(4), pp. 628–666.
39. Huntsberger, T., Jawerth, B., Lopresto, S., Tirumalai, A. Wavelets on closed sets and image processing, in preparation.
40. Jaffard, S., Laurencot, Ph. (1992). Orthonormal wavelets, analysis of operators, and applications to numerical analysis. In: Chui, C. K. (ed.), Wavelets: A Tutorial in Theory and Applications, Academic Press, pp. 543–602.
41. Jawerth, B. (1977). On Besov spaces, Technical Report 1977:1, Lund.
42. Jia, R.-Q., Micchelli, C. A. (1991). Using the refinement equations for the construction of pre-wavelets II: Powers of two. In: Laurent, P. J., Le Méhauté, A., Schumaker, L. L. (eds.), Curves and Surfaces. Academic Press, New York.
43. Koornwinder, T. H. (1993). The continuous wavelet transform. In: Koornwinder, T. H., (ed.), Wavelets: an elementary treatment of theory and applications, Series in Approximations and Decompositions, 1, World Scientific, Singapore.
44. Latto, A., Tenenbaum, E. (1990). Compactly supported wavelets and the numerical solution of Burgers' equation, Technical Report AD 900307, Aware Inc.
45. Lawton, W.M. (1991). Necessary and sufficient conditions for constructing orthonormal wavelets bases, J. Math. Phys. 32(1), pp. 57–61.
46. Lemarié, P.-G. (1988). Ondelettes a localisation exponentielle, J. de Math. Pures et Appl. 67(3), pp. 227–236.
47. Lemarié, P.-G. (ed.) (1990). Les Ondelettes en 1989, Lecture Notes in Mathematics 1438, Springer-Verlag, Berlin.
48. Lemarié, P.-G., Meyer, Y. (1986). Ondelettes et bases hilbertiennes, Rev. Mat. Iberoamericana 2, pp. 1–18.
49. Maday, Y., Perrier, V., Ravel, J.-C. (1991). Adaptivité dynamique sur bases d'ondelettes pour l'approximation d'équitations aux dérivées partielles, C. R. Académie des Sciences Paris, I(312), pp. 405–410.
50. Mallat, S., Hwang, W. L. (1992). Singularity detection and processing with wavelets, IEEE Trans. on Inf. Theory, 2, pp. 617–643.
51. Mallat, S., Zhong, S. (1992). Characterization of signals from multiscale edges, IEEE Trans. on Patt. Anal. and Mach. Intell. 14, pp. 710–732.
52. Mallat, S., Zhong, S. (1992). Wavelet transform maxima and multiscale edges. In: Ruskai, M. B., et. al. (eds.), Wavelets and Their Applications, Jones and Bartlett Publishers, pp. 67–104.
53. Mallat, S.G. (1989). Multifrequency channel decompositions of images and wavelet models, IEEE Trans. on Acoust., Speech Signal Process. 37(12), pp. 2091–2110.
54. Mallat, S.G. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$, Trans. Am. Math. Soc. 315(1), pp. 69–87.

55. Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation, IEEE Trans. on Patt. Anal. and Mach. Intell. 11(7), pp. 674–693.
56. Meyer, Y. (1990). Ondelettes et Opérateurs I. Ondelettes, Hermann, Paris.
57. Meyer, Y. (1991). Ondelettes sur l'intervalle, Technical Report 9020, CEREMADE, Université Paris–Dauphine, to appear in Rev. Mat. Iberoamer.
58. Micchelli, C. A. (1991). Using the refinement equations for the construction of pre-wavelets, Numerical Algorithms 1(1), pp. 75–116.
59. Micchelli, C. A., Rabut, C., Utretas, F. I. (1991). Using the refinement equations for the construction of pre-wavelets III: Elliptic splines, Numerical Algorithms 1(1), pp. 331–352.
60. Oswald, P. (1991). On discrete norm estimates related to multilevel preconditioners in the finite element method, preprint.
61. Peetre, J. (1976). New Thoughts on Besov Spaces. Duke U[illegible] Math. Series, Durham, N.C.
62. Riemenschneider, S. D., Shen, Z. (1992). Wavelets and pre-wavelets in low dimensions, J. Approx. Th. 71(1), pp. 18–38.
63. Ruskai, M.B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y., Raphael, L. (eds.) (1992). Wavelets and their Applications, Jones and Bartlett.
64. Stöckler, J. (1992). Multivariate wavelets. In: Chui, C. K. (ed.), Wavelets: A Tutorial in Theory and Applications, Academic Press, pp. 325–356.
65. Strang, G. (1989). Wavelets and dilation equations: A brief introduction, SIAM Review 31(4), pp. 614–627.
66. Strang, G., Fix, G. (1973). A Fourier analysis of the finite element variational method. In: Constructive aspects of Functional Analysis, Edizione Cremonese, Rome.
67. Strömberg, J. O. (1981). A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces. In: Beckner, et al. (eds.), Conference on Harmonic Analysis in Honor of Antoni Zygmund, volume II, Chicago, pp. 475–494,
68. Sweldens, W., Piessens, R. (1991). Asymptotic error expansions of wavelet approximations of smooth functions, Technical Report TW164, Department of Computer Science, K.U.Leuven.
69. Sweldens, W., Piessens, R. (1992). Calculation of the wavelet decomposition using quadrature formulae, CWI Quarterly 5(1), pp. 33–52.
70. Unser, M., Aldroubi, A., Eden, M. A family of polynomial spline wavelet transforms, NCRR report 153/90, to be published in Signal Processing.

Fractal Surfaces, Multiresolution Analyses and Wavelet Transforms

Jeffrey S. Geronimo[1], *Douglas P. Hardin*[2], *and Peter R. Massopust*[2]

[1] Georgia Institute of Technology, Atlanta, Georgia, USA
[2] Vanderbilt University, Nashville, Tennessee, USA

Abstract. Certain classes of multivariate vector-valued functions are constructed whose graphs are fractal sets. These functions generalize those introduced earlier by Barnsley and the authors, and they may provide a means of describing the many scale structure of real-world images. This class of functions is then used to generate a sequence of nested spaces as they occur in wavelet theory. Wavelet expansions, decomposition and reconstruction algorithms for the function representing the image are given.

Keywords: wavelets, fractal functions, fractal surfaces, multiresolution analysis, scaling function, decomposition and reconstruction algorithms, affine reflection group.

1 Introduction

Real-world images very often exhibit structures on many scales that are fractal-like. Mathematical representations of such images often involve piecewise continuous but not differentiable functions whose graphs are fractal sets in $\mathbb{R}^3$. It is therefore natural to use methods from fractal geometry in the mathematical representation of these images.

Recently, wavelets have been used to describe images containing both fine and coarse structures. This may be done by introducing a *multiresolution analysis on* $\mathbf{L}^2(\mathbb{R}^n)$ consisting of a sequence of nested function spaces that contain the representation of the image at various scales. These function spaces are generated by the translates and dilates of a single *scaling function* [5, 15, 16]. This method also provides fast decomposition and reconstruction algorithms.

In this paper certain classes of functions are constructed whose graphs are fractal sets that may be used in the mathematical representation of images. These functions are generalizations of those constructed in [1, 6, 9, 14]. In particular, these functions are used to generate a slightly modified multiresolution analysis on $\mathbf{L}^2(\mathbb{R}^n)$: the nested function spaces are generated by the translates and dilates of *several* scaling functions. This approach extends earlier results presented in [10].

The structure of the paper is as follows: The next section summarizes some facts from the theory of Coxeter groups and buildings necessary to understand the constructions in Sect. 4. In Sect. 3 the construction of fractal surfaces defined on simplicial and cubical domains is presented. Wavelet expansions based on fractal surfaces will be considered in Sect. 4. There a construction of a multiresolution analysis of $\mathbf{L}^2(\mathbb{R}^n)$ is presented using a finite set of scaling functions whose graphs are the surfaces described in the previous section.

2 Coxeter Groups

This section is a compendium of definitions and theorems from the theory of Coxeter groups, or more generally, the theory of buildings. Only definitions and results relevant to the set-up in this paper are presented. The interested reader is referred to the following, albeit incomplete, list of references: [2, 3, 4, 11, 17].

An *affine hyperplane* of $\mathbb{R}^n$ is a subset of the form $x + H_0$ with $x \in \mathbb{R}^n$ and H_0 a linear hyperplane of $\mathbb{R}^n$, i.e., H_0 is a codimension 1 linear subspace of $\mathbb{R}^n$. Affine hyperplanes may be defined via linear equations of the form $f = c$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a nonzero linear mapping and $c \in \mathbb{R}$. An *affine isometry* is a norm-preserving affine map. Let H_0 be a linear hyperplane. A linear transformation $r_{H_0} : \mathbb{R}^n \to \mathbb{R}^n$ is called a *reflection with respect to* H_0 iff $r_{H_0}|_{H_0} = id_{H_0}$ and $r_{H_0}(h^\perp) = (-1)h^\perp$, for all $h^\perp \in H_0^\perp$, the one-dimensional orthogonal complement of H_0. Now let H be an affine hyperplane and H_0 the linear hyperplane parallel to it. Then there exists an $x \in \mathbb{R}^n$ such that $H = x + H_0$. The reflection r_H with respect to H is defined by $r_H = T_x \circ r_{H_0} \circ T_{-x}$, where r_{H_0} is the orthogonal reflection with respect to H_0 and $T_x : \mathbb{R}^n \to \mathbb{R}^n$ is the translation $y \mapsto y + x$. Clearly, r_H is an (affine) isometry.

Suppose $\mathcal{H}$ is a collection of affine hyperplanes of $\mathbb{R}^n$. Denote by $\mathcal{W}$ the group of affine isometries generated by reflections r_H with $H \in \mathcal{H}$. A collection $\mathcal{H}$ is called $\mathcal{W}$-invariant iff $r_H\mathcal{H} = \mathcal{H}$, for all $r_H \in \mathcal{W}$. The group $\mathcal{W}$ is called an *affine reflection group* if there exists a $\mathcal{W}$-invariant family of hyperplanes $\mathcal{H}$ that is locally finite, that is, every point in $\mathbb{R}^n$ has a neighbourhood which intersects only finitely many $H \in \mathcal{H}$. A given collection $\mathcal{H}$ of hyperplanes partitions $\mathbb{R}^n$ into convex *cells*; these cells being defined by $f = c$, $f > c$, or $f < c$, where $f = c$ defines a hyperplane $H \in \mathcal{H}$, $c \in \mathbb{R}$. The *support* of a cell is the linear subspace L defined by the inequalities $f = 0$ that occur in its description. (If there are no such equalities, then $L := \mathbb{R}^n$.) The dimension of a cell is the dimension of its support. A *chamber* is a cell of maximal dimension n. The chambers are the connected components of the complement of $\mathcal{H}$ in $\mathbb{R}^n$. The *walls* of a chamber are the supports of the faces of codimension 1.

Now choose a chamber C and denote by R the set of all reflections with respect to the walls of C. Then the following facts hold true:

Theorem 1. *1. $\mathcal{W}$ is simply-transitive on the chambers.*
2. R generates $\mathcal{W}$.
3. $\mathcal{H}$ is the set of all hyperplanes $H \in \mathbb{R}^n$ with $r_H \in \mathcal{W}$.

4. *The closure $\overline{C}$ of C is a fundamental domain for the action of $\mathcal{W}$ on $\mathbb{R}^n$, i.e., no $r_H \in \mathcal{W}$ maps a point of $\overline{C}$ to another point of $\overline{C}$, and for all $x \in \mathbb{R}^n$ there exists an $r_H \in \mathcal{W}$ such that $r_H(x) \in \overline{C}$.*

The next theorem addresses some finiteness questions.

Theorem 2. *1. C has only finitely many walls, and thus R is finite.*
2. *There are only finitely many linear hyperplanes H_0 such that $\mathcal{H}$ contains a translate of H_0.*
3. *Denote by $\overline{\mathcal{W}}$ the set of linear parts of all elements in $\mathcal{W}$. Then $\overline{\mathcal{W}}$ is a finite reflection group.*

Remark. A finite reflection group, such as $\overline{\mathcal{W}}$, is called a *Coxeter group.*

The following theorem deals with the structure of C. But first a few more definitions: The group $\mathcal{W}$ is called *essential* iff its associated finite reflection group $\overline{\mathcal{W}}$ is essential, that is, if the origin is the only point fixed by all $r \in \overline{\mathcal{W}}$. A reflection group is *irreducible* iff it cannot be expressed as a product of reflection groups.

Theorem 3. *Suppose that $\mathcal{W}$ is essential and irreducible. Then the chamber C has exactly $n+1$ walls and is a simplex in $\mathbb{R}^n$. Furthermore, $\mathcal{W}$ is infinite.*

An essential irreducible infinite affine reflection group is called a *Euclidean reflection group* or a *Weyl group.*

Finally, consider the structure of $\mathcal{W}$. Since $\mathcal{W}$ acts on $\mathbb{R}^n$, the *stabilizer of* $x \in \mathbb{R}^n$ is the subgroup $\mathcal{W}_x$ of $\mathcal{W}$ given by $\{r_H \in \mathcal{W} : r_H x = x\}$.

Theorem 4. *1. There exist points $x \in \mathbb{R}^n$ such that the stabilizer $\mathcal{W}_x$ is isomorphic to $\overline{\mathcal{W}}$.*
2. *Let $\Gamma := \{x \in \mathbb{R}^n : T_x \in \mathcal{W}\}$. By an appropriate choice of the origin it may be assumed that $0 \in \Gamma$. Then Γ is a lattice in $\mathbb{R}^n$, that is, a subgroup of $\mathbb{R}^n$ of the form $\mathbb{Z}e_1 \oplus \cdots \oplus \mathbb{Z}e_n$, for some $\mathbb{R}$-basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$. Furthermore, $\mathcal{W}$ is isomorphic to the semi-direct product $\Gamma \bowtie \overline{\mathcal{W}}$.*

Remark. It follows from the second statement in Theorem 4 that the finite reflection group $\overline{\mathcal{W}}$ leaves the lattice Γ invariant. Such groups are called *crystallographic.* There is a classification of these crystallographic Coxeter groups in terms of so-called *Coxeter diagrams.* It can be shown that for each n there exists only a finite number of such groups.

3 Construction of Fractal Surfaces

In this section a class of fractal surfaces is introduced using methods from the theory of *iterated function systems.* In particular, the cases where the domains are n-simplices and n-cubes are considered.

Let $\mathcal{D}$ be a closed convex polyhedral region in $\mathbb{R}^n$ with non-empty interior made up of a finite number of affine copies of itself with non-overlapping interiors, that is, there exists a finite set of affine maps $u_i : \mathcal{D} \to \mathcal{D}$,

$$u_i(x) := A_i x + D_i \ , \tag{1}$$

where A_i is a nonsingular $n \times n$ matrix and $D_i \in \mathbb{R}^n$, $i = 1, \ldots, N$, such that $\mathcal{D} = \bigcup_{i=1}^N u_i(\mathcal{D})$ and $u_i(\mathcal{D}^\circ) \cap u_j(\mathcal{D}^\circ) = \emptyset$, $i \neq j$.

Denote by $\mathcal{V}$ the collection of all vertices of $\mathcal{D}$, and let $\mathcal{V}_i$ be the set of vertices of $u_i(\mathcal{D})$, $i = 1, \ldots, N$. Suppose there exists a function $\ell : \bigcup \mathcal{V}_i \to \mathcal{V}$, called a *labelling*, such that for all $i = 1, \ldots, N$, $u_i(\ell(v)) = v$, for all $v \in \mathcal{V}_i$. In this case, $(\mathcal{D}; u_i,\ i = 1, \ldots, N)$ is said to have *Property* (ℓ).

Let B be a nonsingular $m \times m$ matrix whose spectral radius s is less than 1. Throughout this paper the value s is fixed. Note that there exists a norm $\|\cdot\|_B$ on $\mathbb{R}^m$ such that the induced matrix norm of B equals s.

Let $\lambda_i \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$, $i = 1, \ldots, N$, and assume that whenever $u_i(\mathcal{D})$ and $u_j(\mathcal{D})$ have a common face E_{ij}, then $\lambda_i(x) = \lambda_j(x)$ for all $x \in u_i^{-1}(E_{ij}) = u_j^{-1}(E_{ij})$. This last assumption is referred to as *Property (m)*.

Define a norm $\|\cdot\|$ on $\mathbf{C}(\mathcal{D}, \mathbb{R}^m)$ by $\|f\| := sup\{\|f(x)\|_B : x \in \mathcal{D}\}$ and let $\Phi : \mathbf{C}(\mathcal{D}, \mathbb{R}^m) \to \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$ be defined as follows:

$$(\Phi f)(x) := \lambda_i(u_i^{-1}(x)) + Bf(u_i^{-1}(x)) \ , \tag{2}$$

for all $x \in u_i(\mathcal{D})$, $i = 1, \ldots, N$.

Theorem 5. *Suppose that* $(\mathcal{D}; u_i,\ i = 1, \ldots, N)$ *has Property* (ℓ) *and that* $\{\lambda_i : i = 1, \ldots, N\}$ *has Property (m). Then the mapping* Φ *given by (2) is well-defined and contractive with contractivity factor* s *in* $\|\cdot\|$. *Hence, it possesses a unique fixed point* $f^* \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$.

Proof. Suppose E_{ij} is a common face of $u_i(\mathcal{D})$ and $u_j(\mathcal{D})$. If v is one of the vertices of E_{ij} then $v \in \mathcal{V}_i \cap \mathcal{V}_j$. Note that by Property (ℓ) $\ell(v) = u_i^{-1}(v) = u_j^{-1}(v)$. Since each point of E_{ij} is a convex combination of the vertices of E_{ij}, $u_i^{-1}(x) = u_j^{-1}(x)$, for all $x \in E_{ij}$, and thus by Property (m) $\lambda_i(u_i^{-1}(x)) + Bf(u_i^{-1}(x)) = \lambda_j(u_j^{-1}(x)) + Bf(u_j^{-1}(x))$, for all $x \in u_i(\mathcal{D}) \cap u_j(\mathcal{D})$. Hence, for any $f \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$ $\Phi(f)$ is well-defined and continuous on $\mathcal{D}$.

Now let $f, g \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$. Then

$$\begin{aligned} \|\Phi(f) - \Phi(g)\| &= \sup\{B(f(u_i^{-1}(x)) - g(u_i^{-1}(x)))\|_B : x \in \mathcal{D}, i = 1, \ldots, N\} \\ &\leq s\|f - g\| \ . \end{aligned}$$

Therefore, by the Banach Fixed-Point Theorem, Φ has a unique fixed point $f^* \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$. □

The fixed point $f^* \in \mathbf{C}(\mathcal{D}, \mathbb{R}^m)$ is called a *fractal function* and its graph a *fractal surface* [6, 9, 14]. Note that the mapping

$$(\lambda_1, \ldots, \lambda_N) \overset{\Phi}{\mapsto} f^* \tag{3}$$

is linear.

Remark. The graph G of f^* is the *attractor* of the iterated function system $(\mathcal{D} \times \mathbb{R}^m, w_i : i = 1, \ldots, N)$, where $w_i : \mathcal{D} \times \mathbb{R}^m \to \mathcal{D} \times \mathbb{R}^m$ is defined by

$$w_i(x, y) := (u_i(x), \lambda_i(x) + By) \ ,$$

that is, $G = \bigcup_{i=1}^{N} w_i(G) =: \mathbf{w}(G)$. It can be shown that the iterates of any compact set K under the set-valued map $\mathbf{w}$ converge in the Hausdorff metric to G. By taking K to be a compact neighbourhood of G, it is easy to see that G is the inverse limit of the iterates of K under $\mathbf{w}$. Thus, from a shape-categorical point-of-view, all fractal surfaces have the same shape [18].

3.1 Simplicial Domains

Now suppose that $\mathcal{D}$ is an n-simplex, that is, $\mathcal{D} = \{x \in \mathbb{R}^n : x = \sum_{k=0}^{n} \alpha_k v_k, 0 \leq \alpha_k \leq 1, \sum_{k=0}^{n} \alpha_k = 1\}$, where $\{v_0, \ldots, v_n\}$ is a set of geometrically independent points in $\mathbb{R}^n$, with the property that $u_i(\mathcal{D})$ is similar to $\mathcal{D}$ and that $(\mathcal{D}; u_i, \ i = 1, \ldots, N)$ satisfies Property (ℓ). The existence of such simplices follows from well-known results in the theory of Coxeter groups (see Sect. 1 for more details.) Therefore, A_i has to be a similitude with scaling factor $a_i < 1$, $i = 1, \ldots, N$. Let

$$\lambda_i(x) := C_i x + E_i \ , \tag{4}$$

where C_i is an $m \times n$ matrix and $E_i \in \mathbb{R}^m$, $i = 1, \ldots, N$. The quantities C_i and E_i can be uniquely determined by prescribing values $z_v \in \mathbb{R}^m$ for f^* at each $v \in \bigcup \mathcal{V}_i$, that is,

$$z_v = \lambda_i(\ell(v)) + B z_{\ell(v)} \ , \tag{5}$$

for all $v \in \mathcal{V}_i$. Thus, if $u_i(\mathcal{D})$ and $u_j(\mathcal{D})$ have a common face E_{ij}, then $\lambda_i(v) = \lambda_j(v)$, for the vertices v of $u_i^{-1}(E_{ij}) = u_j^{-1}(E_{ij})$. Since the λ_is are affine Property (m) holds.

These arguments prove the following theorem.

Theorem 6. *Suppose that $\mathcal{D}$ is an n-simplex, $\{u_i : i = 1, \ldots, N\}$ consists of similitudes, and that $(\mathcal{D}; u_i, \ i = 1, \ldots, N)$ satisfies Property (ℓ). Assume that $\{z_v \in \mathbb{R}^m : v \in \bigcup \mathcal{V}_i\}$ is a given set of interpolation values. Let $\lambda_i : \mathcal{D} \to \mathbb{R}^m$ be the unique affine map satisfying (5), $i = 1, \ldots, N$. Then $\{\lambda_i : i = 1, \ldots, N\}$ satisfies Property (m), and the resulting continuous fractal surface interpolates $\{(v, z_v) : v \in \bigcup \mathcal{V}_i\}$.*

Let c denote the cardinality of $\bigcup \mathcal{V}_i$, and let $A(\mathbb{R}^n, \mathbb{R}^m)$ be the linear space of affine mappings from $\mathbb{R}^n$ into $\mathbb{R}^m$. It is easy to see that the mapping

$$\tau : (\mathbb{R}^m)^c \to (A(\mathbb{R}^n, \mathbb{R}^m))^N \ , \qquad z_v \mapsto (\lambda_1, \ldots, \lambda_N) \ , \tag{6}$$

is linear. The range of τ will be denoted by Λ_s. Define a linear function space $\mathcal{F}_s$ by setting $\mathcal{F}_s := \Psi(\Lambda_s)$. Note that $\mathcal{F}_s$ is exactly the collection of fractal surfaces in Theorem 6.

Example. Let $\mathcal{D} \subseteq \mathbb{R}^2$ be the triangle with vertex set $\{(0,0),(1,0),(0,1)\}$. Define maps u_i, $i = 1,2,3$, by

$$u_1(x,y) := (x/2,(y+1)/2) \ , \qquad u_2(x,y) := ((y+1)/2,x/2) \ ,$$
$$u_3(x,y) := (x/2,(1-y)/2) \ , \qquad u_4(x,y) := ((1-y)/2,x/2) \ ,$$

and the labelling ℓ by

$$\ell(1,0) = \ell(0,1) = \ell(0,0) := (0,1) \ ,$$

$$\ell(1/2,0) = \ell(0,1/2) := (0,0) \ , \quad \text{and} \quad \ell(1/2,1/2) := (1,0) \ .$$

Let $z_{(0,0)} := 1$, $z_{(1/2,0)} := 0.35$, $z_{(1,0)} := 0.25$, $z_{(0,1/2)} := -0.1$, $z_{(1/2,1/2)} := 0.3$,

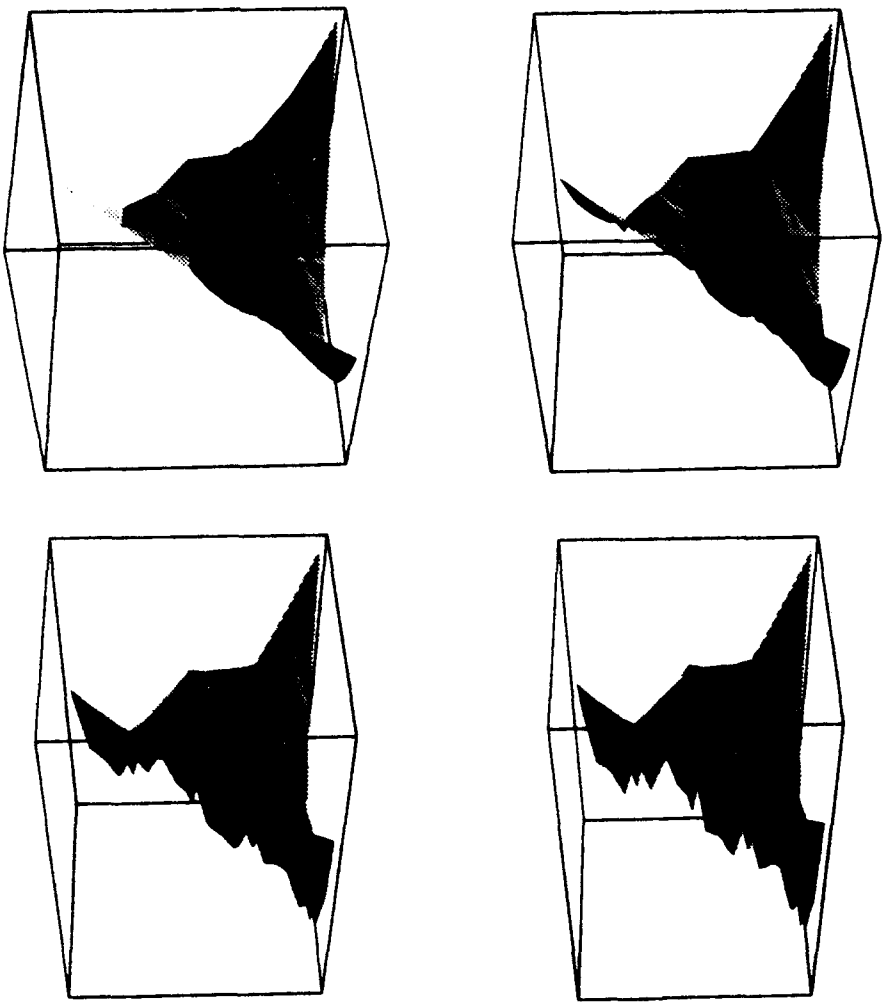


Fig. 1. Fractal surfaces defined on a triangle.

and $z_{(0,1)} := 0.45$. Figure 1 shows the fractal surfaces constructed for $s = 1/4 + (1/10)j$, $j = 0, 1, 3, 4$.

For the rest of this section B is assumed to be a similitude.

Next a notion of *oscillation* of a surface in $\mathbb{R}^{n+m}$ is introduced that gives a measure for the regularity of the surface. It can be shown that this oscillation is related to the fractal (box) dimension of the surface. Recall that the box dimension, dim_B, of a bounded set $S \subseteq \mathbb{R}^n$ is defined as

$$dim_B(S) := \lim_{\epsilon \to 0} \frac{\log \mathcal{N}_\epsilon(S)}{-\log \epsilon} \quad , \quad \text{provided this limit exists,} \tag{7}$$

where $\mathcal{N}_\epsilon(S)$ is the minimum number of ϵ-balls necessary to cover S.

In order to state the next theorem a definition is necessary.

Definition 7. 1. Let $0 < \epsilon < 1$ be given. An ϵ-cover $\mathcal{C}_\epsilon$ of a bounded set $S \subseteq \mathbb{R}^n$ is called admissible iff it is of the form

$$\mathcal{C}_\epsilon = \{B_\epsilon(r_\alpha) : \ |r_\alpha - r_\beta| \geq \frac{\epsilon}{2\sqrt{n}}, \text{ for all } r_\alpha \neq r_\beta\} \ ,$$

where $B_\epsilon(r_\alpha)$ denotes an n-dimensional ball of radius ϵ centreed at $r_\alpha \in S$.

2. Let $f : S \to \mathbb{R}^m$ be a function. The oscillation of f over $B \subseteq S$ is defined as

$$\omega(f; B) := \sup_{x, x' \in B} \|f(x) - f(x')\| \ ,$$

and the oscillation over S as

$$\Omega_\epsilon(f; S) := \inf \sum_{B \in \mathcal{C}_\epsilon} \omega(f; B) \ ,$$

where the infimum is taken over all admissible ϵ-covers $\mathcal{C}_\epsilon$ of S.

The following result concerning the oscillation of a fractal function $f : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and the box dimension of $graph(f^*)$ in the case $m = 1$ is proved in [9]. For the rather lengthy and involved proofs the reader is referred to the above-mentioned paper.

Theorem 8. *Suppose that*

1. *The set of interpolation points $\{(v_j, z_j) : j = 1, \ldots, n+1\}$ is not contained in any $(n+m-1)$-dimensional hyperplane of $\mathbb{R}^{n+m}$;*
2. $\sum_{i=1}^{N} s a_i^{n-1} > 1$.

Then there exist positive constants ϵ_0, k_1, and k_2 such that

$$k_1 \epsilon^{-\delta} \leq \Omega_\epsilon(f^*; \mathcal{D}) \leq k_2 \epsilon^{-\delta} \ , \tag{8}$$

for all $0 < \epsilon < \epsilon_0$, where δ is the unique positive solution of

$$\sum_{i=1}^{N} s a_i^{\delta} = 1 \ . \tag{9}$$

Furthermore, if $m = 1$ then the box dimension of $graph(f^*)$ is equal to $\delta + 1$.

3.2 Cubical Domains

Here $\mathcal{D}$ is the n-dimensional unit cube $[0,1]^n$. A construction of C^0 fractal surfaces on this domain is given.

Denote the vertices of $\mathcal{D}$ by $v := (v_1, \ldots, v_n)$, where each v_j is either 0 or 1. Let $1/a$ be an integer greater than 1. The set $\bigcup \mathcal{V}_i$ can be chosen to be as follows:

$$\bigcup \mathcal{V}_i = \{(ai_1, \ldots, ai_n) : i_k \in \{0, \ldots, 1/a\},\ k = 1, \ldots, n\} \ . \tag{10}$$

Then define the labelling ℓ by

$$\ell(ai_1, \ldots, ai_n) := (i_1, \ldots, i_n)\, mod(2) \ . \tag{11}$$

Now the maps u_i can be chosen in such a way that Property (ℓ) holds. The maps A_i are then of the form $A_i = a\Theta_i$, where Θ_i is a diagonal matrix whose diagonal elements are ± 1 and $a < 1$.

Let $\Pi_1([0,1])$ denote the set of all real polynomials $p : [0,1] \to \mathbb{R}^m$ of degree less than or equal to 1, and $\bigotimes_{k=1}^n \Pi_1([0,1])$ the n-fold tensor product of $\Pi_1([0,1])$. Choose

$$\lambda_i \in \bigotimes_{k=1}^{n} \Pi_1([0,1]) \ , \tag{12}$$

where $i = 1, \ldots, N$. It follows from the theory of multidimensional interpolation that, given a set of interpolation values $z_v \in \mathbb{R}^m$ at the vertices v of $\mathcal{D}$, there exists a unique interpolant in $\bigotimes_{k=1}^n \Pi_1([0,1])$ through $\{(v, z_v) : v \in \mathcal{V}\}$. Furthermore, if F is a face of $\mathcal{D}$ then $\lambda_i|_F \in \bigotimes_{k=1}^{n-1} \Pi_1([0,1])$ and is uniquely determined by $\{z_v : v \text{ is a vertex of } F\}$. Therefore, (5), with B being a nonsingular $m \times m$ matrix whose spectral radius s is less than 1, can be used to obtain the λ_is from $\{z_v : v \text{ is a vertex of } \mathcal{D}\}$. Clearly, $\{\lambda_i : i = 1, \ldots, N\}$ then satisfies Property (m). Hence the following result holds.

Theorem 9. *Let $\mathcal{D}$ be the cube $[0,1]^n$ and let $\{u_i : i = 1, \ldots, N\}$ be the set of similitudes determined by the labelling given in (11). Assume that $\{z_v \in \mathbb{R}^m : v \in \bigcup \mathcal{V}_i\}$ is a given set of interpolation values. Let $\lambda_i \in \bigotimes_{k=1}^n \Pi_1([0,1])$, $i = 1, \ldots, N$, be the unique function satisfying (5), $i = 1, \ldots, N$. Then $\{\lambda_i; i = 1, \ldots, N\}$ satisfies Property (m), and the resulting continuous fractal surface interpolates $\{(v, z_v) : v \in \bigcup \mathcal{V}_i\}$.*

As in the previous section one can define linear function spaces Λ_c and $\mathcal{F}_c$. Again, $\mathcal{F}_c$ consists precisely of the fractal surfaces in Theorem 9.

Example. Let $\mathcal{D} \subseteq \mathbb{R}^2$ be the square with vertices at $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$. Choose $1/a := 2$, and define a labelling ℓ as in (11). The maps $u_i : \mathcal{D} \to \mathcal{D}$ are then given by

$$u_1(x,y) = (x/2, y/2) \ , \qquad u_2(x,y) = (1 - x/2, y/2) \ ,$$
$$u_3(x,y) = (1 - x/2, 1 - y/2) \ , \qquad u_4(x,y) = (x/2, 1 - y/2) \ .$$

Choose $z_{(0,0)} := 0$, $z_{(1/2,0)} := 1/2$, $z_{(1,0)} := 3/5$, $z_{(0,1/2)} := 3/10$, $z_{(1/2,1/2)} := 3/4$, $z_{(0,1)} := 3/5$, $z_{(1,1/2)} := 3/10$, and $z_{(1,1)} := 7/10$ as interpolation values. Figure 2 shows the fractal surfaces constructed for $s = 1/4$ and $s = 3/5$, respectively.

If a grey-scale is associated with each z-value on the fractal surface, the following grey-scale images are obtained (see Fig. 3). Notice that the larger s-value introduces a more noticeable texture.

4 Wavelet Expansions Generated by Fractal Functions

As seen in the previous section, fractal surfaces can be used for interpolation purposes. Here it is shown that they may also be used to obtain nested function spaces forming a multiresolution analysis of $\mathbf{L}^2(\mathbb{R}^n)$ whose associated wavelets are fractal surfaces. For simplicity the case $m = 1$ is considered.

A *Multiresolution Analysis* (MRA for short) consists of a collection $\{V_\nu\}_{\nu\in\mathbb{Z}}$ of subspaces of $\mathbf{L}^2(\mathbb{R}^n)$ such that

1. $V_{\nu+1} \subset V_\nu$, for all $\nu \in \mathbb{Z}$;
2. $\bigcap_{\nu\in\mathbb{Z}} V_\nu = \{0\}$;
3. $cl_{\mathbf{L}^2(\mathbb{R}^n)}(\bigcup_{\nu\in\mathbb{Z}} V_\nu) = \mathbf{L}^2(\mathbb{R}^n)$;
4. There exists a finite number of *scaling functions* $\varphi^1, \ldots, \varphi^A \in V_0$ such that $\{\varphi^\alpha(\cdot - \ell) : \alpha = 1, \ldots, A;\ \ell \in \Gamma\}$, Γ a lattice in $\mathbb{R}^n$, is an orthonormal basis for V_0;
5. $f(x) \in V_\nu$ if and only if $f(\kappa x) \in V_{\nu-1}$, for some natural number κ greater than 1.

Remark. The above definition generalizes the usual notion of an MRA in that more than one scaling function is assumed.

Remark. The condition that $\{\varphi^\alpha(\cdot - \ell) : \alpha = 1, \ldots, A;\ \ell \in \Gamma\}$ is an orthonormal basis for V_0 is sometimes replaced by requiring that $\{\varphi^\alpha(\cdot - \ell) : \alpha = 1, \ldots, A;\ \ell \in \Gamma\}$ is only an *unconditional* or *Riesz basis* of V_0, i.e., there exist positive constants R_1 and R_2, called the *Riesz bounds*, such that

$$R_1 \sum_{\alpha=1}^{A} \sum_{\ell\in\Gamma} |c_\ell^\alpha|^2 \le \| \sum_{\alpha=1}^{A} \sum_{\ell\in\Gamma} c_\ell^\alpha \varphi^\alpha(\cdot - \ell)\|_2^2 \le R_2 \sum_{\alpha=1}^{A} \sum_{\ell\in\Gamma} |c_\ell^\alpha|^2 \ . \qquad (13)$$

Remark. MRAs can also be defined for function spaces other than $\mathbf{L}^2(\mathbb{R}^n)$ (see [6, 10]).

Let W_ν be the orthogonal complement of $V_{\nu+1}$ in V_ν. A collection of functions $\{\psi^1, \ldots, \psi^G\}$ is called a *set of wavelets associated with the MRA* if $\{\psi^\gamma(\cdot - \ell) : \gamma = 1, \ldots, G;\ \ell \in \Gamma\}$ is an orthonormal basis for W_0.

Let $\mathcal{D}$ denote a compact connected subset of $\mathbb{R}^n$. In order to use fractal surfaces defined on $\mathcal{D}$ as scaling functions for an MRA of $\mathbf{L}^2(\mathbb{R}^n)$, $\mathcal{D}$ has to be a *foldable* figure.

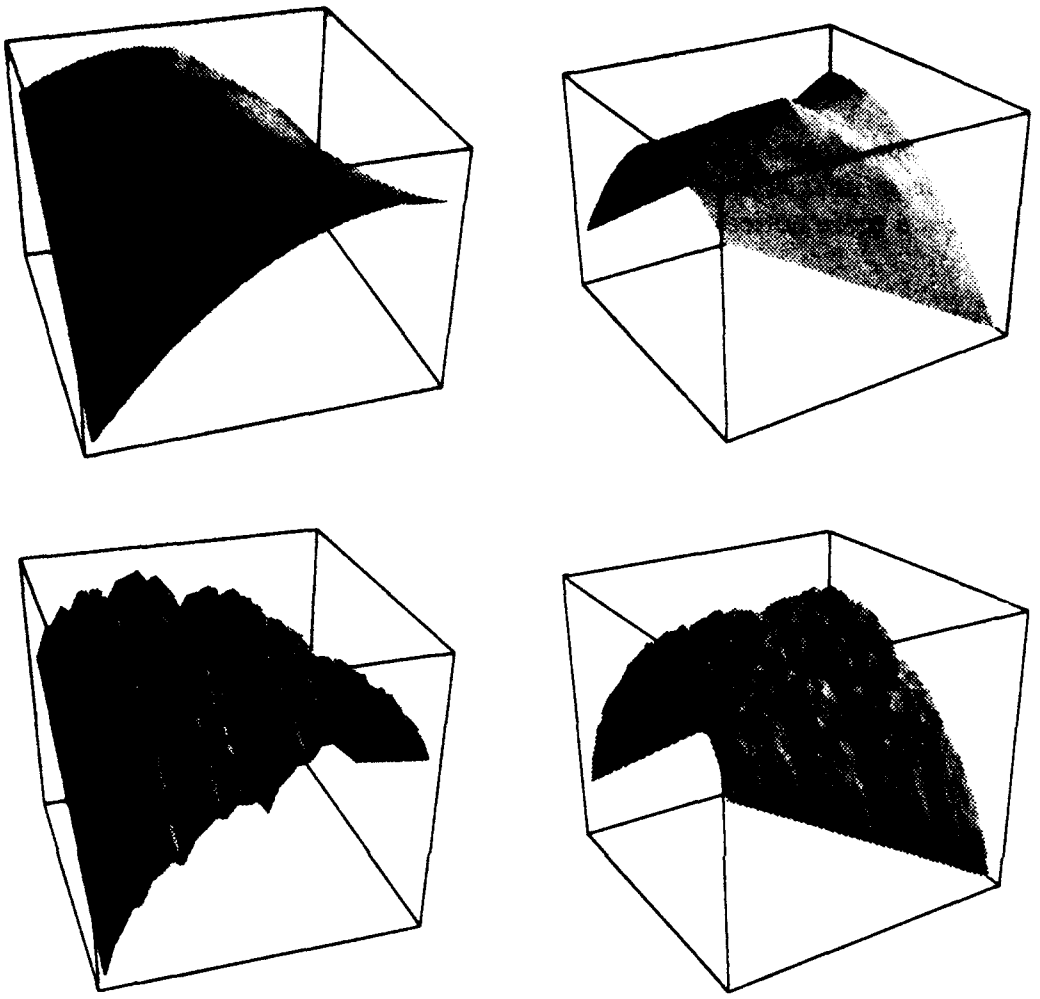


Fig. 2. Fractal surfaces defined on a square.

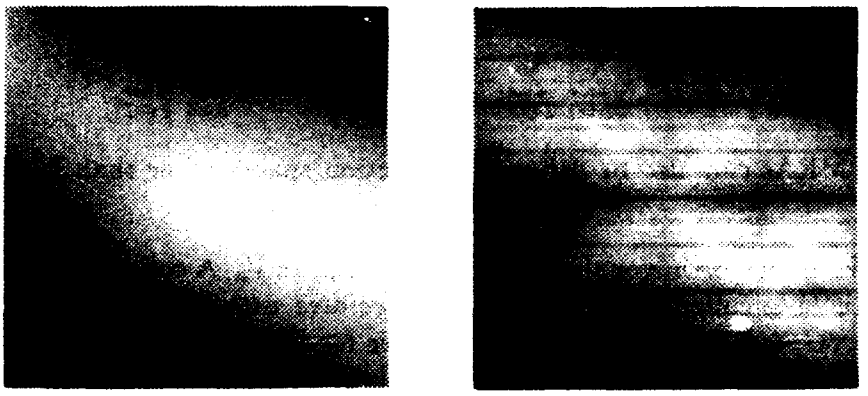


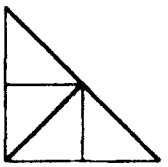
Fig. 3. Grey-scale images for the fractal surfaces constructed in Fig. 2.

Definition 10. A compact and connected subset F of $\mathbb{R}^n$ is called a foldable figure iff there exists a finite set of hyperplanes in $\mathbb{R}^n$ that cuts F into finitely many congruent subfigures $F_1, \ldots, F_M$ each similar to F, so that the reflection in any of these hyperplanes bounding F_m takes it into some $F_{m'}$.

The following result is proven in [12].

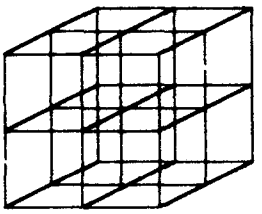
Theorem 11. *A foldable figure in $\mathbb{R}^n$ is a convex polytope that tessellates $\mathbb{R}^n$ by reflections in hyperplanes. Moreover, foldable figures are in one-to-one correspondence with crystallographic Coxeter groups.*

Example. A foldable figure in $\mathbb{R}^2$.



The following is example of a foldable figures in $\mathbb{R}^3$ whose associated (crystallographic) Coxeter groups is reducible.

Example. The three-dimensional unit cube $[0,1] \times [0,1] \times [0,1]$.



Let F be a foldable figure in $\mathbb{R}^n$. Let Σ be the tessellation and $\mathcal{H}$ be the set of hyperplanes associated with F, and let $\mathcal{W}$ be the affine reflection group generated by $\mathcal{H}$, i.e., $\mathcal{W}$ is the group of affine isometries in $\mathbb{R}^n$ generated by the reflections r_H for $H \in \mathcal{H}$. Then the following properties of $\mathcal{H}$ and $\mathcal{W}$ hold (see Sect. 1):

1. $\mathcal{H}$ consists of the translates of a finite set of linear hyperplanes.
2. $\mathcal{W}$ is simply-transitive on Σ, i.e., for any $\sigma, \sigma' \in \Sigma$ there exists a unique element $r_{\sigma'\sigma} \in \mathcal{W}$ mapping σ onto σ'.
3. For a proper choice of the origin, $\kappa\mathcal{H} \subseteq \mathcal{H}$ for any $\kappa \in \mathbb{N}$, where $\kappa\mathcal{H} := \{\kappa H : H \in \mathcal{H}\}$.

Now let $\mathcal{D} := \kappa F$. Clearly, $\mathcal{D}$ is also a foldable figure whose subfigures $\mathcal{D}_i$ are in Σ. Note that the tessellation and the set of hyperplanes associated with $\mathcal{D}$ are $\kappa\Sigma$ and $\kappa\mathcal{H}$, respectively. Furthermore, the affine reflection group generated by $\kappa\mathcal{H}$ is an isomorphic subgroup of $\mathcal{W}$. Using the fact that the group $\mathcal{W}$ is simply-transitive on Σ, one obtains a set of similitudes $u_i : \mathcal{D} \to \mathcal{D}_i$, $i = 1, \ldots, N$, $N = \kappa^n$, such that $(\mathcal{D}; u_i,\ i = 1, \ldots, N)$ satisfies Property (ℓ). More precisely,

let $\mathcal{D}_j$ be one of the subfigures. A bijection $\tilde{\ell}$ mapping the set $\mathcal{V}_j$ of vertices of $\mathcal{D}_j$ onto $\mathcal{V}$, can be extended to a mapping $\ell : \bigcup \mathcal{V}_j \to \mathcal{V}$ by setting

$$\ell(r_{\mathcal{D}_{j'}\mathcal{D}_j}(v)) := \tilde{\ell}(v) \ , \tag{14}$$

for $v \in \mathcal{V}_j$ and all $j' = 1, \ldots, N$, where $r_{\mathcal{D}_{j'}\mathcal{D}_j}$ is the unique element in $\mathcal{W}$ mapping $\mathcal{D}_j$ into $\mathcal{D}_{j'}$. Furthermore, $\tilde{\ell}$ defines a unique similitude $u_j : \mathcal{D} \to \mathcal{D}_j$ by

$$u_j(\tilde{\ell}(v)) := v \ , \quad \text{for all } v \in \mathcal{V}_j \ , \tag{15}$$

which can then be used to define $u_{j'} : \mathcal{D} \to \mathcal{D}_{j'}$, by

$$u_{j'} := r_{\mathcal{D}_{j'}\mathcal{D}_j} \circ u_j \ , \tag{16}$$

for all $j' \in \{1, \ldots, N\}$, $j' \neq j$. It is easy to see that u_i, $i = 1, \ldots, N$, is well-defined and, clearly, $(\mathcal{D}; u_i,\ i = 1, \ldots, N)$ satisfies Property (ℓ).

In what follows $\mathcal{D}$ denotes either an n-cube or a foldable n-simplex. Let

$$\Lambda := \begin{cases} \Lambda_c & \text{if } \mathcal{D} \text{ is an } n\text{-cube,} \\ \Lambda_s & \text{if } \mathcal{D} \text{ is a foldable } n\text{-simplex,} \end{cases}$$

and let

$$\mathcal{F} := \begin{cases} \mathcal{F}_c & \text{if } \mathcal{D} \text{ is an } n\text{-cube,} \\ \mathcal{F}_s & \text{if } \mathcal{D} \text{ is a foldable } n\text{-simplex.} \end{cases}$$

Define function spaces V_ν by

$$V_0 := \{f \in \mathbf{L}^2(\mathbb{R}^n) : \ f|_{\mathcal{D}'} \circ r_{\mathcal{D}'\mathcal{D}} \in \mathcal{F} \ , \text{ for all } \mathcal{D}' \in \kappa\Sigma\} \ , \tag{17}$$

and

$$V_\nu := \{f \in \mathbf{L}^2(\mathbb{R}^n) : \ f(\kappa^\nu \cdot) \in V_0\} \ , \text{ for } \nu \in \mathbb{Z},\ \nu \neq 0 \ . \tag{18}$$

Theorem 12. *The function spaces $\{V_\nu\}_{\nu\in\mathbb{Z}}$ have the following properties:*

1. *(Nestedness) $V_\nu \subset V_{\nu+1}$, for all $\nu \in \mathbb{Z}$;*
2. *(Separation) $\bigcap_{\nu\in\mathbb{Z}} V_\nu = \{0\}$;*
3. *(Density) $cl_{\mathbf{L}^2(\mathbb{R}^n)}(\bigcup_{\nu\in\mathbb{Z}} V_\nu) = \mathbf{L}^2(\mathbb{R}^n)$;*

Proof. Nestedness. Suppose $f \in V_1$, i.e., $f(\kappa\cdot)|_{\mathcal{D}'} \circ r_{\mathcal{D}'\mathcal{D}} \in \mathcal{F}$, for all $\mathcal{D}' \in \kappa\Sigma$. Therefore,

$$f(\kappa r_{\mathcal{D}'\mathcal{D}} \circ u_i(x)) = \lambda_i(x) + sf(\kappa r_{\mathcal{D}'\mathcal{D}}(x)) \ , \tag{19}$$

for $i = 1, \ldots, N$. For $\mathcal{D}' \in \kappa\Sigma$ and $j = 1, \ldots, N$, let $\mathcal{D}'_j := r_{\mathcal{D}'\mathcal{D}} \circ u_j(\mathcal{D})$. Note that for any $\mathcal{D}'' \in \kappa\Sigma$ there exists $\mathcal{D}' \in \kappa\Sigma$ and $j \in \{1, \ldots, N\}$ so that $\mathcal{D}'' = \kappa\mathcal{D}'_j$. It needs to be shown that $g := f|_{\kappa\mathcal{D}'_j} \circ r_{\kappa\mathcal{D}'_j\mathcal{D}} \in \mathcal{F}$. But (19) implies

$$\begin{aligned} g(u_i(x)) &= \lambda_j(u_i(x)) + sf(\kappa r_{\mathcal{D}'\mathcal{D}} \circ u_i(x)) \\ &= \lambda_j \circ u_i(x) + s(\lambda_i(x) + f(\kappa r_{\mathcal{D}'\mathcal{D}}(x)) \\ &= \lambda_j \circ u_i(x) + s(\lambda_i(x) - \lambda_j(x)) + sg(x) \ . \end{aligned}$$

Since A is invariant under the operation $(\lambda_1, \ldots, \lambda_N) \mapsto (\lambda_j \circ u_i - s(\lambda_i - \lambda_j))_{i=1}^N$, for all $j = 1, \ldots, N$, it follows that $g \in \mathcal{F}$.

Separation. Let $U_{\mathcal{D}} := \{f|_{\mathcal{D}} : f \in V_0\}$. Since $U_{\mathcal{D}}$ is finite-dimensional the norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are equivalent on $U_{\mathcal{D}}$. The translation invariance of V_0 implies that these norms are also equivalent on any $\mathcal{D}' \in \kappa\Sigma$. Hence

$$\|f\|_\infty \leq c\kappa^{-\nu/2}\|f\|_2 \ ,$$

for all $f \in V_\nu$. Thus, if $f \in \bigcap_{\nu \in \mathbb{Z}} V_\nu$, then $\|f\|_\infty = 0$.

Density. Due to the translation and dilation invariance of $\bigcup_{\nu \in \mathbb{Z}} V_\nu$ it suffices to show that $\chi_{\mathcal{D}} \in cl_{\mathbf{L}^2(\mathbb{R}^n)}(\bigcup_{\nu \in \mathbb{Z}} V_\nu)$. (Here $\chi_{\mathcal{D}}$ denotes the characteristic function of $\mathcal{D}$.) This, however, follows immediately from choosing $z_v := 1$, for all $v \in \bigcup \mathcal{V}_i$, and using the linearity of Ψ and τ (see (3) and (6), respectively). □

Next a basis for V_0 is introduced. Recall that $\mathcal{W} = \Gamma \bowtie \overline{\mathcal{W}}$ (see Sect. 1), where Γ is the lattice $\{x \in \mathbb{R}^n : T_x \in \mathcal{W}\}$ and $T_x : \mathbb{R}^n \to \mathbb{R}^n$ the translation $y \mapsto y + x$. Furthermore, recall that $\Gamma = \mathbb{Z}e_1 \oplus \cdots \oplus \mathbb{Z}e_n$, for some $\mathbb{R}$-basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$. Let $\{1, \ldots, A\}$ be an enumeration of the vertices of the smallest collection $\mathcal{C}$ of elements in $\kappa\Sigma$ containing $\mathcal{D}$ such that each element in the tessellation is an Γ-translate of some element in $\mathcal{C}$. Using Theorem 5, a fractal function $\varphi^\alpha \in \mathcal{F}$ satisfying

$$\varphi^\alpha(x) := \begin{cases} 1 & \text{for } x = v_\alpha \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

for all v_α, $\alpha = 1, \ldots, A$, can be constructed. The linearity of τ and Ψ implies that every $f \in \mathcal{F}$ is a linear combination of the functions in $\{\varphi_\alpha : \alpha = 1, \ldots, A\}$. It now follows that $\{\varphi^\alpha(\cdot - \ell) : \alpha \in \{1, \ldots, A\}, \ell \in \Gamma\}$ is a basis for V_0. The Gram-Schmidt Orthonormalization Algorithm yields then an orthonormal basis which—in order to ease notation— will also be denoted by $\{\varphi^\alpha(\cdot - \ell) : \alpha \in \{1, \ldots, A\}, \ell \in \Gamma\}$. This procedure requires the calculation of the inner product between two fractal functions. It is a well-known fact that such an inner product can be expressed in terms of the moments of the fractal functions, and that these moments can be calculated recursively, explicitly, and uniquely. (see [10] for the univariate case and [14] for the bivariate case. The general case is easy to obtain.)

Setting $\varphi^\alpha_{\nu,\ell} := \varphi^\alpha(\kappa^\nu \cdot - \ell)$, for all $\alpha \in \{1, \ldots, A\}$, $\ell \in \Gamma$, it is easy to see that the collection $\{\varphi^\alpha_{\nu,\ell} : \alpha \in \{1, \ldots, A\}, \ell \in \Gamma\}$ is an orthonormal basis for V_ν, $\nu \in \mathbb{Z}$. The functions in $\{\varphi^\alpha_{\nu,\ell} : \alpha \in \{1, \ldots, A\}, \ell \in \Gamma\}$ are obviously compactly supported. Theorem 12 now implies that the function spaces $\{V_\nu : \nu \in \mathbb{Z}\}$ form an MRA of $\mathbf{L}^2(\mathbb{R}^n)$ with compactly supported and orthonormal scaling functions $\{\varphi^\alpha_{\nu,\ell} : \alpha \in \{1, \ldots, A\}, \ell \in \Gamma\}$.

Next the construction of the wavelets associated with the above MRA is given. Denote by W_0 the orthogonal complement of V_0 in V_{-1}, i.e., $V_{-1} = V_0 \oplus W_0$. Suppose $\mathcal{D}' \in \kappa\mathcal{C}$, where $\mathcal{C}$ is as above. Note that $\{1, \ldots, A\}$ is a labelling of the vertices of $\kappa\mathcal{C}$. Since $\mathcal{D}'$ consists of κ^n subfigures there exist $\kappa^n A$ scaling functions $\varphi^1_{1,\ell}, \ldots, \varphi^{\kappa^n A}_{1,\ell}$ forming an orthonormal basis for $\mathbf{L}^2(\mathbb{R}^n)$ -functions defined on

the subfigures of the elements in $\kappa\mathcal{C}$. Hence the wavelets $\{\psi^1, \ldots, \psi^{(\kappa^n-1)A}\}$ can be defined by

$$\psi^\gamma_{0,\ell} := \varphi^\gamma_{-1,\ell} \ , \tag{21}$$

for all $\gamma \in G := \{1, \ldots, \kappa^n A\} - \{1, \ldots, A\}$. Clearly, $\{\psi^\gamma_{0,\ell} : \gamma \in G, \ell \in \Gamma\}$ is an orthonormal basis for W_0 whose elements have compact support. Thus, if W_ν is such that $V_\nu \oplus W_\nu = V_{\nu-1}$, the κ^ν-dilates and Γ-translates of $\{\psi^\gamma_{0,\ell} : \gamma \in G, \ell \in \Gamma\}$, $\{\psi^\gamma_{\nu,\ell} : \gamma \in G, \ell \in \Gamma\}$, form an orthonormal compactly supported basis of W_ν, $\nu \in \mathbb{Z}$. The function spaces $\{W_\nu\}_{\nu\in\mathbb{Z}}$ are usually called the *wavelet spaces*.

Theorem 13. *The function spaces defined by (17) and (18) form an MRA of* $\mathbf{L}^2(\mathbb{R}^n)$ *with orthonormal, compactly supported scaling functions. Furthermore, there exists a finite set of compactly supported and orthonormal wavelets that are fractal functions in* $\mathcal{F}$.

Theorem 13 provides a means of decomposing $\mathbf{L}^2(\mathbb{R}^n)$ into subspaces. More precisely, let $\{W_\nu : \nu \in \mathbb{Z}\}$ be the wavelet spaces. Since by construction $W_\nu \cap W_{\nu'} = \emptyset$ and $W_\nu \perp W_{\nu'}$, $\nu \neq \nu'$, the following orthogonal direct sum decomposition of $\mathbf{L}^2(\mathbb{R}^n)$ is obtained:

$$\mathbf{L}^2(\mathbb{R}^n) = \bigoplus_{\nu\in\mathbb{Z}} W_\nu \ . \tag{22}$$

Hence every function $f \in \mathbf{L}^2(\mathbb{R}^n)$ has a unique representation as a *wavelet series* in the form

$$f(x) = \sum_{\gamma\in G}\sum_{\nu\in\mathbb{Z}}\sum_{\ell\in\Gamma} c^\gamma_{\nu,\ell}\psi^\gamma_{\nu,\ell}(x) \ , \tag{23}$$

where the sum is understood in the $\mathbf{L}^2$ sense. A more compact representation is obtained by using vector notation. Let $\mathbf{c}_{\nu,\ell} := (c^1_{\nu,\ell}, \ldots, c^G_{\nu,\ell})^t$, where t denotes the transpose, and let $\boldsymbol{\psi}_{\nu,\ell} := (\psi^1_{\nu,\ell}, \ldots, \psi^G_{\nu,\ell})^t$. Then (23) can be expressed as

$$f(x) = \sum_{\nu\in\mathbb{Z}}\sum_{\ell\in\Gamma} \mathbf{c}^t_{\nu,\ell}\boldsymbol{\psi}_{\nu,\ell} \ . \tag{24}$$

The coefficient vector $\mathbf{c}_{\nu,\ell}$ is given by

$$\mathbf{c}_{\nu,\ell} = (\langle f, \psi^1_{\nu,\ell}\rangle, \ldots, \langle\psi^G_{\nu,\ell}\rangle)^t \ . \tag{25}$$

(Here $\langle\cdot,\cdot\rangle$ denotes the $\mathbf{L}^2$ inner product on $\mathbb{R}^n$.) Let $\psi_{a,b} := \psi(\frac{\cdot - b}{a})$, for $a \in \mathbb{R} - \{0\}$ and $b \in \mathbb{R}^n$, and define the *(integral) wavelet transform* W_ψ on $\mathbf{L}^2(\mathbb{R}^n)$ by

$$(W_\psi f)(a,b) := (|a|^{-n/2}\int_{\mathbb{R}^n} f(x)\psi^1_{a,b}(x)\,dx, \ldots, |a|^{-n/2}\int_{\mathbb{R}^n} f(x)\psi^G_{a,b}(x)\,dx)^t \ . \tag{26}$$

The wavelet transform can be used to write the vector coefficients in (25) as

$$\mathbf{c}_{\nu,\ell} = (W_\psi f)(\frac{1}{\kappa^\nu}, \ell) \ . \tag{27}$$

Since V_1 and W_1 are subspaces of V_0, there exists a sequence of $A \times A$ matrices $\{\mathbf{p}_\ell\}_{\ell \in \Gamma}$ and a sequence of $G \times G$ matrices $\{\mathbf{q}_\ell\}_{\ell \in \Gamma}$ such that $\varphi := (\varphi^1, \ldots, \varphi^A)^t$ and $\psi := (\psi^1, \ldots, \psi^G)^t$ satisfies the following two-scale matrix dilation equations:

$$\varphi(x) = \sum_{\ell \in \Gamma} \mathbf{p}_\ell \varphi(\kappa x - \ell) \ , \tag{28}$$

and

$$\psi(x) = \sum_{\ell \in \Gamma} \mathbf{q}_\ell \varphi(\kappa x - \ell) \ . \tag{29}$$

Since both φ and ψ have compact support, all but a finite number of $\mathbf{p}_\ell$s and $\mathbf{q}_\ell$s are equal to the zero matrix.

Finally, let us have a brief look at the decomposition and reconstruction algorithm associated with the MRA introduced above.

Suppose a function $f_0 \in V_0$ is given. Since $V_0 = V_1 \oplus W_1$, there exist functions $f_1 \in V_1$ and $g_1 \in W_1$ such that the following unique decomposition of f_0 holds:

$$f_0 = f_1 + g_1 \ . \tag{30}$$

Conversely, given functions $f_1 \in V_1$ and $g_1 \in W_1$, one can reconstruct f_0. Since $f_0 \in V_0$ there exists a sequence of vectors $\mathbf{c}(0) := \{c_\ell(0) : \ell \in \mathbb{Z}\} \in (\ell^2(\Gamma))^A$ such that

$$f_0(x) = \sum_{\ell \in \Gamma} c_\ell^t(0) \varphi(x - \ell) \ . \tag{31}$$

In a similar fashion, since $f_1 \in V_1$ and $g_1 \in W_1$, there exist two sequences $\mathbf{c}(1) := \{c_\ell(1) : \ell \in \mathbb{Z}\} \in (\ell^2(\Gamma))^A$ and $\mathbf{d}(1) := \{d_\ell(1) : \ell \in \mathbb{Z}\} \in (\ell^2(\Gamma))^G$ such that

$$f_1(x) = \sum_{\ell \in \Gamma} c_\ell^t(1) \varphi(x/\kappa - \ell) \ , \tag{32}$$

and

$$g_1(x) = \sum_{\ell \in \Gamma} d_\ell^t(1) \psi(x/\kappa - \ell) \ . \tag{33}$$

It is important to note that $\mathbf{d}(1) = (W_\psi f)(\kappa, \ell)$. Using (28) and (29), the following reconstruction algorithm is obtained:

$$c_\ell(0) = \sum_{\ell' \in \Gamma} c_{\ell'}^t(1) \mathbf{p}_{\ell - \kappa\ell'} + d_{\ell'}^t(1) \mathbf{q}_{\ell - \kappa\ell'} \ , \tag{34}$$

for all $\ell \in \mathbb{Z}$. Due to the compact support of φ and ψ this algorithm is finite.

Since $\varphi(\kappa \cdot - \ell) \in V_{-1} = V_0 \oplus W_0$ the following decomposition relation holds:

$$\varphi(\kappa x - \ell) = \sum_{\ell' \in \Gamma} [\mathbf{a}_{\ell - \kappa\ell'} \varphi(x - \ell') + \mathbf{b}_{\ell - \kappa\ell'} \psi(x - \ell')] \ , \tag{35}$$

for some $A \times A$ matrices $\{\mathbf{a}_\ell : \ell \in \mathbb{Z}\}$ and some $G \times G$ matrices $\{\mathbf{b}_\ell : \ell \in \mathbb{Z}\}$. Now given that $f_1(x) = \sum_{\ell \in \Gamma} c_\ell^t(1) \varphi(x/\kappa - \ell)$ one has—after some straight-forward algebra—

$$f_1(x) = \sum_{\ell' \in \Gamma} [\sum_{\ell \in \Gamma} \mathbf{a}_{\ell - \kappa\ell'} c_\ell(0)] \varphi(x - \ell') + \sum_{\ell' \in \Gamma} [\sum_{\ell \in \Gamma} \mathbf{b}_{\ell - \kappa\ell'} c_\ell(0)] \psi(x - \ell') \ . \tag{36}$$

Hence,

$$c_{\ell'}(1) = \sum_{\ell \in \Gamma} a_{\ell - n\ell'} c_\ell(0) \ , \quad \text{and} \quad d_{\ell'}(1) = \sum_{\ell \in \Gamma} b_{\ell - n\ell'} c_\ell(0) \ . \tag{37}$$

Again, this algorithm is finite.

References

1. Barnsley, M.F. (1986). Fractal functions and interpolation, Constr. Approx. 2, pp. 303–329.
2. Bourbaki, N. (1968). Groupes et Algèbres de Lie, Chapitres IV, V, VI, Hermann, Paris.
3. Brown, K. S. (1989). Buildings, Springer-Verlag, New York.
4. Coxeter, H. S. M. (1973). Regular Polytopes, 3rd. ed., Dover, New York.
5. Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, Comm. Pure and Applied Math. XLI, pp. 909–996.
6. Geronimo, J. S., Hardin, D. P. (1993). Fractal interpolation surfaces, J. Math. Anal. and Appl., to appear.
7. Geronimo, J. S., Hardin, D. P., Massopust, P.R. (1993). Fractal functions and wavelet expansions based on several scaling functions, J. of Approximation Theory, to appear.
8. Geronimo, J. S., Hardin, D. P., Massopust, P.R. (1992). An application of Coxeter groups to the construction of wavelet bases in $\mathbb{R}^n$, Contemporary Aspects of Fourier Analysis, Marcel Dekker, to appear.
9. Hardin, D. P., Massopust, P.R. (1993). Fractal interpolation functions from $\mathbb{R}^n \to \mathbb{R}^m$ and their projections, submitted.
10. Hardin, D. P., Kessler, B., Massopust, P.R. (1992). Multiresolution analyses based on fractal functions, J. Approx. Theory 71 (1), pp. 104–120.
11. Hiller, H. (1982). Geometry of Coxeter groups, Pitman, Boston.
12. Hoffman, M., Withers, W.D. (1988). Generalized Chebyshev polynomials associated with affine Weyl groups, Trans. Am. Math. Soc. 308 (1), pp. 91–104.
13. Massopust, P.R. (1990). Fractal surfaces, J. Math. Anal. and Appl. 5 (1), pp. 275–290.
14. Massopust, P.R. (1993). Smooth interpolating curves and surfaces generated by iterated function systems, Zeitschrift für Analysis und ihre Anwendungen, July issue.
15. Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R}^2)$, Trans. Am. Math. Soc. 315, pp. 69–87.
16. Meyer, I. (1990). Ondelettes et Opérateurs, Hermann, Paris.
17. Ronan, R. (1992). Buildings: Main ideas and applications, I. Main ideas, Bull. London Math. Soc. 24, pp. 1–51.
18. Segal, J. (1993). Shape theory: An ANR-Sequence approach, this volume, pp. 111–125.

Interpolation in Multiscale Representations

Charles H. Anderson[1] *and Subrata Rakshit*[2]

[1] Department of Anatomy and Neurobiology, Washington Univ. School of Medicine, St. Louis MO 63110, USA
[2] EE Division, California Institute of Technology, Pasadena, CA 91125, USA

Abstract. The decomposition of images into multiscale representations provides a fundamental starting point for the analysis of grey-scale images. This paper examines some practical implementation issues centered on the traditional tradeoff between computational complexity and storage. Critically sampled wavelet transforms provide compact, complete, orthonormal representations, but interpolation is computationally expensive. Oversampled, non-orthogonal pyramid representations provide computationally simpler interpolation at a modest cost of increased storage, leading to more efficient image analysis systems.

Keywords: pyramids, wavelets, basis functions, interpolation, Lie operators.

1 Introduction

Multiscale representations of images have been known to be important for image analysis for some time [8, 10]. The Burt Laplacian pyramid opened the door for efficient implementation of these concepts [3, 4]. More recently, the discrete subband decompositions that are critically sampled, such as wavelets, have become very popular [1, 9, 11]. These have proved highly efficient in encoding images because they are compact representations which match the scale-invariant structure of most images. Their proponents suggest their mathematically clean properties of orthogonality make them superior to the earlier pyramid representations, which are overcomplete and non-orthogonal [9]. But, in practice, these representations have not been able to replace the filter-based techniques in many areas of image analysis. The primary reason for this turns out to be the classical trade-off between computational complexity and storage. Overcompleteness provides more local interpolation formulae in space, scale, and orientation.

The shortcomings of the critically sampled multiscale representations have been detailed in a paper by Simoncelli *et al.* [12]. They show that, while the basis functions of the wavelets are related to one another by translation, dilation, and in some cases rotation, the coefficients do not show simple invariance along these dimensions. To resolve this problem they have formalized the concept of

"shiftability", which provides the functionality of local interpolation within one of these dimensions without utilizing the parameters in the others. Achieving this computational compactness carries the cost of overcompleteness.

2 Interpolation of Uniformly Sampled Data

This paper first explores the interaction between interpolation complexity, sample density and precision for the simple case of uniform sampling in one dimension. Assume a band-limited continuous function $F(x)$ uniformly sampled at the Nyquist frequency. The function at location a can be reconstructed using the familiar form

$$F(a) = \sum_n F_n \text{sinc}(k_s(a - n\Delta)/2.0) \ , \tag{1}$$

where

$\Delta = 2\pi/k_s$ is the sample spacing,
$k_s = 2k_m$ is twice the maximum frequency of $F(x)$,
$F_n = F(n\Delta)$ is the sample values.

Equation (1) can be interpreted in three ways: as an interpolation formula, a complete orthonormal basis representation where the F_n are the amplitudes, or the sinc() as an analog postfilter operation that recreates the original continuous function from the samples.

Practical issues are: (1) how many terms in the sum are required to achieve a desired level of precision; (2) are there better interpolating formulae; (3) how does oversampling help? An examination of the root mean square (rms) error, in reconstructing a sinewave, $F(x) = \sin(kx)$, using a variety of standard interpolating functions shows that performance becomes very bad as the frequency approaches the Nyquist limit (Fig. 1). Simple linear interpolation does quite well with an rms error less than 5%, up to $k = k_s/8$. A cubic spline pushes a similar level of performance to $k = k_s/4$. Finally, for a given precision the width of the interpolation function must be increased dramatically as the Nyquist limit is approached, which seems to imply that high-frequency information is encoded non-locally.

This behaviour can be formalized by considering (1) for a frequency near the Nyquist limit, $k = k_s/2 - \epsilon$. The sample values in this case are values alternating in sign with an envelope at the beat frequency between the input and half the sampling frequency,

$$F_n = \sin(nk\Delta) = \sin(n\Delta(k_s/2 - \epsilon)) = -\cos(n\pi)\sin(n\Delta\epsilon) \ , \tag{2}$$

as shown in Fig. 2.

A stringent test is the reconstruction of this function at the point $a = 0.5\Delta$, where the value should be close to 1.0. However the sample values in the region around 0.5 become vanishingly small as the beat frequency ϵ approaches zero.

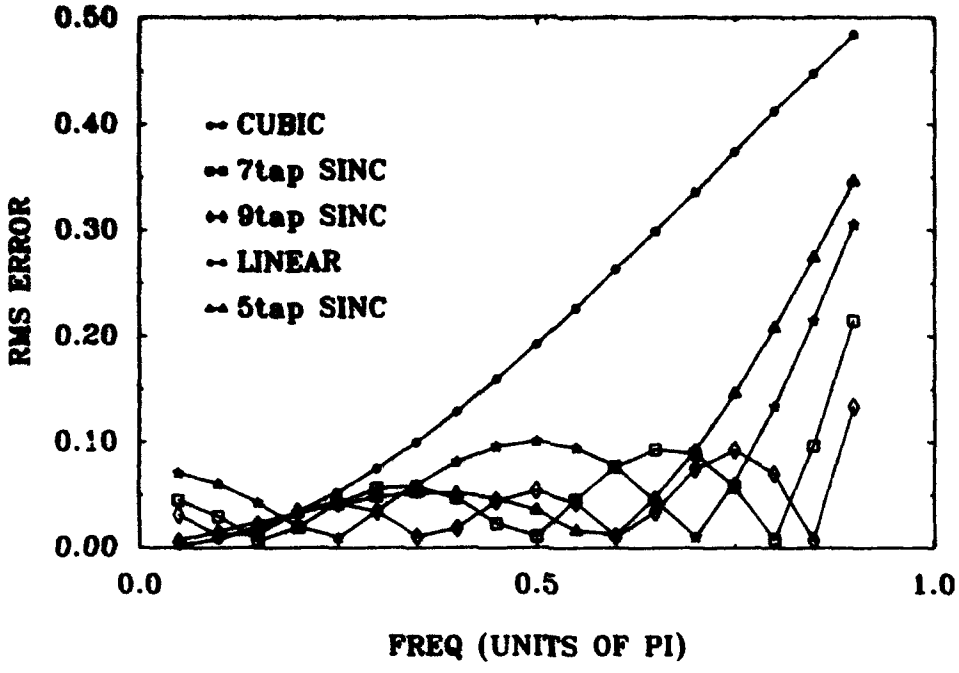


Fig. 1. Root mean square errors for different interpolation formulae.

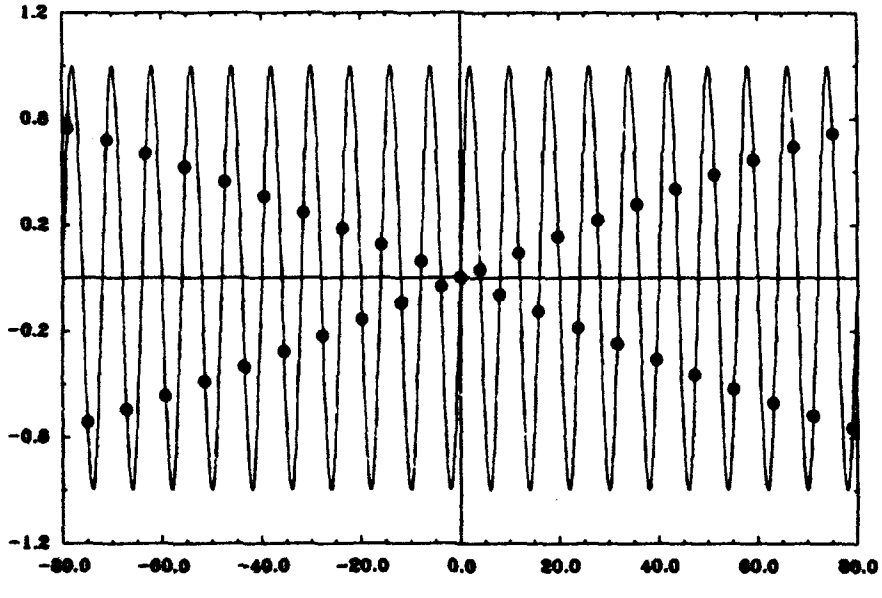


Fig. 2. Sampling close to the Nyquist limit.

Equation (1) gives

$$F(0.5\Delta) = \sum_n F_n \sin(\pi(0.5-n))/\pi(0.5-n)$$
$$= -\sum_n \cos^2(n\pi)\sin(n\epsilon\Delta)/(\pi(0.5-n)) \ . \quad (3)$$

The factor of 0.5π in the denominator can be ignored since most of the contributions to the sum comes from large n.

$$F(0.5\Delta) \approx \epsilon\Delta/\pi \sum_n \sin(n\epsilon\Delta)/(n\epsilon\Delta) \ . \quad (4)$$

The $\epsilon\Delta$ times the sum is approximately equal to the integral of the sinc() function between $-\infty$ to $+\infty$, whose value is π. The major contribution to the integral lies in the interval $[-\pi, \pi]$, hence the range of summation must be of the order of $\pi/(\epsilon\Delta) = k_s/(k_s - 2k)$ to give a reasonable estimate of $F(0.5\Delta)$. Note that this diverges as k_s approaches $2k$ and provides an analytic basis for the behaviour observed in Fig. 1.

Consider the case of re-interpolating a sinewave sampled over a fixed interval L with Lk_s/π sample values. Using the result given above for the width of the interpolation formula, the total number of operations scales as

$$N \approx Lk_s(k_s/(k_s - 2k)) \ . \quad (5)$$

This diverges at $k_s = 2k$ and $k_s = \infty$, and has a minimum at $k_s = 4k$, or twice the Nyquist rate. Oversampling clearly pays in this case.

3 Oversampling in Laplacian Pyramids

How to carry out local interpolation using information in multiresolution representations is an important issue. Consider the Laplacian pyramid which separates information into bands whose width increases by a constant multiple factor, usually 2. The encoding of the information at the edges of the subbands is a problem. The previous discussion should make it clear that hard, square boundaries between bands remove all hope for locality of interpolating formulae. Instead the band edges should extend into the neighbouring bands as shown in Fig. 3, which raises the problem of encoding these extended bands.

A simple approach would be to sample each band at a rate where aliasing of the information is eliminated and interpolation can be done locally. In general this is difficult because the convolution kernels of the filters generally lose the property of translational invariance in the sense that the kernel values become dependent on absolute spatial location. Oversampling by a factor of 2 is a case which does not have this difficulty and, as discussed above, there are reasons to believe this is a desirable sampling rate. The following discussion reviews two simple Laplacian pyramid techniques and extensions that allow them to achieve double sampling.

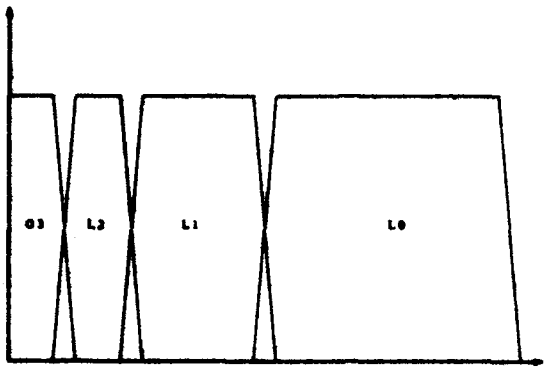


Fig. 8. Spatial frequencies of Laplacian pyramid subbands.

The original Burt Laplacian pyramid is generated by the following recursive rules, starting with the original image defined as G_0:

$$G_{n+1} = \text{Reduce } G_n,$$
$$L_n = G_n - \text{Expand } G_{n+1}.$$

The Reduce operation is defined as a low-pass filter followed by Decimation or subsampling by removing every other row and column. The Expand operation creates an enlarged image by inserting zeros at the points removed by the Reduce operation, multiplying the retained values by 4, and then smoothing the results with a low-pass filter operation. The Burt Laplacian pyramid, composed of the band-pass L_n, plus a residue G_N, is a formal wavelet-like transform with an exact reconstruction rule that is non-orthogonal and oversampled by a factor of 4/3.

A variant, the Burt double density pyramid [5], provides a practical method for over-sampling the bands by a factor of 2 along each spatial dimension. This procedure begins with creating G_1 by low-pass filtering the G_0, but not subsampling. The next levels of a Gaussian pyramid are then created by convolving with a spread-tap filter, followed by subsampling in the usual fashion. A spread-tap filter is created by interleaving the coefficients in the original low-pass filter with zeros, which produces a low-pass response with half the bandwidth of the original, plus a band-pass response in the region where the low-pass images have minimal response. Laplacian or band-pass components can be generated from the Gaussian levels by subtracting them from one another before the decimation process. All the levels are sampled at twice the linear density of a normal Laplacian pyramid except for the first level. The overall computational cost of forming a double density pyramid is the same per coefficient as constructing a normal pyramid, so the overall increase in cost in both computation and storage is (1+ 4/3)/(4/3)=7/4.

The *Filter Subtract Decimate* [2], FSD pyramid provides an algorithm for locally computing twice oversampled coefficients. The FSD pyramid is generated

by the recursive rules:

$$\begin{aligned} \tilde{G}_{n+1} &= H * G_n && \text{Low-pass } \textit{Filter}, \\ L_n &= G_n - \tilde{G}_{n+1} && \textit{Subtract}, \\ G_{n+1} &= \text{Decimate } \tilde{G}_{n+1} && \textit{Decimate}. \end{aligned}$$

This has the advantage over the original Burt Laplacian pyramid of not having aliased information introduced into the middle of the Laplacian bands. Aliasing does however appear at the high spectral end of the Laplacian components in both of these pyramids.

Perfect reconstruction of the FSD pyramid is generally not possible because the Expand interpolation process always removes some high frequency information. However, in this case the information lost in expanding G_{n+1} can be recovered to a good approximation from low-pass filtering L_n;

$$\begin{aligned} G_n &= L_n + \tilde{G}_{n+1} \\ &\approx L_n + H * L_n + \text{Expand } G_{n+1}, \end{aligned} \tag{6}$$

where the low-pass filter in the Expand operation is the same as the H used in the formation of the FSD pyramid. Using this observation, it is possible to create a doubly sampled Laplacian from an FSD pyramid using

$$\tilde{L}_n = H * L_{n-1} + \text{Expand } L_n. \tag{7}$$

This operation can be done locally, which is advantageous since most real-time pyramid applications operate on windowed regions of selected levels of the Laplacian pyramids.

4 Performance Comparisons

A hierarchical database of small templates, less than 16 pixels in width and height, of selected features of an object on multiple scales is an efficient and robust way to encode complex objects [6]. This allows coarse to fine searches where rough outlines are first located and then precise identification is carried out by focusing on those key features that uniquely define the object. Such a procedure depends on a method for comparing the stored templates with the selected windowed areas of a Laplacian pyramid. The comparison however must allow for small changes in the scale, rotation and other warpings between the two image patches being compared. A standard means for doing this is the least squares correlator

$$\mathbf{E} = \sum_{i,j} \left[L[i,j] - T[i,j] - \sum_n p_n \mathbf{D}_n T[i,j] \right]^2 , \tag{8}$$

where the i, j sum is over the entire Laplacian patch L, which is taken to be the same size as the template T. The p_n are distortion parameters for the corresponding Lie group operators $\mathbf{D}_n$. The p_n which minimize $\mathbf{E}$ provide estimates

for their values and the value of E at the minimum provides a measure of how close L and T are to one another.

The Lie operator for simple translations is the spatial derivative, where an estimate of the amount of a simple translation in the absence of other distortions is given by

$$\delta x = \frac{\sum_{i,j}(L-T)dT/dx}{\sum_{i,j}(dT/dx)^2} \quad . \tag{9}$$

This equation was used to compare the performance of the various Laplacian pyramids discussed above where the derivative was estimated using the simple equation

$$dT[i,j]/dx = (T[i+1,j] - T[i-1,j])/2 \quad . \tag{10}$$

This is equivalent to the assumption that the Laplacian values can be found by linear interpolation, an assumption that the first result of this paper suggests is very suspect if the functions have been critically sampled.

The test was carried out by creating a floating point image of random numbers 1027 x 1025 pixels in size followed by a low-pass filter using the separable 5-tap filter given in Table 2. Three 1025 x 1025 images, A, B and C, were extracted at the points (0,0), (1,0), and (2,0) respectively and Laplacian pyramids generated from them. The comparison was then computed at various pyramid levels using

$$\delta[n,m] = \frac{\sum_{i,j}(L_C[n+i,m+j] - L_A[n+i,m+j])dL_B[n+i,m+j]/dx}{\sum_{i,j}(dL_B[n+i,m+j]/dx)^2} \quad , \tag{11}$$

where the sum was taken over a 7 x 7 local window around each n, m point. The mean and standard deviation of the estimated displacement, as normalized to the known values, are summarized in Table 1 for the pyramids discussed above using two low-pass filters. The first filter (Table 2) is a separable, efficient and

Table 1. Displacement estimates

Pyramid	FSD5	Burt5	FSD7	FSDDbl5	BurtDbl5
Simple dx	1.60(0.17)	1.52(0.14)	1.40(0.10)	1.16(0.08)	1.19(0.06)
Filter dx	1.10(0.10)	1.02(0.10)	1.08(0.06)	..	..

compact 5-tap filter that minimizes aliasing in the pyramids. The non-separable 7-tap filter (Table 3) is designed to provide 7 bits of precision in reconstructing the FSD pyramid and to give strong low-frequency rejection in the Laplacian bands to compensate for the $1/f$ spatial-frequency structure of images. There is a specialized pyramid chip [13] that utilizes the 5-tap filter, and the 7-tap design can be utilized on commercially available convolvers designed for image processing.

Table 2. Separable 5-tap filter

1/16	1/4	3/8	1/4	1/16

Table 3. 7-tap filter, normalisation factor = 1024

0	5	13	16	13	5	0
5	25	52	64	52	25	5
13	52	102	126	102	52	13
16	64	126	156	126	64	16
13	52	102	126	102	52	13
5	25	52	64	52	25	5
0	5	13	16	13	5	0

A second set of tests were run using a 9 x 9 filter for the derivative in the FSD5, Burt5 and FSD7 pyramids. The derivative was found using the following procedure. The basis functions in these discrete representations can be generated by inserting the value 1.0 in the centre of the lowest band-pass level, say L_5, and setting all the other coefficients in the representation to zero and then reconstructing until the function is defined with sufficient samples, G_0 in this case. The resulting basis function can be then be differentiated by taking finite differences. The construction of a Laplacian pyramid on this sampled version of the derivative will create a representation for the derivative in the pyramid. However this will be spread across several pyramid levels, which increases the computational load. Instead the result was subsampled at the density of L_5 and then band-pass filtered to create an approximation that is restricted to a single level. These kernels are given in Tables 4, 5, and 6.

The first thing to note in Table 1 is that all cases using the simple form of the derivative provide estimates of the displacement that are too high; 50% in the case of the simple pyramids. This is because finite differencing always

Table 4. FSD5 dx kernel

.000	.000	.001	.001	.000	-.001	-.001	.000	.000
.001	.004	.007	.007	.000	-.007	-.007	-.004	-.001
.004	.012	.021	.019	.000	-.019	-.021	-.012	-.004
.007	.022	.039	-.132	.000	.132	-.039	-.022	-.007
.008	.027	.048	-.625	.000	.625	-.048	-.027	-.008
.007	.022	.039	-.132	.000	.132	-.039	-.022	-.007
.004	.012	.021	.019	.000	-.019	-.021	-.012	-.004
.001	.004	.007	.007	.000	-.007	-.007	-.004	-.001
.000	.000	.001	.001	.000	-.001	-.001	.000	.000

Table 5. Burt5 dx kernel

.000	.001	.002	.001	.000	-.001	-.002	-.001	.000
.001	.005	.008	.005	.000	-.005	-.008	-.005	-.001
.003	.013	.020	.013	.000	-.013	-.020	0.013	-.003
.007	.026	.039	-.140	.000	.140	-.039	-.026	-.007
.008	.034	.051	-.633	.000	.633	-.051	-.034	-.008
.007	.026	.039	-.140	.000	.140	-.039	-.026	-.007
.003	.013	.020	.013	.000	-.013	-.020	0.013	-.003
.001	.005	.008	.005	.000	-.005	-.008	-.005	-.001
.000	.001	.002	.001	.000	-.001	-.002	-.001	.000

Table 6. FSD7 dx kernel

.004	.006	.007	.005	.000	-.005	-.007	-.006	-.004
.008	.014	.016	.011	.000	-.011	-.016	-.014	-.008
.015	.024	.025	-.009	.000	.009	-.009	-.025	-.015
.020	.033	-.025	-.327	.000	.327	.025	-.033	-.020
.023	.036	-.101	-.678	.000	.678	.101	-.036	-.023
.020	.033	-.025	-.327	.000	.327	.025	-.033	-.020
.015	.024	.025	-.009	.000	.009	-.009	-.025	-.015
.008	.014	.016	.011	.000	-.011	-.016	-.014	-.008
.004	.006	.007	.005	.000	-.005	-.007	-.006	-.004

underestimates the derivative for oscillatory functions such as the Laplacian bands. The double density pyramids are much better than the single density ones as expected. The second line in Table 1 shows that the derivative kernels provide better absolute estimates of shifts. In practice it has been found that a good working estimate for the translation parameters can be found by using the simple estimate (10) for the derivative and then rescaling the results by the values given in the first line of Table 1.

5 Summary

Interpolation and the closely associated microscopic distortions described by Lie groups are fundamental to robust image analysis in multiscale representations. Critically sampled representations have computationally expensive interpolating formulae and when total system costs are computed using the total number of operations then oversampled representations can be less expensive. The simple, computationally efficient, Laplacian pyramids can be easily extended to double density representations where simple linear interpolating works quite well.

References

1. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M. (1984). Pyramid methods in image processing, RCA Engineer 29, pp. 33–41.
2. Anderson, C.H. (1990). Pyramids in machine vision at JPL, Proc. First International Symposium on Measurement and Control in Robotics, pp. F1.3.1–F1.3.4, Houston, Texas.
3. Burt, P.J. (1983). Fast algorithms for estimating local image properties, Computer Graphics and Image Processing 21, pp. 368–382.
4. Burt, P.J., Adelson, E.H. (1983). The laplacian pyramid as a compact image code, IEEE Trans. on Comm. COM 31, pp. 532–540.
5. Burt, P.J. (1985). Private Communication.
6. Burt, P.J. (1988). Smart sensing within a pyramid vision machine, Proc. of the IEEE 76, No. 8.
7. Daubechies, I. (1990). The wavelet transform, time-frequency localisation and signal analysis, IEEE Trans. Information Theory 36, pp. 961–1005.
8. Koenderink, J.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
9. Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation, IEEE Patt. Anal. Machine Intell. 11, pp. 674–693.
10. Marr, D. (1982). Vision: A computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, San Francisco.
11. Simoncelli, E.P., Adelson, E.H. (1991). Subband Transforms. In: Woods, J.W. (ed.), Subband Image Coding, Kluwer Academic Publishers, Norwell, Mass., pp. 143–192.
12. Simoncelli, E P., Freeman, W.T., Adelson, E.H., Heeger, D.J. (1992). Shiftable multi-scale transforms, IEEE Trans. Information Theory, 2(38):587–607.
13. Van der Wal, G.S. (1991). The Sarnoff pyramid chip, Proc. Computer Architecture for Machine Perception (CAMP-91), Paris, pp. 69–79.

Discrete Stochastic Growth Models for Two-Dimensional Shapes *

Scott Thompson and Azriel Rosenfeld

Center for Automation Research, University of Maryland, College Park, MD 20742, USA

Abstract. Discrete models for growth of a shape from a point on a two-dimensional Cartesian grid are described. By *growth* is meant an accretionary process occurring at the boundary of the shape. Three types of growth models are discussed: deterministic (periodic), probabilistic (stochastic), and probabilistic mixing of deterministic processes. Each type is defined and illustrated with examples. It is shown that probabilistically mixing deterministic processes can produce smooth isotropic or elongated regions, concavities, and protrusions. The paper emphasizes empirical results; analytical studies are in progress.

Keywords: shape, shape models, growth processes.

1 Introduction

Examples of deterministic and stochastic parallel growth models have appeared in the literature on cellular automata (CA), mathematical morphology (MM), L-systems, and fractals.

Cellular automata (e.g. [12]) are described by cell-states and transition functions. A transition function assigns a new state to a cell based on the cell's current state and possibly the states of neighbouring cells. In general, the transition function may lead to finite, infinite, periodic, or chaotic growth phenomena. Research on such phenomena has emphasized their ability to generate complex patterns (e.g. [9]). For a famous example, see Conway's "Game of Life" [1].

In mathematical morphology ([10]), structuring elements are used to perform dilations and erosions of shapes or patterns. Combining dilation and erosion is a powerful technique for shape analysis (e.g., skeletonization).

In L-systems and fractals [6, 7, 11], single cells are recursively replaced by patterns of cells. Such processes produce self-similar scale-invariant patterns.

While CA, MM, L-systems, and fractal processes have been shown to be useful in many applications, they afford a degree of generality unnecessary to model the growth of compact shapes. Little or no work in these areas has dealt

* The support of the Air Force Office of Scientific Research under Grant F49620-93-1-0039 is gratefully acknowledged.

with processes for generating simple shapes. An example is [3]; but it used a sequential, rather than parallel, growth process.

In this paper, three simple deterministic growth processes are described that produce compact regions, elongated regions, and concavities. A procedure called *probabilistic mixing* is also introduced, which allows composition of deterministic processes. These models are special cases of CA and MM, but they are purely dilational; and, unlike L-systems and fractals, they produce simple "solid" shapes.

2 Definitions

A (digital) *shape* is a non-empty, finite set of grid points $S = \{P : P \in \mathbb{Z}^2\}$. The grid points belonging to S will be called *cells*. The *background* (complement) of a shape S is the set of grid points $\bar{S} = \{P : P \notin S\}$.

For $P, Q \in \mathbb{Z}^2$, the *City Block* and *Chessboard* metrics are respectively:

$$d_4(P,Q) = |x_1 - x_2| + |y_1 - y_2| \ ,$$

$$d_8(P,Q) = max(|x_1 - x_2|, |y_1 - y_2|) \ ,$$

where $P = (x_1, y_1)$ and $Q = (x_2, y_2)$. If $d_i(P,Q) = 1$ (i=4 or 8), P and Q are called *i-adjacent* or *i-neighbours*. The reflexive, transitive closure of i-adjacency is called *i-connectedness*; in other words, P and Q are called *i-connected* ($i = 4$ or 8) if there exists a sequence of grid points $P = P_0, P_1, \cdots, P_n = Q$, such that P_k and P_{k+1} are i-adjacent, $0 \leq k < n$.

The *boundary* S' of a shape S is the set of cells of S that are 8-adjacent to points of the background:

$$S' = \{P : P \in S,\ d_8(P,Q) = 1 \text{ for some } Q \in \bar{S}\} \ .$$

Similarly, the *coboundary* $\bar{S}'$ of S is the set of grid points of $\bar{S}$ that are 8-adjacent to points of S:

$$\bar{S}' = \{P : P \in \bar{S},\ d_8(P,Q) = 1 \text{ for some } Q \in S\} \ .$$

A *growth process* applied to a shape adds to that shape some (or all) of the points in its coboundary. In other words, if $\mathcal{M}$ is a growth process and S is a shape, then applying $\mathcal{M}$ to S yields a set of cells $\mathcal{M}(S)$ such that

$$S \subseteq \mathcal{M}(S) \text{ and } \mathcal{M}(S) - S \subseteq \bar{S}' \ .$$

Since the points in $\bar{S}'$ are 8-adjacent to S, the result of applying a growth process to an 8-connected shape is still 8-connected. In particular, shapes grown from a single cell by iteratively applying a growth process are 8-connected.

The remainder of this paper is organized into two main sections, dealing respectively with isotropic growth and nonisotropic growth. In each of these sections, three types of growth models are considered: deterministic, probabilistic, and probabilistic mixing of deterministic processes.

3 Isotropic Growth

Isotropic growth implies that the rate of growth should be the same in all directions. The experiments reported in this section attempt to simulate isotropic growth on a Cartesian two-dimensional grid. S is approximately a disc if

$$S = \{P \in \mathbb{Z}^2 : d_e(O, P) \leq r,\ r \in \mathbb{Z}\} \ ,$$

where $d_e(P,Q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, $P = (x_1, y_1)$, $Q = (x_2, y_2)$, and $O = (0,0)$. The goal is to define a growth process capable of growing shapes that are approximately discs, starting from a single point O at the origin.

3.1 Deterministic 4/8 Mixing

The simplest growth processes on the grid are the processes that repeatedly add to a shape all the points that are i-adjacent to it ($i = 4$ or 8):

$$D_4(S) = S \cup \{P : d_4(P,Q) = 1 \text{ for some } Q \in S\} \ ,$$

$$D_8(S) = S \cup \{P : d_8(P,Q) = 1 \text{ for some } Q \in S\} \ ,$$

where S is any shape. D_4 causes growth in the horizontal and vertical directions, and D_8 also causes growth in the diagonal directions. It is easy to see that D_4 alone produces diamonds, and D_8 produces squares. Combinations of D_4 and D_8 produce octagons whose sides have slopes that are multiples of 45°. It can be shown that $D_4^s(D_8^t(O))$ (t D_8–steps followed by s D_4–steps) yields an octagon with vertices

$$\{(\pm(s+t), t), (\pm(s+t), -t), (\pm t, (s+t)), (\pm t, -(s+t))\} \ ,$$

and that the order in which the D_4 and D_8 processes are applied makes no difference.

Since the length of the horizontal and vertical sides is $2t$, and the length of the diagonal sides is $\sqrt{2}s$, the octagon can be made arbitrarily close to regular by letting s/t approach $\sqrt{2}$ (Fig. 1). Note that this near-regular octagon has a circumscribed circle of radius $s+t$. It has been shown that octagons are the best "discs" which can be obtained using only the D_4 and D_8 operations [8].

3.2 Probabilistic Growth

Much rounder but slightly ragged discs are obtained by making the growth process probabilistic. The probabilistic growth model associates a probability of creating a new cell with each 8-neighbour of an existing cell. This structure is called an *8-neighbour probability kernel*:

$$\begin{matrix} p_3 & p_2 & p_1 \\ p_4 & \square & p_0 \\ p_5 & p_6 & p_7 \end{matrix}$$

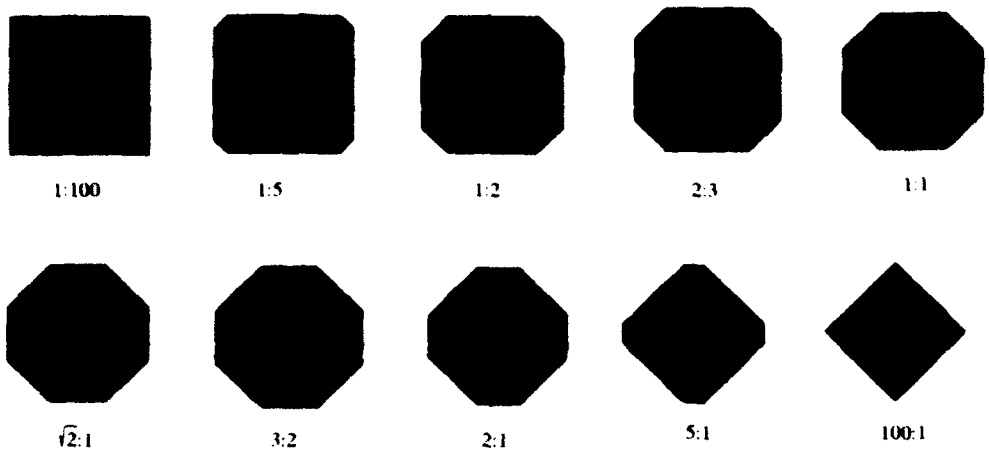


Fig. 1. $D_4^s(D_8^t(O))$ best approximates a disc when $s/t \approx \sqrt{2}$.

This growth process iterates over a sequence of discrete time steps; at each time step the jth neighbour of a cell becomes a cell with probability p_j. If a point is a neighbour of more than one cell, the probabilities combine independently. Setting all the p_js to 1 produces a square, but lowering the diagonal probabilities produces growth that is more nearly isotropic. The corresponding kernel is defined to be

$$\begin{array}{ccc} p & 1 & p \\ 1 & \square & 1 \\ p & 1 & p \end{array}$$

Experimental results for this growth model are shown in Fig. 2. Setting $p = 0.2$ gives the most disc-like shape. (For the measure of circularity, the ratio of mean radius (μ_r) to standard deviation of radius (σ_r), proposed in [4], is used. The higher the ratio, the more circular the shape.)

3.3 Probabilistic 4/8 Mixing

Ragged discs can also be obtained by probabilistically mixing the D_4 and D_8 growth processes, independently for each cell. Recall that in deterministic growth all cells on the boundary of S are treated the same,

$$D_i(S) = S \cup \bigcup_{P \in S'} D_i(P) \ ,$$

and each iteration involves applying either D_4 or D_8 to all cells on the boundary. In probabilistic mixing, the cells are treated independently; at each iteration, a separate choice between D_4 and D_8 is made for each boundary cell. Subscripts (4 or 8) are randomly assigned to the cells of S'; let S'_4 and S'_8 denote the sets

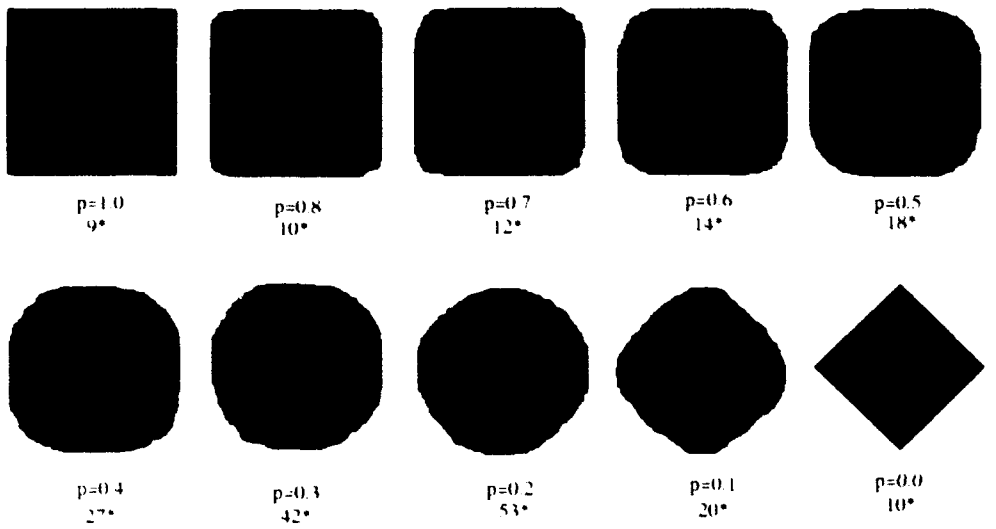


Fig. 2. Probabilistic growth with horizontal and vertical probability 1 and diagonal probability p. The starred numbers are values of a circularity measure [4].

of cells that receive subscripts 4 and 8 respectively. Then a probabilistic mix of D_4 and D_8 applied to S is

$$\mathcal{PM}(D_4, D_8; S) = S \cup \bigcup_{P \in S'_4} D_4(P) \cup \bigcup_{P \in S'_8} D_8(P) \ .$$

When 4 is assigned with probability p and 8 with probability $1 - p$, for various values of $p/(1 - p)$, the results are as shown in Fig. 3. Apparently, the most disc-like shapes result when

$$p = \frac{2\sqrt{2}}{1 + 2\sqrt{2}} \ , \quad \text{that is,} \quad \frac{p}{(1-p)} = 2\sqrt{2} \ .$$

3.4 Smoothed Growth

Evidently, probabilistic models allow, with low probability, "arbitrary" growth at the boundary: this is why the boundary may become ragged and "hairy", as illustrated in the first part of Fig. 4 (the process that generates this figure will be described in Sect. 4.4). The raggedness occurs because each cell generates new cells without knowledge of its neighbours' offspring. By requiring "local support" for new cells, the degree of boundary jaggedness can be controlled. (This notion of "local support" is not new. If we think of cells as requiring bonds with neighbouring cells, the notion of local support (i.e., requiring more than one bond) is consistent with biological embryology [5]. Other growth models in the past have imposed similar constraints on the addition of new cells [3].) One way to achieve this is to require that a new cell must have at least k pre-existing

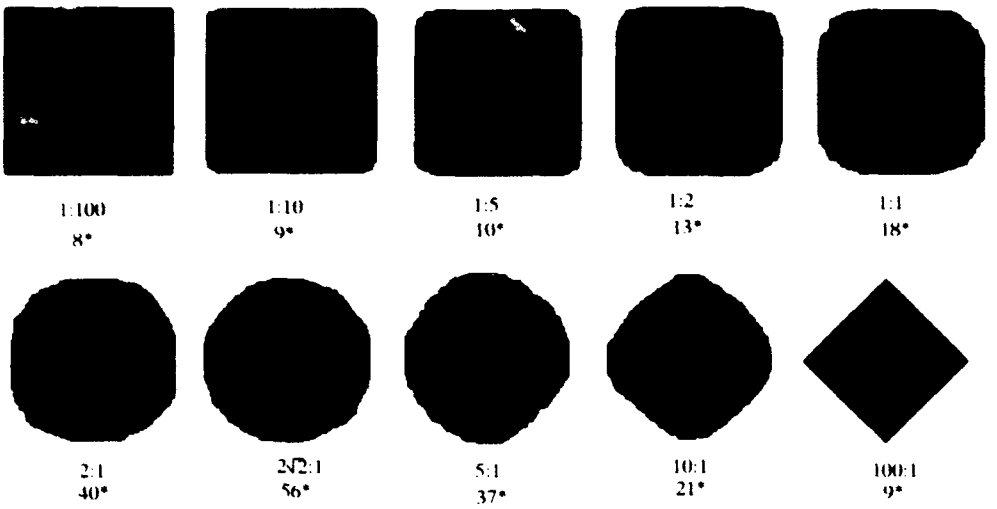


Fig. 3. Probabilistic mixing of D_4 and D_8 growth processes.

cells as neighbours after some time t, where clearly $t > 1$ since growth begins with a single cell. Figure 4 shows the results of requiring k bonds during the last five iterations of the growth process, for $k = 1, 2, 3, 4$ and 5.

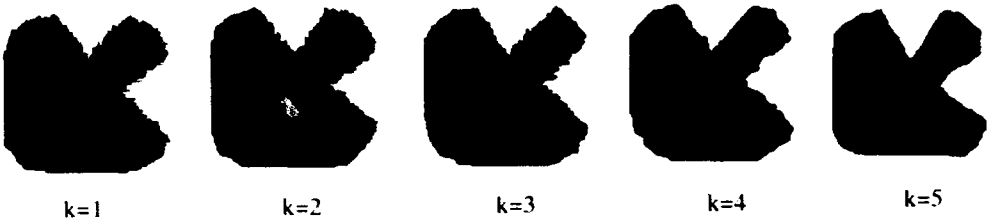


Fig. 4. Smoothing by requiring multiple support for growth: in the last 5 (of 60) steps, k neighbours are required.

3.5 Growth from a Skeleton

So far, only examples of growth from a single point have been shown. It is clear, however, that growth processes can be initiated at multiple points and allowed to proceed in parallel. As a simple example, Fig. 5 shows a curve-like "skeleton." At each point of this skeleton a probabilistic 4/8 growth process with $p/(1-p) = 2\sqrt{2}$ is initiated. The envelope of these processes gradually "fleshes out" the skeleton. No topological constraints are imposed on the parallel growth process; thus the growing shape will eventually fuse with itself. This paper will not investigate growth from a skeleton any further, but will concentrate only on growth from a single point.

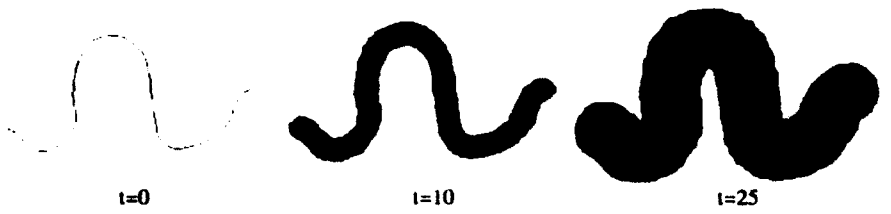


Fig. 5. Growth from a skeleton.

4 Nonisotropic Growth

The remainder of this paper describes methods for achieving "controlled" nonisotropic growth from a point where the control allows the production of elongation and concavities. It will be seen that simple probabilistic growth is quite limited in the types of shapes it can produce; but a richer class of shapes is produced by probabilistic mixing of nonisotropic deterministic growth processes.

4.1 Probabilistic Models

The simplest class of nonisotropic shapes is symmetric around the horizontal and vertical axes; ellipses and rectangles are classic examples. Three possible ways of growing such shapes using probabilistic growth are defined by the following kernels, all of which have the highest probability (1) of growth in the horizontal direction:

- Low-probability vertical growth and no diagonal growth:

$$\begin{matrix} 0\ p\ 0 \\ 1\ \square\ 1 \\ 0\ p\ 0 \end{matrix}$$

- Low-probability diagonal growth and no vertical growth:

$$\begin{matrix} p\ 0\ p \\ 1\ \square\ 1 \\ p\ 0\ p \end{matrix}$$

- Low-probability vertical and diagonal growth:

$$\begin{matrix} p\ p\ p \\ 1\ \square\ 1 \\ p\ p\ p \end{matrix}$$

The results obtained using these three kernels, for various values of p, are shown in Figs 6, 7, and 8, respectively. It appears that such processes can yield only pointed or blunted "ellipses." Note in particular that they do not yield significant concavities, even in directions in which the growth probability is zero.

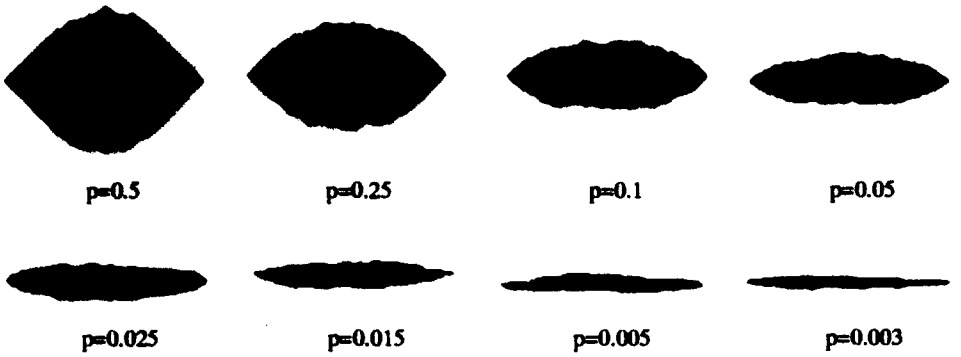


Fig. 6. Horizontal elongation: diagonal probability 0 and vertical probability p.

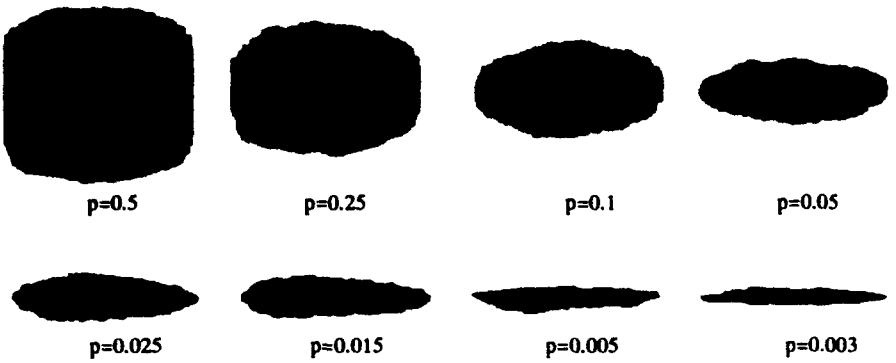


Fig. 7. Horizontal elongation: vertical probability 0 and diagonal probability p.

4.2 Deterministic Models with Phase Control

In this section, it will be shown that better-controlled nonisotropic growth can be achieved by using deterministic periodic growth processes in which the phases of the processes can be adjusted.

In directional periodic growth, the ith neighbour of a cell becomes a cell after t_i time steps. (If a point has more than one neighbouring cell, it becomes a cell at the earliest of the times t_i.) Thus, directional periodic growth is controlled by an 8-tuple of t_is, representing the time delays (positive integers) in the eight possible directions. This 8-tuple is called a *time delay kernel* (TDK) $T = (t_0, t_2, \ldots, t_7)$, $t_i \in \mathbb{Z}^+$:

$$\begin{array}{ccc} t_3 & t_2 & t_1 \\ t_4 & \square & t_0 \\ t_5 & t_6 & t_7 \end{array}$$

Periodic growth is implemented by decrementing the t_is of all cells by 1 at each time step. The ith neighbour of a cell becomes a cell one time step after t_i counts down to 1. When a new cell is created it is given a TDK, which in turn

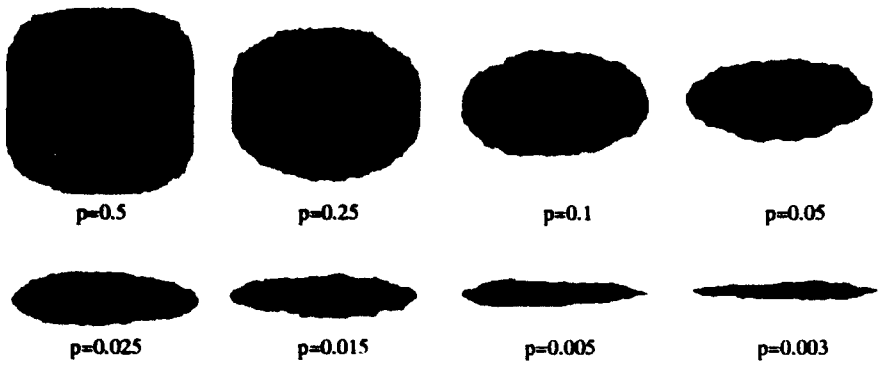


Fig. 8. Horizontal elongation: vertical and diagonal probability p.

controls the growth from that cell. The TDK of a new cell can be defined in at least two ways:

- The cell is given the same TDK that its parent *started* with. This is called the phase-resetting, or *restart clocks* method.
- The cell is given the *decremented* TDK of its parent. In other words, the new cell copies the t_is from its parent at the time step at which the cell is created (except for t_i, which is reset to its original value). This is the phase-preserving, or *copy clocks* method.

Both the restart clocks and the copy clocks methods are capable of growing octagons. However, the copy clocks method has the added capability of growing arbitrarily elongated octagons, as we shall now see.

Restart Clocks As in Sect. 4.1, T is restricted to be symmetric with respect to the x and y axes:

$$\begin{array}{ccc} b & c & b \\ a & \square & a \\ b & c & b \end{array}$$

The first three steps of a simple restart clocks process are shown in Fig. 9. In general, it is not hard to see that this process yields convex polygons whose vertices lie on the axes and on the diagonals. Thus, after n iterations, the kernel shown above yields the polygon shown in Fig. 10, where

$$\alpha = \left\lfloor \frac{n}{a} \right\rfloor, \quad \beta = \left\lfloor \frac{n}{b} \right\rfloor, \quad \gamma = \left\lfloor \frac{n}{c} \right\rfloor, \quad \delta = \left\lfloor \frac{n}{a+c} \right\rfloor .$$

Since the vertices lie on the lines $x = 0$, $y = 0$, $y = \pm x$, the polygon must be a square, hexagon, octagon, or rhombus (without loss of generality, assume $a \leq c$):

1. When $b \leq a$, the polygon is a square with vertices $(\pm\beta, \pm\beta), (\pm\beta, \mp\beta)$.

2. When $a < b \leq c$, it is a hexagon with vertices $\{(\pm\beta, \pm\beta), (\pm\beta, \mp\beta), (\pm\alpha, 0)\}$.
3. When $c < b < a+c$, it is an octagon with vertices $\{(\pm\alpha, 0), (0, \pm\gamma), (\pm\beta, \pm\beta), (\pm\beta, \mp\beta)\}$.
4. When $a + c \leq b$, it is a rhombus with vertices $\{(\pm\alpha, 0), (0, \pm\gamma)\}$.

Note that the polygon can be arbitrarily elongated in the horizontal direction (or the vertical direction, if $c < a$), but only by coming to a sharp point in that direction; this method cannot generate rectangles because the off-axis vertices must lie on the diagonals $y = \pm x$. (Some examples will be shown below.)

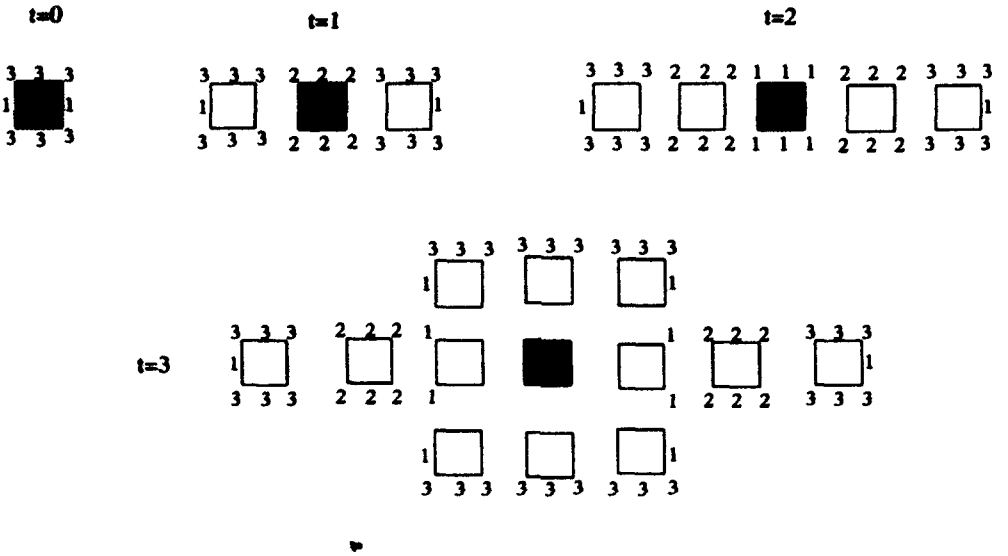


Fig. 9. A simple restart clocks growth process. The starting cell is shaded.

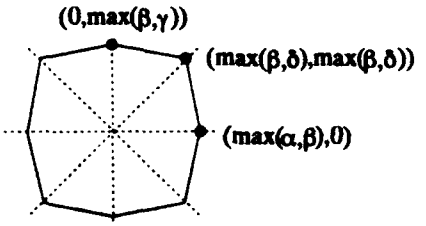


Fig. 10. The polygon produced by a restart clocks process.

Copy Clocks The first three steps of a simple copy clocks process are shown in Fig. 11. It can be shown that when horizontally and vertically symmetric time delay kernels are used, as in Sect. 4.2, convex polygons having horizontal,

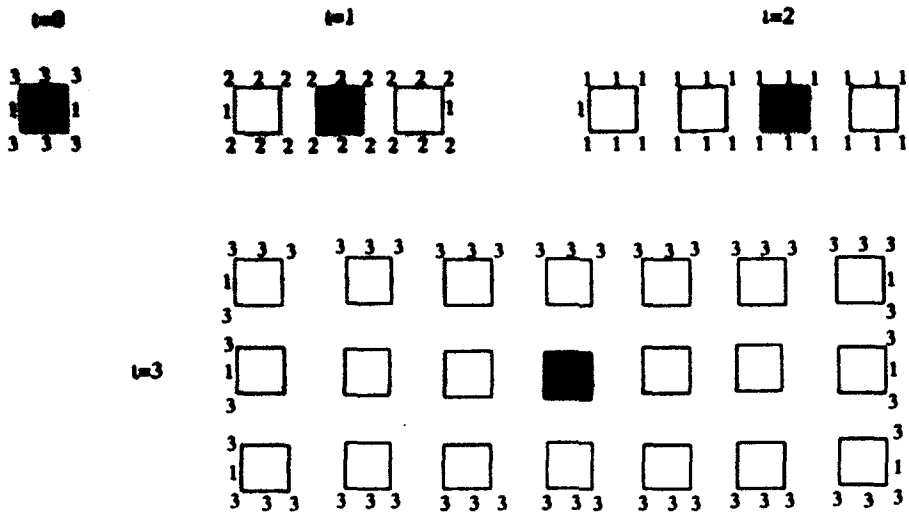


Fig. 11. A simple copy clocks growth process. The starting cell is shaded.

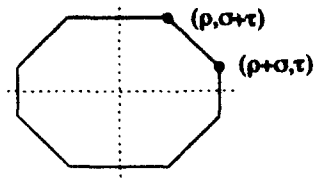


Fig. 12. The polygon produced by a copy clocks process.

vertical, and diagonal sides are obtained, as shown in Fig. 12. Let α, β, γ be as above, and let

$$\kappa = \left\lfloor \frac{n}{lcm(a,b)} \right\rfloor, \quad \lambda = \left\lfloor \frac{n}{lcm(b,c)} \right\rfloor, \quad \mu = \left\lfloor \frac{n}{lcm(a,c)} \right\rfloor, \quad \nu = \left\lfloor \frac{n}{lcm(a,b,c)} \right\rfloor,$$

where $lcm(x,y)$ is the least common multiple of x and y. Then it can be shown that

$$\begin{aligned} \rho &= (\alpha - \mu) + (\beta - \kappa) + \nu, \\ \sigma &= \mu - \nu, \\ \tau &= (\gamma - \mu) + (\beta - \lambda) + \nu. \end{aligned}$$

Since the slopes of the sides of a convex polygon must be monotonic, and here they are multiples of 45°, the polygon has at most eight sides. In fact, the polygon must be a rectangle, hexagon, diamond, or octagon (without loss of generality, assume $a \leq c$):

1. The polygon is a rectangle (with vertices $\{(\pm\rho, \pm\tau), (\pm\rho, \mp\tau)\}$) if and only if $\sigma = 0$; this means that $\mu = \nu$, i.e., that b divides $lcm(a, c)$.
2. The polygon is a hexagon (with vertices $\{(\pm\rho, \pm\sigma), (\pm\rho, \mp\sigma), (\pm(\rho+\sigma), 0)\}$) if and only if $\tau = 0$. It is not hard to see that this is equivalent to $\gamma = \mu$,

$\beta = \lambda$, and $\nu = 0$. Evidently, $\nu = 0$ requires $n < lcm(a, b, c)$. Since c divides $lcm(a, c)$, $\gamma = \mu$ implies $c = lcm(a, c)$, which implies that a divides c. Similarly, $\beta = \lambda$ implies that c divides b, so that $lcm(a, b, c) = b$. Thus a hexagon is obtained if and only if a divides c, c divides b, and $n < b$.

3. The polygon is a diamond (with vertices $\{(\pm\sigma, 0), (0, \pm\sigma)\}$) if and only if $\rho = \tau = 0$; as in the previous case, this is equivalent to $a = c$, a divides b, and $n < b$.
4. Otherwise, the polygon is an octagon (with vertices $\{(\pm\rho, \pm(\sigma + \tau)), (\pm\rho, \mp(\sigma + \tau)), (\pm(\rho + \sigma), \pm\tau), (\pm(\rho + \sigma), \mp\tau)\}$).

Note that the rectangle, hexagon, or octagon can be arbitrarily elongated, but the diamond is square.

Comparisons The restart and copy clocks methods are compared in Figs 13–14 for two TDKs:

- No diagonal growth, slow vertical growth;
- Slow diagonal growth, no or slow vertical growth.

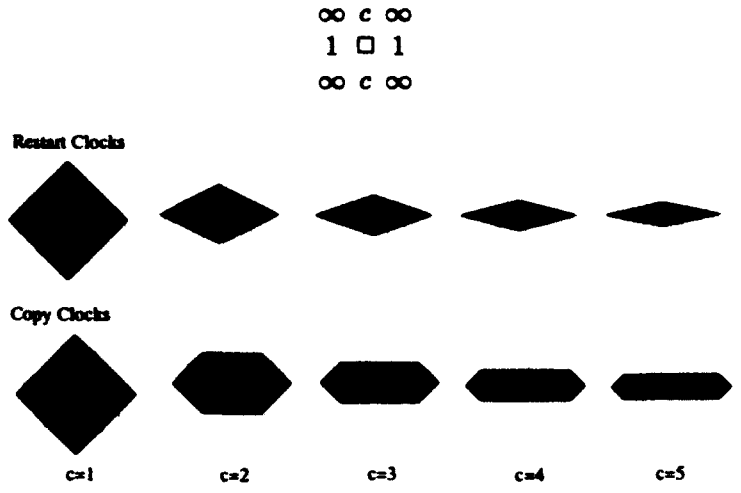


Fig. 13. Horizontal elongation (no diagonal growth, slow vertical growth).

Note that these TDKs all yield rhombuses, hexagons, or rectangles. Recall that the restart clocks method yields octagons (for $a \leq c$) when $c < b < a + c$, which requires $a > 1$; the copy clocks method yields octagons when b does not divide $lcm(a, c)$, and when either a does not divide c, c does not divide b, or $b \leq n$, for example as in Fig. 15. In the restart cases, the same hexagon is obtained for all $c \geq 3$; in the copy cases, rectangles are obtained when 3 divides c, and octagons are obtained otherwise.

$$\begin{matrix} b\ \infty\ b \\ 1\ \square\ 1 \\ b\ \infty\ b \end{matrix} \quad \text{or} \quad \begin{matrix} b\ b\ b \\ 1\ \square\ 1 \\ b\ b\ b \end{matrix}$$

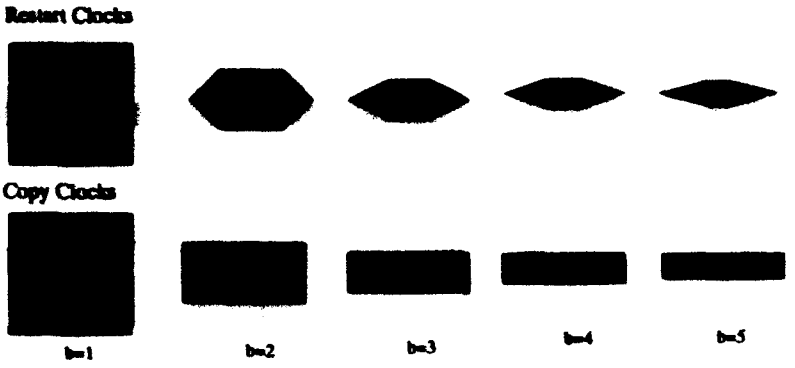


Fig. 14. Horizontal elongation (slow diagonal growth, no or slow vertical growth).

$$\begin{matrix} 3\ c\ 3 \\ 1\ \square\ 1 \\ 3\ c\ 3 \end{matrix}$$

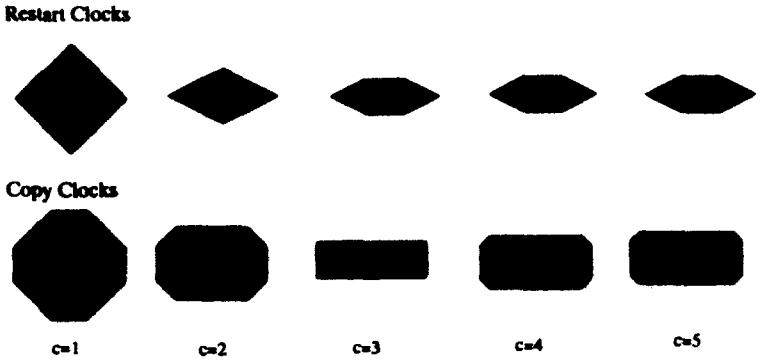


Fig. 15. Horizontal elongation (slow diagonal growth, very slow vertical growth).

Probabilistic Mixing Shapes with curved, but ragged boundaries can be produced by probabilistically mixing these deterministic processes; two examples are shown in Figs 16–17, using, respectively, the kernels

$$\begin{matrix} \infty\ 5\ \infty \\ 1\ \mathcal{R}\ 1 \\ \infty\ 5\ \infty \end{matrix} + \begin{matrix} 5\ 5\ 5 \\ 1\ \mathcal{C}\ 1 \\ 5\ 5\ 5 \end{matrix} \quad \text{and} \quad \begin{matrix} \infty\ 5\ \infty \\ 1\ \mathcal{C}\ 1 \\ \infty\ 5\ \infty \end{matrix} + \begin{matrix} 5\ 5\ 5 \\ 1\ \mathcal{C}\ 1 \\ 5\ 5\ 5 \end{matrix}$$

where $\mathcal{R}$ and $\mathcal{C}$ denote "restart" and "copy." Thus, this approach makes it possible to obtain elongated shapes that have rounded ends.

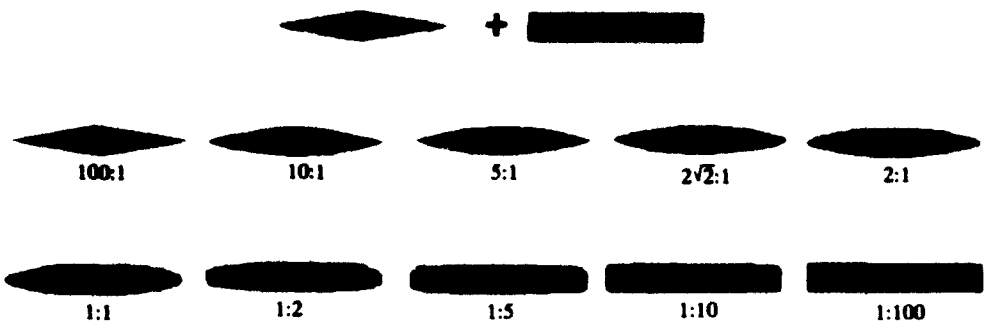


Fig. 16. Probabilistically mixing a rhombus with a rectangle.

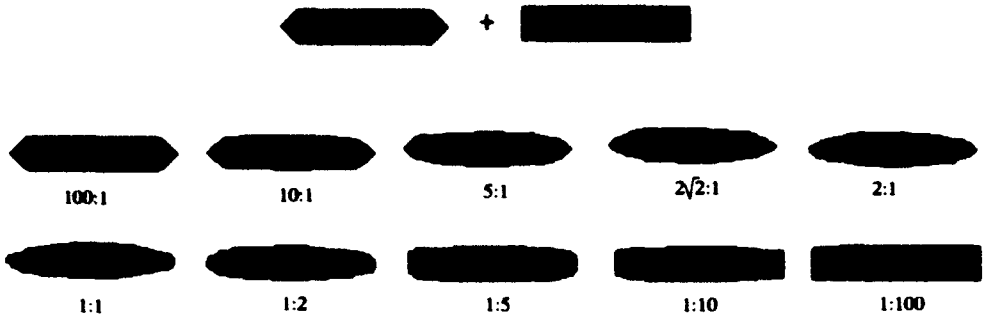


Fig. 17. Probabilistically mixing a hexagon with a rectangle.

4.3 Time-Varying Growth

In all of the previous examples, the kernels have remained the same throughout the growth process. In this section two examples are shown where the kernels change during the process. In the first example, the kernels change abruptly half way through the growth (Fig. 18); in the second example, the first pair of kernels gradually changes into the second over a period of 15 iterations (Fig. 19). The kernels used were

$$\begin{matrix} \infty\ 5\ \infty \\ 1\ \mathcal{R}\ 5 \\ \infty\ 5\ \infty \end{matrix} + \begin{matrix} 5\ 5\ 5 \\ 1\ \mathcal{C}\ 5 \\ 5\ 5\ 5 \end{matrix} \quad \text{changing to} \quad \begin{matrix} 5\ \infty\ 5 \\ \infty\ \mathcal{R}\ \infty \\ 1\ \infty\ 5 \end{matrix} + \begin{matrix} 5\ 5\ 5 \\ 5\ \mathcal{C}\ 5 \\ 1\ 5\ 5 \end{matrix}$$

4.4 Concavities

All the time-invariant growth processes described thus far can produce only (ragged) convex shapes. Setting the probability in a particular direction to zero, or the time delay to infinity, does not create a concavity because growth from neighbouring cells fills it in. It will now be shown that, by making appropriate changes in the kernel, concavities can be produced.

Fig. 18. Changing the kernels abruptly; the resulting elongated shape takes an abrupt "turn."

Fig. 19. A gradual change from one pair of kernels to another produces a smoother "turning" effect.

To prevent a concavity from filling in from the sides, a growth process must be introduced in which some child cells can have longer periods (or lower probabilities) than their parents. (Note that in both the restart and copy clocks methods, the time delays in a child's TDK never exceed those in its parent's TDK.) A concavity can then be formed by giving certain child cells long periods in the appropriate directions. An example of this method is shown in Fig. 20. In this example, when children are produced by a cell whose TDK has a local maximum (i.e., a long delay flanked by short ones), indicating slow growth in a certain direction, these children are given TDKs having long delays on the sides that face that direction.

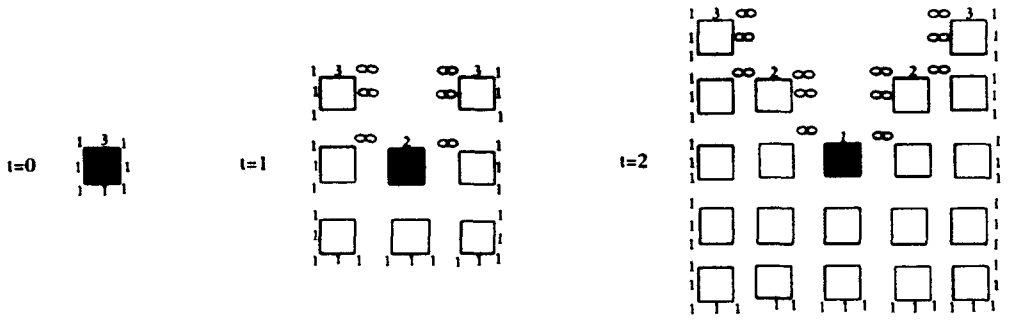


Fig. 20. The first two steps of a concavity-producing process.

Concavities can be produced in this way in any of the eight directions. The depth of a concavity is controlled by the length of the time delay(s) specified in the TDK (Fig. 21). Lobes or protrusions can be produced by creating pairs of concavities. Probabilistic mixing controls the concavity sharpness and the lobe shape (Figs 22–23).

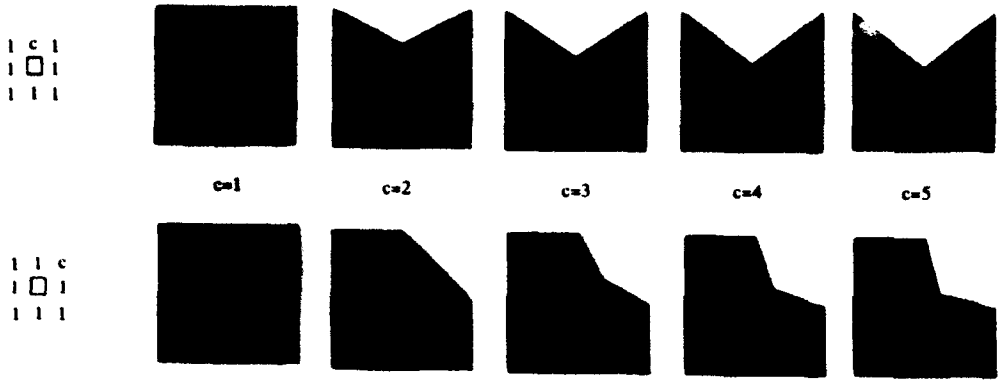


Fig. 21. Vertical and diagonal concavities. (Only the original TDKs are shown.)

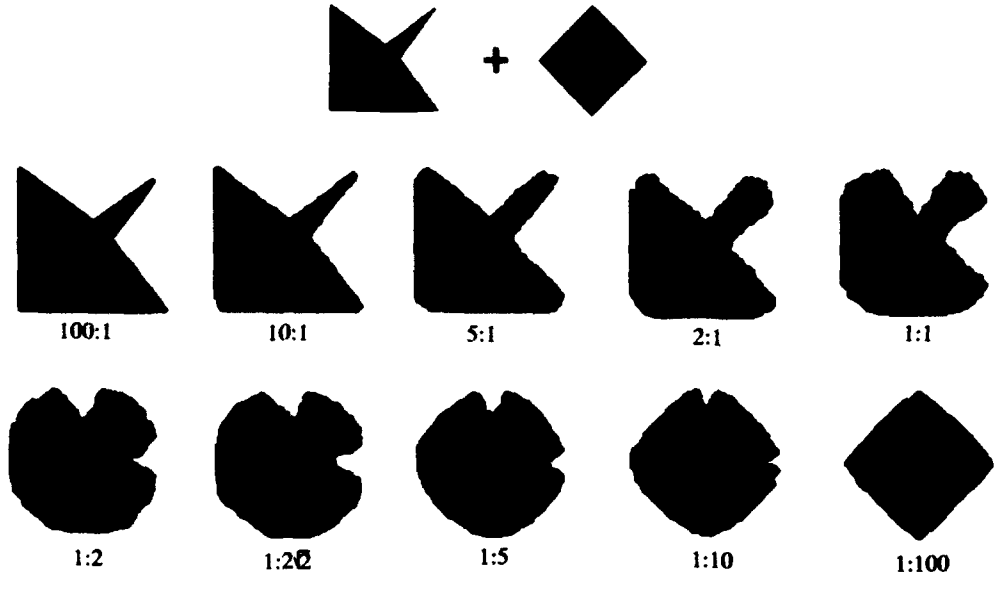


Fig. 22. Producing lobes by combining pairs of concavities. Probabilistic mixing controls the lobe shape.

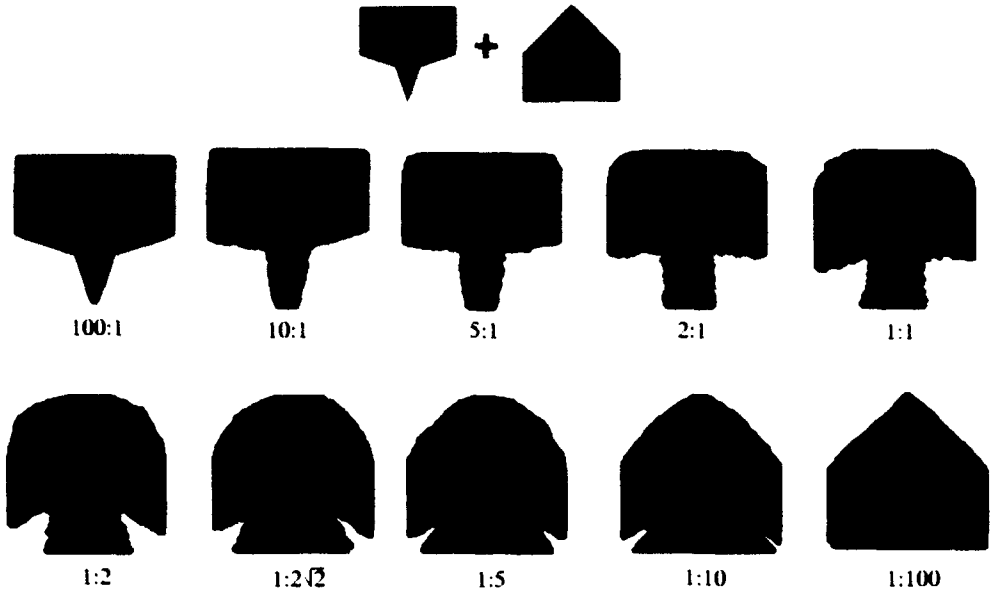


Fig. 23. A second example of lobe generation.

5 Concluding Remarks

In this paper, simple deterministic (periodic) and probabilistic models have been described for growth of a shape from a point on a two-dimensional Cartesian grid. It was found that "natural"-looking shapes can be produced by probabilistically combining deterministic growth processes which would individually yield simple polygons. In particular, it was found that long parallel-sided "ribbons" with rounded ends can be generated by mixing a process that produces rectangles with a process that produces elongated hexagons or rhombuses; and that a "ribbon" can be made to turn smoothly by gradually changing the directional bias of the growth periods (or probabilities). (Evidently ribbons of varying width can be produced by using a varying growth rate.) It was also found that these models can be made to yield shapes with concavities by allowing some of a child cell's growth periods to be longer than those of its parents (or some of its probabilities to be lower than its parents'). It was shown that the raggedness of the shapes can be reduced by requiring "local support" (k-adjacency) for new cells. A more detailed paper, including proofs of the results stated in this paper, is in preparation.

The models introduced in this paper have many possible generalizations. It would be of interest to study the general class of processes in which the periods (or probabilities) have Markovian dependencies on the previous generations' periods. Evidently a wide variety of complex, hierarchically-structured "natural" shapes can be generated in this way.

It would also be of interest to study models in which "environmental" factors can influence the growth (in analogy with the effects of gravity, illumination, nutrition, etc., on the growth of organisms), and to study how two growing

shapes might interact (or how a growing shape might interact with itself, e.g. to avoid "fusing" with itself). In addition, it would be of interest to study processes in which cells can "die"; this allows the growth of shapes that have holes.

In addition, various modifications of the ideas in this paper could be investigated. This paper dealt only with a discrete, two-dimensional Cartesian grid and with a discrete, synchronous sequence of time steps; other grids (in two or more dimensions), or the growth of clusters of cells in Euclidean space, where the time between the creation of a cell and the creation of its children is a random variable, could be considered. Hierarchical or "multigrid" growth processes, which take place at more than one "scale", could be studied (see, e.g., [2]). Finally, it would be of considerable interest to investigate the recovery of growth models from examples of their output.

References

1. Conway, J. (1985). Winning Ways for Mathematical Plays, Academic Press, London.
2. Edelman, G. (1988). Topobiology: An Introduction to Molecular Embryology, Basic Books, New York, pp. 3–55.
3. Eden, M. (1961). A two-dimensional growth process. In: Neyman, F. (ed.), Proc. 4th Berkeley Symposium on Mathematics, Statistics, and Probability, Vol. 4, University of California Press, Berkeley, pp. 223–239.
4. Haralick, R.M. (1974). A measure for circularity of digital figures, IEEE Trans. on Systems, Man, and Cybernetics 4, pp. 394–396.
5. Langman, J. (1977). Medical Embryology, Williams and Wilkins, Baltimore.
6. Prusinkiewicz, P., Hanan, J. (1989). Lindenmayer Systems, Fractals, and Plants, Springer-Verlag, New York.
7. Prusinkiewicz, P., Lindenmayer, A. (1990). The Algorithmic Beauty of Plants, Springer-Verlag, New York.
8. Rosenfeld, A., Pfaltz, J.L. (1968). Distance functions on digital pictures, Pattern Recognition 1, pp. 33–61.
9. Schrandt, R.G., Ulam, S. (1967). On recursively defined geometrical objects and patterns of growth, Technical Report LA-3762, Los Alamos Scientific Laboratory, University of California. Reprinted in: Bednarek, A.R., Ulam, F. (eds.) (1990). Analogies between Analogies, The Mathematical Reports of S.M. Ulam and his Los Alamos Collaborators, University of California Press, Berkeley, Chapter 12.
10. Serra, J. (1982). Image Analysis and Mathematical Morphology, Academic Press, London.
11. Vicsek, T. (1992). Fractal Growth Phenomena, World Scientific, Singapore.
12. Wolfram, S. (1986). Theory and Applications of Cellular Automata, World Scientific, Singapore.

Classical and Fuzzy Differential Methods in Shape Analysis*

David H. Foster

Department of Communication and Neuroscience, Keele University, Staffordshire ST5 5BG, UK

Abstract. This study considers four means of defining differential operators for extracting local aspects of shape in ill-specified environments: fuzzy differentiation as kernel smoothing; differentiation in the sense of weak or generalized derivatives; differentiation for fuzzy functions between normed spaces; and fuzzy differentiation for mappings between fuzzy manifolds. More consideration is given to the last, norm-free approach, which involves the notions of an abstract fuzzy topological vector space, fuzzy differentiation between fuzzy topological vector spaces, fuzzy atlases, and tangent vectors of fuzzy manifolds.

Keywords: shape description, differential geometry, fuzzy set, fuzzy derivative, fuzzy topological vector space, fuzzy manifold, tangent vector, tangent space.

1 Introduction

A common technique for characterizing shape in an image is to use some kind of differential operator to extract the critical local variations in the light distribution. For images of two-dimensional objects, and their boundaries in particular, one might determine the positions of curvature extrema [1, 38, 9]; and, for images of three-dimensional objects, the positions of extrema in, for example, principal curvatures [17].

Yet in real vision systems, whether machine or human, imprecisions are inherent in the spatial and intensity characterization of the image. At the lowest, most immediate levels of image representation, there are effects of noise in sensory transduction and of limits on sampling frequency, both spatial and temporal. At higher, more removed levels of image representation [13], there are more general imprecisions to do with the specification of image qualities [40]. For the human observer it is unclear what geometrical framework is used to form the representation, and indeed whether a metric structure or the structure of

* I am grateful to P. Fletcher, R. Kopperman, S.R. Pratt, and M.G.A. Thomson for critically reading the manuscript and to J.J. Koenderink for comments on Sect. 2. This work was supported by ESPRIT Basic Research Action No. 6448 (VIVA).

a normed space is part of it [11, 14]. How then should differential operators be defined for these ill-specified environments?

The approaches to this problem have differed in the restrictions they have placed on the class of admissible image characterizations and on the analytic machinery assumed to be available at each processing stage. Four of the main approaches may be summarized as follows.

1. Assume a Euclidean framework and smooth the low-level image representation. The classical differential methods of real analysis may then be applied straightforwardly.
2. Assume that the low-level representation is important only in the way that it "interacts" with certain other functions. For a sufficiently large set of such functions, this interaction defines an operator which is differentiable, in the sense of generalized derivatives, and which can be used in place of the representation.
3. Assume that the image representation is "fuzzy" but constrained in such a way that it may be isometrically embedded in a normed space, which then allows classical differential methods to be applied.
4. Assume that the image representation is fuzzy and introduce a natural fuzzy topological vector space structure—or more generally the structure of a fuzzy differentiable manifold—so that the notion of fuzzy differentiation follows naturally without the imposition of a norm.

This article reviews briefly methods (1)–(3), and then more fully method (4), which involves some relatively unfamiliar topological-geometrical notions. The treatment is not complete: topological [19, 23, 24] and graph-based [22] digital-topological approaches are not considered, nor are synthetic methods [20].

It is assumed, with little loss in generality, that the images of interest are monochromatic, viewed monocularly.

2 Fuzzy Differentiation as Kernel Smoothing

Suppose that the image is represented by some luminance distribution $I(\mathbf{x})$, where $\mathbf{x}$ ranges over the real plane $\mathbb{R}^2$, and suppose that I is non-smooth in some way, that is, I or its first or second derivative is discontinuous in the standard differential structure on $\mathbb{R}^2$. There are various ways of smoothing the data defined by I. A *kernel smoother* uses an explicit set of local weights, defined by the *kernel* K, to produce the smoothed estimate $\tilde{I}$ of I at each $\mathbf{x}$ [42, 16]; thus

$$\tilde{I}(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{x}')\, I(\mathbf{x}')\, d\mathbf{x}' .$$

If I is obtained by discrete sampling, that is, determined only on a finite subset $\{\mathbf{x}_i\}_{1 \le i \le n}$ of points in $\mathbb{R}^2$, the integral is replaced by a summation over i [16].

In general the kernel takes the form

$$K(\mathbf{x}) = (c_0/\sigma)\, d(\|\mathbf{x}\|/\sigma) ,$$

where d is a decreasing function; $\|\cdot\|$ is a norm; σ is the window-width or bandwidth; and c_0 is a normalizing constant. There are several criteria for the choice of kernel [31]; in the present context a natural candidate for d is the standard Gaussian function [42].

For functions of $\mathbb{R}^2$, and for luminance distributions in particular, a definition of a fuzzy derivative has been been proposed [21] that may be viewed as a kernel smoother, the kernel being the derivative of a Gaussian function; that is:

Definition 1. The n^{th} (partial) *fuzzy derivative* at $x \in \mathbb{R}$ is the kernel

$$\phi_n(x;s) = \frac{\partial^n}{\partial x^n}\left(\frac{e^{-x^2/4s}}{\sqrt{4\pi s}}\right),$$

where $\sigma = \sqrt{4s}$ sets the *scale* parameter.

The functions ϕ_n have a ready physical and physiological interpretation [21], and show a concatenation property such that the higher-order derivatives are obtained at lower spatial "resolutions", the resolution corresponding to the inverse of the scale parameter value σ. A discretized version of this scale-space approach has been described in [27, 28], where a discrete analogue of the Gaussian kernel is used.

There is, however, a fundamental problem of deciding how appropriately the fitted surface represents the original surface [10, 4]. A critical question, for example, is whether Gaussian smoothing leads to robust derivatives. As has been noted elsewhere [43], there are two conflicting requirements: accuracy (correct derivatives should be obtained, at least for low orders), and smoothing (the effects of noise and discretization should be minimized). Gaussian kernels can lead to "over-smoothing" errors, but other kernels can be derived that achieve a better compromise between these two requirements [43]. The technique of *adaptive* kernel estimation has been reviewed in [42].

The approach summarized in Definition 1 and developed in [21] differs from some others in that it does not assume necessarily that an "original" surface exists, other than that which can be observed through the kernels (see Sect. 3). This foundational issue has been circumvented in an approach [4] that uses a statistical covariance technique [26] for surface descriptors. By analogy with classical differential methods, the technique yields, for discretely sampled data, definitions of the first and second fundamental forms for a surface in $\mathbb{R}^3$, and the Weingarten equations, which relate the rate of change of the unit normal vector and the corresponding chosen direction of a curve on the tangent plane [4].

The next section considers more generally the notion of derivatives as operators.

3 Generalized Derivatives

Suppose that the image luminance distribution $I(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, is such that it can be associated formally with an operator on a set of "test" functions on $\mathbb{R}^2$.

(The association may be through convolution, as in the preceding section; the test functions are defined shortly, after a natural topology for them is introduced.) Although derivatives of the representation may not be defined in the ordinary way, derivatives of this operator *may* be defined, providing that certain conditions are satisfied.

The set of test functions is given a topology based on a family of seminorms. A *seminorm* on a vector space E is a mapping $\rho : E \to [0, \infty)$ such that:

1. $\rho(\xi + \eta) \leq \rho(\xi) + \rho(\eta)$, for all $\xi, \eta \in E$.
2. $\rho(\alpha\xi) = |\alpha|\,\rho(\xi)$, for all $\xi \in E, \alpha \in \mathbb{C}$ (or $\mathbb{R}$).

A family $\{\rho_\gamma\}_{\gamma \in \Gamma}$ of seminorms *separates points* if

3. $\rho_\gamma(\xi) = 0$ for all $\gamma \in \Gamma$ implies $\xi = 0$.

The *natural topology* on a vector space with a family $\{\rho_\gamma\}_{\gamma \in \Gamma}$ of seminorms separating points is the weakest topology in which all the ρ_γ are continuous and in which the operation of addition is continuous.

The set of test functions on $\mathbb{R}^2$ (or, more generally, $\mathbb{R}^n$) is the set $\mathcal{S}$ of *functions of rapid decrease*; that is, the set of infinitely differentiable functions ϕ on $\mathbb{R}^2$ for which

$$\sup_{(x_1,x_2)\in\mathbb{R}^2} \left| (x_1^{\alpha_1}, x_2^{\alpha_2}) \frac{\partial^{(\beta_1+\beta_2)}}{\partial x_1^{\beta_1} \partial x_2^{\beta_2}} \phi(x_1, x_2) \right| < \infty, \tag{1}$$

for all non-negative integers $\alpha_1, \alpha_2, \beta_1, \beta_2$. The functions in $\mathcal{S}$ are thus those that together with their derivatives fall off more rapidly than the inverse of any polynomial. The quantity on the left-hand side of (1) defines a seminorm $\|\cdot\|_{\alpha,\beta}$ on $\mathcal{S}$. These seminorms give $\mathcal{S}$ the natural topology.

The space of operators can now be defined as the (topological) dual of $\mathcal{S}$; that is, the set of all continuous linear functions ("functionals") on $\mathcal{S}$. It is denoted by $\mathcal{S}'$ and called the *space of tempered distributions* [41]. The derivative of a tempered distribution is defined as follows.

Definition 2. Let T be a tempered distribution. The *weak* or *generalized derivative* $D^{(\alpha_1,\alpha_2)}T$ (or the *derivative in the sense of distributions*) is given by

$$(D^{(\alpha_1,\alpha_2)}T)(\phi) = (-1)^{|\alpha_1+\alpha_2|}\, T\left(\frac{\partial^{(\alpha_1+\alpha_2)}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \phi\right), \quad \text{for all } \phi \in \mathcal{S}.$$

There is a natural way to associate a certain class of functions f on $\mathbb{R}^2$ with tempered distributions T_f such that if $T_f = T_g$ then $f = g$ almost everywhere. For an image luminance distribution that falls into this class, its derivative may thus be defined as the derivative of the corresponding tempered distribution. The connection between this approach to the differentiation of image functions and the smoothing approach of Sect. 2 is discussed in [10].

4 Normed Spaces of Fuzzy Sets

Suppose, now, that the sampling of the image is less precisely specified. For example, consider a function that assigns to each point $\mathbf{x}$ in $\mathbb{R}^2$ with luminance $I(\mathbf{x})$ some measure of the "goodness" of this characterization of the image at that point, or, more generally, consider a function that assigns to each element of some set X of image attributes, possibly including an estimate of spatial position, a number that specifies the extent to which that attribute is associated with the image or part of the image. Both of these functions are examples of "fuzzy sets", the formal notion of which was introduced by Zadeh [44]. Thus, given an arbitrary set X, a *fuzzy set* (or *fuzzy subset*) in X is a function $A : X \to [0,1]$ such that the value $A(x)$ of A at the point $x \in X$ gives the "grade of membership" of x in A. (Fuzzy set theory should not be confused with probability theory; for discussion of this and related issues, see [45, 33, 34].) For a classical set the grade of membership would be either 0 or 1 (and A would then coincide with its characteristic function). The grade of membership of a fuzzy set may be taken in a complete lattice [15]—that is, a lattice in which every subset has a supremum and an infimum—rather than in the unit interval $[0,1]$; see [33] for examples. A kind of fuzziness for which there is *no* greatest element has been considered in [32], but this weaker structure limits the definition of a topology (Sect. 6).

The set $\mathcal{F}(X)$ of all fuzzy sets in X is a complete distributive lattice. For any fuzzy set A and any number $\alpha \in [0,1]$, the *α-cut* A_α of A is the set $\{x \in X \mid A(x) \geq \alpha\}$. If X is a vector space, a *convex* fuzzy set A in $\mathcal{F}(X)$ has the property that

$$A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{A(x_1), A(x_2)\},$$

for every x_1, x_2 in X, and λ in $[0,1]$.

The next section considers the differentiation of a "fuzzy" function from a normed vector space into a set of fuzzy sets in a reflexive Banach space Y with norm $\|\cdot\|$. It is possible to introduce a norm on a subset of $\mathcal{F}(Y)$, the set of all fuzzy sets in Y. Recall that the Hausdorff distance $d_{\mathrm{H}}(P,Q)$ between non-empty bounded (classical) subsets P, Q of Y is given by

$$d_{\mathrm{H}}(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q}\|p-q\|, \sup_{q\in Q}\inf_{p\in P}\|p-q\|\right\}.$$

This distance can be extended to the subset $\mathcal{F}_0(Y)$ of $\mathcal{F}(Y)$ containing those fuzzy sets A with the following properties [36]:

1. A is upper semicontinuous;
2. A is convex;
3. A_α is compact for every α.

For $A, B \in \mathcal{F}_0(Y)$, define the distance $d(A,B)$ between A and B by

$$d(A,B) = \sup_{\alpha>0}\{d_{\mathrm{H}}(A_\alpha, B_\alpha)\}.$$

Then it can be shown [36] that $(\mathcal{F}_0(Y), d)$ is a complete metric space.

The subset $\mathcal{F}_0(Y)$ can be given a linear structure in the following way [36]. For $A, B \in \mathcal{F}_0(Y)$, the *sum* $C = A + B$ of A, B (sometimes denoted by $A \oplus B$) is the fuzzy set in Y defined by

$$C(y) = \sup_{\alpha \in [0,1]} \{ \alpha \mid y \in (A_\alpha + B_\alpha) \}, \quad \text{for all } y \in Y,$$

where $A_\alpha + B_\alpha$ is the (classical) subset $\{ z \in Y \mid z = a + b, a \in A_\alpha, b \in B_\alpha \}$. For any scalar $\alpha \in \mathbb{R}$, the *scalar product* αA of α and A is the fuzzy set in Y defined by

$$(\alpha A)(y) = \begin{cases} A(y/\alpha), & \text{if } \alpha \neq 0, \\ 0, & \text{if } \alpha = 0 \text{ and } y \neq 0, \\ \sup_{z \in Y} A(z), & \text{if } \alpha = 0 \text{ and } y = 0. \end{cases}$$

Although $\mathcal{F}_0(Y)$ is not a vector space with this sum and product [37, 36], the embedding theorem of Rådström [37] may be used to embed $\mathcal{F}_0(Y)$ isometrically in a normed vector space. Let $\mathcal{Y}$ be this normed space and let $j : \mathcal{F}_0(Y) \to \mathcal{Y}$ denote the embedding.

5 Differentiation of a Fuzzy Function between Normed Spaces

One definition [36] of a *fuzzy function* f from an arbitrary set X to an arbitrary set Y is that it is a set-valued mapping or multifunction [3] that assigns to each point $x \in X$ a fuzzy set $f(x) \in \mathcal{F}(Y)$ (but see e.g. [33] for other interpretations). Suppose that X is a normed vector space; U a (classical) open subset of X; Y a reflexive Banach space, as in Sect. 4; and f a fuzzy function from U into Y such that $f(x) \in \mathcal{F}_0(Y)$; that is, for each $x \in X$, the fuzzy set $f(x)$ has the properties (1)–(3) of Sect. 4. Then the differentiability of f at a point in U may be defined [36] by the differentiability of its composition with the embedding j in the normed vector space $\mathcal{Y}$; thus:

Definition 3. The fuzzy function $f : U \to \mathcal{F}_0(Y)$ is *differentiable at a point* $x_0 \in U$ if the composition $\hat{f} = j \circ f$ is differentiable at x_0; that is, if there exists a linear bounded mapping $\hat{f}'(x_0)$ from X into $\mathcal{Y}$ such that

$$\lim_{x \to x_0} \left\{ \frac{\| \hat{f}(x) - \hat{f}(x_0) - \hat{f}'(x_0)(x - x_0) \|}{\| x - x_0 \|} \right\} = 0.$$

Further details are given in [36, 3], where the Hukuhara differential is also discussed.

By definition [46, 30], a *type 2* fuzzy set A in a set X is a fuzzy set characterized by a fuzzy membership function whose values are each fuzzy sets in the unit interval $[0,1]$; that is, for each $x \in X$, the grade of membership $A(x) : J \to [0,1]$, where $J \subset [0,1]$. Type 2 fuzzy sets are a special case of

the fuzzy functions just defined. An application of Definition 3 might thus be to those image characterisations that form type 2 fuzzy sets; that is, as in Sect. 4, where each point x in $\mathbb{R}^2$ of the image is associated with a fuzzy estimate $f(\mathbf{x})$ of (normalised) luminance.

6 Fuzzy Topology and Fuzzy Topological Vector Spaces

Consider, next, fuzzy sets in a set X where there is *no* norm. As will become clear later, all that is needed for a basic definition of differentiation is that X should be equipped with an appropriately fuzzy version of the structure of a topological vector space.

Note. In fact an even simpler framework is possible. R. Kopperman has considered (1992, personal communication) the equivalent definition: $f'(x)$ is a *derivative* for f at x if $f(y) = f(x) + m(x,y)(y-x)$ and $\lim_{y\to x} m(x,y) = f'(x)$, with $m(x,y)$, the *slope* of f between x and y, defined for $x, y \in \mathrm{Dom}(f)$. This definition extends easily to any category of topological abelian groups such that if X, Y are topological abelian groups, then $\mathrm{Hom}(X,Y)$ is also a topological abelian group and $[(f,x) \to f(x)] : \mathrm{Hom}(X,Y) \times X \to Y$ and $[(f,g) \to f \circ g] : \mathrm{Hom}(X,Y) \times \mathrm{Hom}(Z,X) \to \mathrm{Hom}(Z,Y)$ are jointly continuous. In this situation, functions are continuous at points of differentiability and the chain rule and sum rule hold; further, theorems on partial derivatives and the inverse and implicit function theorems, among others, can be formulated and shown in natural settings (see Sect. 8).

For the sake of completeness, some elementary properties of fuzzy sets are briefly recalled [44, 33, 35, 7]. For each $c \in [0,1]$, let k_c denote the *constant* fuzzy set in X, that is, $k_c(x) = c$ for all $x \in X$; and let x_c denote the *fuzzy point* in X, where

$$x_c(y) = \begin{cases} c, & \text{for } y = x; \\ 0, & \text{otherwise.} \end{cases}$$

For a fuzzy set A in X, one writes $x_c \in A$ when $c \leq A(x)$. The set X is identified with the constant fuzzy set k_1 and the empty set is identified with k_0. The inclusion, intersection, union, and complement of two arbitrary fuzzy sets are defined in an obvious fashion [44, 7]; for example, for fuzzy sets A, B in X, the intersection $A \cap B$ is given by $(A \cap B)(x) = \min\{A(x), B(x)\}$, for all $x \in X$.

Let f be a mapping from a set X to a set Y. Let B be a fuzzy set in Y. Then the *inverse image* $f^{-1}[B]$ of B is the fuzzy set in X defined by $f^{-1}[B](x) = B(f(x))$, for all $x \in X$. Conversely, let A be a fuzzy set in X. Then the *image* $f[A]$ of A is the fuzzy set in Y defined by

$$f[A](y) = \begin{cases} \sup_{z \in f^{-1}(y)} A(z), & \text{if } f^{-1}(y) \text{ is nonempty,} \\ 0, & \text{otherwise.} \end{cases}$$

Notice that although f takes fuzzy sets into fuzzy sets, it is not a set-valued mapping in the sense of Sect. 5, where (classical) points are taken into fuzzy sets.

The following definition of a fuzzy topological space is due to Lowen [29]. A *fuzzy topology* on a set X is a family $\mathcal{T}$ of fuzzy sets in X that satisfies the following conditions:

1. For all $c \in [0,1]$, $k_c \in \mathcal{T}$.
2. If $A, B \in \mathcal{T}$, then $A \cap B \in \mathcal{T}$.
3. If $A_j \in \mathcal{T}$ for all $j \in J$ (J some index set), then $\bigcup_{j \in J} A_j \in \mathcal{T}$.

In the definition of a fuzzy topology due to Chang [5], the condition (1) is

1'. $k_0, k_1 \in \mathcal{T}$.

The inclusion in $\mathcal{T}$ of all fuzzy sets that are constant functions on X is required for the fuzzy continuity of the constant functions from X to any other set Y equipped with a fuzzy topology (fuzzy continuity is defined shortly). A fuzzy topology that satisfies condition (1) is called a *proper fuzzy topology*. The pair $(X, \mathcal{T})$ is called a *fuzzy topological space*. An *open* fuzzy set A in X is one which is in $\mathcal{T}$, and a *closed* fuzzy set is one whose complement $\bar{A} = 1 - A$ is in $\mathcal{T}$. A fuzzy set B is a *neighbourhood* of a fuzzy point x_c in X if there is a fuzzy set A in $\mathcal{T}$ such that $x_c \in A \subset B$. A fuzzy topological space is called a *fuzzy* T_1 space if every fuzzy point is a closed fuzzy set.

Let $(X, \mathcal{T}), (Y, \mathcal{V})$ be two fuzzy topological spaces. A mapping f of $(X, \mathcal{T})$ into $(Y, \mathcal{V})$ is *fuzzy continuous* if for each open fuzzy set V in $\mathcal{V}$ the inverse image $f^{-1}[V]$ is in $\mathcal{T}$. Conversely, f is *fuzzy open* if for each open fuzzy set U in $\mathcal{T}$, the image $f[U]$ is in $\mathcal{V}$. For related properties, including the notions of an induced fuzzy topology on a fuzzy set, and relatively fuzzy continous and relatively fuzzy open mappings, see [12].

Suppose that E is a vector space over $\mathbb{K}$ (the real field $\mathbb{R}$ or complex field $\mathbb{C}$). Let A, B be fuzzy sets in E. The definitions of the sum and scalar product (Sect. 4) may be reformulated thus. The *sum* $C = A + B$ of A, B is the fuzzy set in E defined by

$$C(x) = \sup_{a+b=x} \min\{A(a), B(b)\}, \quad \text{for all } x \in E;$$

and, for any scalar $\alpha \in \mathbb{K}$, the *scalar product* αA of α and A is the fuzzy set in E defined by

$$(\alpha A)(x) = \begin{cases} A(x/\alpha), & \text{for } \alpha \neq 0, \\ 0_c(x), & \text{otherwise,} \end{cases}$$

for all $x \in E$, where 0_c is the fuzzy point at 0 in E with $c = \sup_{y \in E} A(y)$.

Suppose that E is equipped with a fuzzy topology $\mathcal{T}$ and that $\mathbb{K}$ is equipped with the usual topology $\mathcal{K}$. A *fuzzy topological vector space* (*ftvs*) is a vector space E over $\mathbb{K}$ such that [18] the two mappings

1. $(x, y) \mapsto x + y$ of $(E, \mathcal{T}) \times (E, \mathcal{T})$ into $(E, \mathcal{T})$,
2. $(\alpha, x) \mapsto \alpha x$ of $(\mathbb{K}, \mathcal{K}) \times (E, \mathcal{T})$ into $(E, \mathcal{T})$,

are fuzzy continuous. Notice that the fuzzy topological vector space E may be proper or improper, but $\mathbb{K}$ is a special case of an improper fuzzy topological vector space. In the sequel, E denotes a ftvs with scalar field $\mathbb{K}$.

7 Fuzzy Differentiation Between Fuzzy Topological Vector Spaces

The following definition of a fuzzy derivative is a generalization of the classical definition for topological vector spaces [25]. Let E, F be two fuzzy ftvs's and let ϕ be a mapping from E into F. Let $o(t)$ be any function of a real variable t such that $\lim_{t\to 0} o(t)/t = 0$. Then ϕ is *tangent to* 0 if given a neighbourhood W of 0_δ in F, $0 < \delta \leq 1$, there exists a neighbourhood V of 0_λ in E, $0 < \lambda < \delta$, such that

$$\phi[tV] \subset o(t)W\,,$$

for some function $o(t)$. If both V, W are classical sets and E, F are normed, then this amounts [25] to the usual condition

$$\|\phi(x)\| \leq \|x\|\psi(x)\,,$$

where $\lim_{\|x\|\to 0} \psi(x) = 0$.

Let E, F be two ftvs's, each endowed with a fuzzy T_1 topology. Let $f : E \to F$ be fuzzy continuous. The fuzzy differentiability of f at a point in E may be defined [6] thus:

Definition 4. The mapping $f : E \to F$ is *fuzzy differentiable at a point* $x \in E$ if there exists a linear fuzzy continuous mapping $f'(x)$ of E into F such that

$$f(x+y) = f(x) + f'(x)(y) + \phi(y), \quad \text{for all } y \in E\,,$$

where ϕ is tangent to 0.

The mapping $f'(x)$ is the *fuzzy derivative* of f at x; it is an element of $L(E, F)$, the set of all linear fuzzy continuous mappings of E into F. The mapping f is *fuzzy differentiable* if it is fuzzy differentiable at every point of E. That $f'(x)$ is unique depends [6] on the fuzzy topology being fuzzy T_1.

An application of Definition 4 might be to those image characterizations which associate with each image point $\mathbf{x}$ in $\mathbb{R}^2$, say, a fuzzy estimate of location (Sect. 4), and with each point $f(\mathbf{x})$ in $\mathbb{R}$, say, a fuzzy estimate of an attribute value such as contour curvature.

The next section considers a generalization of this notion of differentiation to spaces which are only locally like fuzzy topological vector spaces.

8 Fuzzy Differentiation Between Fuzzy Manifolds

Let E, F, G be ftvs's. It may be shown [6] that the composition $g \circ f$ of two fuzzy differentiable mappings $f : E \to F$, $g : F \to G$ is fuzzy differentiable, and that the fuzzy derivative of $g \circ f$ at $x \in E$ is $g'(f(x)) \circ f'(x)$. It may also be shown [6] that if f, g are two fuzzy continuous mappings of E into F that are each fuzzy differentiable at $x \in E$, then $f + g$ is fuzzy differentiable and so is αf for all $\alpha \in \mathbb{K}$. A bijection f of E onto F is a *fuzzy diffeomorphism of class* C^1 if f and its inverse f^{-1} are fuzzy differentiable, and f' and $(f^{-1})'$ are fuzzy continuous.

Classically, one can glue together the open subsets of a topological vector space (more commonly a Banach space) to form a manifold. Fuzzy differentiable manifolds can be defined in the same way; the glue is a family of (local) fuzzy diffeomorphisms between fuzzy topological vector spaces.

Let X be a set. A *fuzzy atlas* $\mathcal{A}$ *of class* C^1 on X is a collection of pairs (A_j, ϕ_j) (here and subsequently j ranges in some index set) that satisfies the following conditions:

1. Each A_j is a fuzzy set in X and $\sup_j A_j(x) = 1$, for all $x \in X$.
2. Each ϕ_j is a bijection, defined on the support of A_j, which maps A_j onto an open fuzzy set $\phi_j[A_j]$ in some ftvs E_j, and, for each l in the index set, $\phi_j[A_j \cap A_l]$ is an open fuzzy set in E_j.
3. For each l in the index set, the mapping $\phi_l \circ \phi_j^{-1}$, which maps $\phi_j[A_j \cap A_l]$ onto $\phi_l[A_j \cap A_l]$, is a C^1 fuzzy diffeomorphism.

Each pair (A_j, ϕ_j) is a *fuzzy chart* of the fuzzy atlas. If a point $x \in X$ lies in the support of A_j then (A_j, ϕ_j) is a *fuzzy chart at* x.

It is then possible to show [8] that given a C^1 fuzzy atlas $\mathcal{A}$ on a set X, the set X may be endowed with a fuzzy topology such that each A_j in $\mathcal{A}$ is an open fuzzy set and each ϕ_j is fuzzy continuous. In fact, the family $\{A_j\}$ of fuzzy sets forms a base for a proper fuzzy topology on X and in this topology the ϕ_j are fuzzy continuous.

Let $(X, \mathcal{T})$ be a fuzzy topological space. Suppose that A is an open fuzzy set in X and that ϕ is a fuzzy continuous bijective mapping which is defined on the support of A and which maps A onto an open fuzzy set V in some ftvs E. The pair (A, ϕ) is *compatible* with the C^1 atlas $\{(A_j, \phi_j)\}$ if each mapping $\phi_j \circ \phi^{-1}$ of $\phi[A \cap A_j]$ onto $\phi_j[A \cap A_j]$ is a fuzzy diffeomorphism of class C^1. Two C^1 fuzzy atlases are compatible if each fuzzy chart of one atlas is compatible with each fuzzy chart of the other atlas. Compatibility between C^1 fuzzy atlases is obviously an equivalence relation. An equivalence class of C^1 fuzzy atlases on X defines a C^1 *fuzzy manifold* on X. In the following, reference is made simply to fuzzy manifolds.

Suppose that X, Y are fuzzy manifolds and that f is a mapping of X into Y. The fuzzy differentiability of f at a point x in X may be defined [8] by its fuzzy differentiability in fuzzy charts at x and $f(x)$; that is:

Definition 5. The mapping $f : X \to Y$ is *fuzzy differentiable at a point* $x \in X$ if there is a fuzzy chart (U, ϕ) at $x \in X$ and a fuzzy chart (V, ψ) at $f(x) \in Y$ such that the mapping $\psi \circ f \circ \phi^{-1}$, which maps $\phi[U \cap f^{-1}[V]]$ into $\psi[V]$, is fuzzy differentiable at $\phi(x)$.

It is obvious that this definition does not depend on the choice of fuzzy chart at x and $f(x)$. The mapping f is *fuzzy differentiable* if it is fuzzy differentiable at every point of X; it is a C^1 *fuzzy diffeomorphism* if it is a bijection and both it and its inverse f^{-1} are fuzzy differentiable.

Let X, Y, Z be fuzzy manifolds. The composition $g \circ f$ of two fuzzy differentiable mappings $f : X \to Y$, $g : Y \to Z$ is fuzzy differentiable, and, as a

corollary, if f, g are C^1 fuzzy diffeomorphisms, then the composition $g \circ f$ is a C^1 fuzzy diffeomorphism [8].

9 Tangent Vectors in a Fuzzy Manifold

The notion of a directional derivative in Euclidean (or affine) space leads to the classical notion of a tangent vector of a differentiable manifold. A tangent vector of a fuzzy manifold may be defined as follows. Let X be a fuzzy manifold and let x be a (classical) point in X. Consider triples (U, ϕ, v_λ), where (U, ϕ) is a fuzzy chart at x and v_λ is a fuzzy point of the ftvs in which $\phi[U]$ lies. Two such triples (U, ϕ, v_λ), (V, ψ, w_λ) are *related*, written $(U, \phi, v_\lambda) \sim (V, \psi, w_\lambda)$, if the fuzzy derivative of $\psi \circ \phi^{-1}$ at $\phi(x)$ maps v_λ into w_λ; that is,

$$(\psi \circ \phi^{-1})'(\phi(x))v_\lambda = w_\lambda .$$

It is straightforward to show that the relation $(U, \phi, v_\lambda) \sim (V, \psi, w_\lambda)$ is an equivalence relation. The equivalence class of triples (U, ϕ, v_λ) constitutes a *tangent vector* of the fuzzy manifold X at x. The *tangent space* $T_x(X)$ at x is the set of all tangent vectors at x.

The set $T_x(X)$ can be given the structure of a vector space. Define the *sum* of two tangent vectors at $x \in X$ as

$$(U_1, \phi_1, v_{1\lambda}) + (U_2, \phi_2, v_{2\lambda}) = (U_2, \phi_2, (\phi_2 \circ \phi_1^{-1})'(\phi_1(x))v_{1\lambda} + v_{2\lambda}) ;$$

and the *product* of a tangent vector with a scalar α as

$$\alpha \cdot (U, \phi, v_\lambda) = (U, \phi, \alpha v_\lambda) .$$

These two operations do not depend on the choice of fuzzy chart [8]; thus if $(U_1, \phi_1, v_{1\lambda}) \sim (V_1, \psi_1, w_{1\lambda})$ and $(U_2, \phi_2, v_{2\gamma}) \sim (V_2, \psi_2, w_{2\gamma})$, then $(U_1, \phi_1, v_{1\lambda}) + (U_2, \phi_2, v_{2\gamma}) \sim (V_1, \psi_1, w_{1\lambda}) + (V_2, \psi_2, w_{2\gamma})$; and if $(U, \phi, v_\lambda) \sim (V, \psi, w_\lambda)$, then $\alpha \cdot (U, \phi, v_\lambda) \sim \alpha \cdot (V, \psi, w_\lambda)$.

10 Conclusion

Of the possible approaches to defining differential operators in ill-specified environments, the four considered here vary, necessarily, in the directness of their application to image representations. The definitions of differentiation based on convolving image luminance distributions have an immediate applicability, but they may be less suited to the analysis of higher-level image representations. The definitions of differentiation based on fuzzy sets make weaker assumptions about the nature of image representations and the extent of the analytic machinery available; but, for practical applications, they require the construction of an explicit relationship between the physically measurable properties of images and the fuzzy sets that, at some processing level, represent them.

The last issue may be addressed with the aid of a *fuzzy location*; that is, the kind of fuzzy set that, as introduced in Sect. 4, associates with each point $\mathbf{x}$ in $\mathbb{R}^2$

with luminance $I(\mathbf{x})$ a measure of the adequacy of that characterization of the image at that point. At least one experimental procedure has been described [2] for estimating the reliability of visual positional sense, and this procedure could be used to determine a fuzzy location. Based on the notion of fuzzy location and fuzzy orientation, the elements of a fuzzy geometry for visual space have been set out in [7] (see also [39]), where the notions of fuzzy locations for lines and curves have been introduced, and some of the fuzzy relations among them, including fuzzy collinearity, straightness, and tangency.

References

1. Attneave, F. (1954). Some informational aspects of visual perception, Psychological Review 61, pp. 183–193.
2. Attneave, F. (1955). Perception of place in a circular field, American Journal of Psychology 68, pp. 69–82.
3. Banks, H.T., Jacobs, M.Q. (1970). A differential calculus for multifunctions, Journal of Mathematical Analysis and Applications 29, pp. 246–272.
4. Berkmann, J., Caelli, T. (1993) On the relationship between surface covariance and differential geometry, this volume, pp. 343–352.
5. Chang, C.L. (1968). Fuzzy topological spaces, Journal of Mathematical Analysis and Applications 24, pp. 182–190.
6. Ferraro, M., Foster, D.H. (1987). Differentiation of fuzzy continuous mappings on fuzzy topological vector spaces, Journal of Mathematical Analysis and Applications 121, pp. 589–601.
7. Ferraro, M., Foster, D.H. (1993). Elements of a fuzzy geometry for visual space, this volume, pp. 333–342.
8. Ferraro, M., Foster, D.H. (1993). C^1 fuzzy manifolds, Fuzzy Sets and Systems 54, pp. 99–106.
9. Fischler, M.A., Bolles, R.C. (1986). Perceptual organization and curve partitioning, IEEE Transactions on Pattern Analysis and Machine Intelligence 8, pp. 100–105.
10. Florack, L.M.J., ter Haar Romeny, B.M., Viergever, M.A., Koenderink, J.J. (1993). Images: regular tempered distributions, this volume, pp. 651–659.
11. Foster, D.H. (1975). An approach to the analysis of the underlying structure of visual space using a generalized notion of visual pattern recognition, Biological Cybernetics 17, pp. 77–79.
12. Foster, D.H. (1979). Fuzzy topological groups, Journal of Mathematical Analysis and Applications 67, pp. 549–564.
13. Foster, D.H. (1980). A description of discrete internal representation schemes for visual pattern discrimination, Biological Cybernetics 38, pp. 151–157.
14. Foster, D.H. (1991). Operating on spatial relations. In: Watt, R.J. (ed.), Pattern Recognition by Man and Machine, Macmillan, Basingstoke, Hampshire, pp. 50–68.
15. Goguen, J.A. (1967). L-fuzzy sets, Journal of Mathematical Analysis and Applications 18, pp. 145–174.
16. Hastie, T.J., Tibshirani, R.J. (1990). Generalized Additive Models, Chapman & Hall, London.
17. Hoffman, D.D., Richards, W.A. (1984). Parts of recognition, Cognition 18, pp. 65–96.

18. Katsaras, A.K., Liu, D.B. (1977). Fuzzy vector spaces and fuzzy topological vector spaces, Journal of Mathematical Analysis and Applications 58, pp. 135–146.
19. Khalimsky, E.D. (1986). Pattern analysis of N-dimensional digital images, Proceedings 1986 IEEE International Conference on Systems, Man, and Cybernetics, Atlanta, GA, pp. 1559–1562.
20. Kock, A. (1981). Synthetic Differential Geometry (London Mathematical Society Lecture Note Series 51), Cambridge University Press, Cambridge.
21. Koenderink, J.J., van Doorn, A.J. (1987). Representation of local geometry in the visual system, Biological Cybernetics 55, pp. 367–375.
22. Kong, T.Y. Rosenfeld, A. (1991). Digital topology: a comparison of the graph-based and topological approaches. In: Reed, G.M., Roscoe, A.W., Wachter, R.F. (eds.), Topology and Category Theory in Computer Science, Clarendon Press, Oxford, pp. 273–289.
23. Kopperman, R. (1993). The Khalimsky line as a foundation for digital topology, this volume, pp. 3–20.
24. Kovalevsky, V.A. (1993). Topological foundations of shape analysis, this volume, pp. 21–36.
25. Lang, S. (1967). Introduction to Differentiable Manifolds, Interscience (John Wiley and Sons), New York.
26. Liang, P., Todhunter, J.S. (1990). Representation and recognition of surface shapes in range images: a differential geometry approach, Computer Vision, Graphics, and Image Processing 52, pp. 78–109.
27. Lindeberg, T.P. (1990). Scale-space for discrete signals, IEEE Transactions on Pattern Analysis and Machine Intelligence 12, pp. 234–254.
28. Lindeberg, T. (1992). Discrete derivative approximations with scale-space properties: a basis for low-level feature extraction, Technical Report TRITA-NA-P9212 Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, pp. 1–53.
29. Lowen, R. (1976). Fuzzy topological spaces and fuzzy compactness, Journal of Mathematical Analysis and Applications 56, pp. 621–633.
30. Mizumoto, M., Tanaka, K. (1976). Some properties of fuzzy sets of type 2, Information and Control 31, 312–340.
31. Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes, Annals of Statistics 12, pp. 766–774.
32. Noest, A.J. (1993). Neural processing of overlapping shapes, this volume, pp. 383–392.
33. Novák, V. (1989). Fuzzy Sets and their Applications, Adam Hilger, Bristol.
34. Pedrycz, W. (1989). Fuzzy Control and Fuzzy Systems, Research Studies Press, Taunton, Somerset.
35. Pu, P.-M., Liu, Y.-M. (1980). Fuzzy topology. I. Neighborhood structure of a fuzzy point and Moore-Smith convergence, Journal of Mathematical Analysis and Applications 76, pp. 571–599.
36. Puri, M.L., Ralescu, D.A. (1983). Differentials of fuzzy functions, Journal of Mathematical Analysis and Applications 91, pp. 552–558.
37. Rådström, H. (1952). An embedding theorem for spaces of convex sets, Proceedings of the American Mathematical Society 3, pp. 165–169.
38. Richards, W., Dawson, B., Whittington, D. (1986). Encoding contour shape by curvature extrema, Journal of the Optical Society of America A 3, pp. 1483–1491.

39. Rosenfeld, A. (1992). Fuzzy geometry: an overview, IEEE International Conference on Fuzzy Systems, San Diego, CA, pp. 113–117.
40. Russell, B. (1923). Vagueness, Australasian Journal of Psychology and Philosophy 1, p. 84–92.
41. Schwartz, L. (1957, 1959). Théorie des Distributions, Vols I–II, Hermann, Paris.
42. Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.
43. Weiss, I. (1992). Noise resistant invariants of curves. In: Mundy, J.L., Zisserman, A. (eds.), Geometric Invariance in Computer Vision, MIT Press, Cambridge, MA, pp. 135–156.
44. Zadeh, L.A. (1965). Fuzzy sets, Information and Control 8, pp. 338–353.
45. Zadeh, L.A. (1968). Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23, pp 421–427.
46. Zadeh, L.A. (1974). Fuzzy logic and its application to approximate reasoning, Information Processing 74, pp. 591–594.

Elements of a Fuzzy Geometry for Visual Space*

Mario Ferraro[1] *and David H. Foster*[2]

[1] Dipartimento di Fisica Sperimentale, Università di Torino, via Giuria 1, 10125 Torino, Italy

[2] Department of Communication and Neuroscience, Keele University, Staffordshire ST5 5BG, UK

Abstract. This study introduces the notions of *fuzzy location* and *fuzzy proximity* to capture the imprecision associated with judgements of absolute and relative visual position. These notions are used to establish the elements of a fuzzy geometry for visual space, including the *fuzzy betweenness* of points, the *fuzzy orientation* of a pair of points, and the *fuzzy collinearity* of three or more points. Fuzzy orientation and fuzzy collinearity are, in turn, used to define the *fuzzy straightness* of a curve and the *fuzzy tangency* of two curves.

Keywords: shape description, differential geometry, fuzzy topology, fuzzy location, proximity, orientation, collinearity, tangency.

1 Introduction

Any description of perceived visual shape is based on certain assumptions concerning the topology and geometry of visual space and the parts of the shape under consideration. For example, the representation of an image by a scalar field $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(\mathbf{x})$ is the light intensity at the point $\mathbf{x} \in \mathbb{R}^2$, assumes that visual space is a manifold, usually smooth, and that $(\mathbf{x}, f(\mathbf{x}))$, $\mathbf{x} \in \mathbb{R}^2$, is a surface, namely a Monge patch, the characteristics of which can be analysed by geometrical methods. Other types of visual representations, oriented more towards graph-theoretic methods, assume that shapes can be partitioned into elementary geometrical components, such as points and lines, which are connected by certain geometrical relations [5].

Although classical topology and geometry provide powerful tools for investigating shape, they fail, by definition, to acknowledge that visual space is not an abstract space and that its properties are determined by the processes that lead to perception. Any visual measurement—that is, any operation performed to estimate the attributes of an image—is affected by imprecision that arises

* We are grateful to V.A. Kovalevsky for helpful comment, and to P. Fletcher and S.R. Pratt for critical review of the manuscript. This work was supported by the Consiglio Nazionale delle Ricerche and ESPRIT Basic Research Action No. 6448 (VIVA).

from various sources. Thus, if an attribute has numerical values, the accuracy with which these values can be estimated on some absolute scale is limited, by noise and by quantization errors; and, whether or not an attribute has numerical values, the labels used by human observers to characterize those values may still be vague [13]; consider, for example, the notion of the "nearness" of two objects.

This study uses the theory of fuzzy sets [14] as a basis for a more appropriate approach to the geometry of visual space in that it addresses directly the imprecision associated with visual measurements. The structure of fuzzy sets is poorer than that of classical sets since the law of the excluded middle does not hold [10] (see comment after Definition 3). The theory of fuzzy sets has sometimes been interpreted as a part or reformulation of probability theory, but the two theories are distinct, philosophically and operationally. Discussion of related issues can be found in [15, 10, 11], and a review of some other fuzzy geometrical concepts in [12].

2 Fuzzy Sets and Fuzzy Topologies

This section reviews, briefly, some of the basic properties of fuzzy sets. Let X be a set. Any subset W of X has associated with it a *characteristic function* $\chi_W : X \to \{0,1\}$, where $\chi_W(x) = 1$ if $x \in W$ and $\chi_W(x) = 0$ if $x \notin W$. This definition may be generalized to form the notion of a "fuzzy set", which associates with each point $x \in X$ a "grade of membership", usually taken in the unit interval $[0,1]$. Thus a *fuzzy set* A in a set X is a mapping $A : X \to [0,1]$ such that $A(x)$ is the grade of membership of x in A. The grade of membership may be taken in a lattice [6] rather than the interval $[0,1]$.

Definition 1. Let A be a fuzzy set in X. The *support* of A is the classical set $\mathrm{Supp}(A) = \{x \mid A(x) > 0\}$; the *$\alpha$-level* of A for a given $\alpha \in [0,1]$ is the classical set $A^\alpha = \{x \mid A(x) = \alpha\}$; and the *$\alpha$-cut* of A for a given $\alpha \in [0,1]$ is the classical set $A_\alpha = \{x \mid A(x) \geq \alpha\}$.

The following proposition presents a different view of fuzzy sets, namely, as a sequence of classical sets A_α for $\alpha \in [0,1]$.

Proposition 2. *Let A be a fuzzy set. Then*

$$A(x) = \sup_{\alpha \in [0,1]} \{\alpha \mid x \in A_\alpha\}.$$

Proof. See [10].

Relations among fuzzy sets such as equality or inclusion, and operations such as union, intersection, or complement can be defined naturally for fuzzy sets by generalization of the classical definitions. The following summarizes the basic definitions, for the sake of completeness.

Definition 3. Let A, B, C be fuzzy sets in a set X. Then

$$A = B \quad \text{if and only if} \quad A(x) = B(x) \text{ for all } x \in X;$$

$$A \subset B \quad \text{if and only if} \quad A(x) \leq B(x) \text{ for all } x \in X;$$

$$C = A \cup B \quad \text{if and only if} \quad C(x) = \max\{A(x), B(x)\}, \text{ for all } x \in X;$$

$$C = A \cap B \quad \text{if and only if} \quad C(x) = \min\{A(x), B(x)\}, \text{ for all } x \in X;$$

and the complement $\bar{A}$ of A is given by

$$B = \bar{A} \quad \text{if and only if} \quad B(x) = 1 - A(x) \text{ for all } x \in X.$$

More generally, for an arbitrary family $\{A_j\}_{j \in J}$ of fuzzy sets, the union $C = \bigcup_{j \in J} A_j$ and the intersection $D = \bigcap_{j \in J} A_j$ are defined by $C(x) = \sup_{j \in J} A_j(x)$, for all $x \in X$, and $D(x) = \inf_{j \in J} A_j(x)$, for all $x \in X$.

It is easy to verify that if membership functions are replaced by characteristic functions the classical definitions result.

For each $c \in [0, 1]$, denote by k_c the fuzzy set in X with membership function $k_c(x) = c$, for all $x \in X$. The fuzzy set k_1 corresponds to the set X and k_0 to the empty set $\emptyset$. Notice that in general for a fuzzy set A the intersection $A \cap \bar{A} \neq k_0$ and the union $A \cup \bar{A} \neq k_1$.

Definition 4. Let f be a mapping from a set X to a set Y. Let B be a fuzzy set in Y. Then the *inverse image* $f^{-1}[B]$ of B is the fuzzy set in X given by

$$f^{-1}[B](x) = B(f(x)), \text{ for all } x \in X.$$

Conversely, let A be a fuzzy set in X. The *image* $f[A]$ of A is the fuzzy set in Y given by

$$f[A](y) = \begin{cases} \sup_{z \in f^{-1}(y)} A(z), & \text{if } f^{-1}(y) \text{ is nonempty,} \\ 0, & \text{otherwise,} \end{cases}$$

where $f^{-1}(y) = \{x \mid f(x) = y\}$.

Definition 5. Let E be a linear space. A fuzzy set A in E is *convex* if

$$A(\lambda x + (1 - \lambda)y) \geq \min\{A(x), A(y)\},$$

for all $x, y \in E$ and $0 \leq \lambda \leq 1$.

Proposition 6. *A fuzzy set is convex if and only if all its α-cuts are (classical) convex sets.*

Proof. See [10].

Given a family of fuzzy sets it is possible to define a fuzzy topology that is a natural generalization of the classical definition.

Definition 7. A *fuzzy topology* on a set X is a family $\mathcal{T}$ of fuzzy sets that satisfies the following conditions [2]:

1. $k_0, k_1 \in \mathcal{T}$.
2. If $A, B \in \mathcal{T}$, then $A \cap B \in \mathcal{T}$.
3. If $A_j \in \mathcal{T}$ for all $j \in J$ (J some index set), then $\bigcup_{j \in J} A_j \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is called a *fuzzy topological space*, or *fts* for short, and the members of $\mathcal{T}$ are called *open fuzzy sets*. Definition 7 is not completely satisfactory—for instance, it fails to make constant functions between fts's fuzzy continuous—and an alternative definition [7] has been proposed in which condition (1) is replaced by

1′. For all $c \in [0,1]$, $k_c \in \mathcal{T}$.

A fuzzy topology that satisfies condition (1′) is referred to as a *proper* fuzzy topology [3].

Definition 8. A subfamily $\mathcal{B}$ of $\mathcal{T}$ is a *basis* for a fuzzy topology $\mathcal{T}$ if each member of $\mathcal{T}$ can be expressed as the union of members of $\mathcal{B}$.

Proposition 9. *A family $\mathcal{B}$ of fuzzy sets in X is a basis for a proper fuzzy topology on X if it satisfies the following conditions:*

1. $\sup_{B \in \mathcal{B}} B(x) = 1$, *for every* $x \in X$.
2. *If* $B_1, B_2 \in \mathcal{B}$, *then* $B_1 \cap B_2 \in \mathcal{B}$.
3. *For every* $B \in \mathcal{B}$ *and* $c \in [0,1]$, $B \cap k_c \in \mathcal{B}$.

Proof. Let $\mathcal{T}(\mathcal{B})$, or simply $\mathcal{T}$, be the family of fuzzy sets that can each be expressed as a union of elements of $\mathcal{B}$. From condition (1), $k_1 \in \mathcal{T}$, and it is obvious that if $A_j \in \mathcal{T}$ for all $j \in J$ (J some index set), then $\bigcup_{j \in J} A_j \in \mathcal{T}$. Let $\{B_j\}$ and $\{B_l\}$ be subfamilies of $\mathcal{B}$ (j and l ranging in index sets J and L respectively) and let $A = \bigcup_{j \in J} B_j$ and $C = \bigcup_{l \in L} B_l$. Then, for each $x \in X$, $\min\{A(x), C(x)\} = \min\{\sup_j B_j(x), \sup_l B_l(x)\} = \sup_{j,l}\{\min\{B_j(x), B_l(x)\}\}$. Thus if $A, C \in \mathcal{T}$, then $A \cap C \in \mathcal{T}$. Finally, it is necessary to show that k_c belongs to $\mathcal{T}$ for every c, $0 \leq c < 1$. Condition (3) implies that for each such c, the fuzzy set with membership function $\sup_{B \in \mathcal{B}}\{\min\{B(x), c\}\}$, $x \in X$, belongs to $\mathcal{T}$. By condition (1) there exists, for each $x \in X$ and c, $0 \leq c < 1$, a fuzzy set $B \in \mathcal{B}$ with grade of membership $B(x) \geq c$; hence $\sup_{B \in \mathcal{B}}\{\min\{B(x), c\}\} = c$, which shows that $k_c \in \mathcal{T}$. Thus the family generated by unions of $B \in \mathcal{B}$ is a proper fuzzy topology. □

Notice that in a basis for an improper fuzzy topology, condition (3) is unnecessary.

3 Fuzzy Locations and Fuzzy Proximities

Consider a "physical point" p, for instance, a tiny spot of light, that has position x in the space $\mathbb{R}^2$. If an observer attempts to locate p visually he or she obtains

an estimate that is imprecise, for the reasons mentioned in the Introduction; in addition, each such measurement depends on the experimental conditions under which the determination is made. The effects of this imprecision can be modelled by assuming that, for a given experimental condition, the point p is associated with a fuzzy set, thus:

Definition 10. A *fuzzy location* of a physical point p is a fuzzy set $P_p : \mathbb{R}^2 \to [0,1]$. The family of all fuzzy locations of the point p at a given position $\mathbf{x}$ in $\mathbb{R}^2$ is denoted by $\mathcal{P}_\mathbf{x}$, and over all possible positions by $\mathcal{P}$; that is, $\mathcal{P} = \bigcup_{\mathbf{x}\in\mathbb{R}^2} \mathcal{P}_\mathbf{x}$. It is assumed that for every physical point p there exists $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\sup_{P_p\in\mathcal{P}_{\mathbf{x}_0}} P_p(\mathbf{x}_0) = 1$, and that $\sup_{P_p\in\mathcal{P}} P_p(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathbb{R}^2$; that is, the family of all fuzzy locations *covers* $\mathbb{R}^2$.

Sometimes an additional assumption is made; namely, that fuzzy locations are convex (Definition 5). Some relevant properties of convex fuzzy sets are given in the following propositions (where the subscript identifying the physical point has been omitted). Convex sets will be used in the next section in the development of the elements of a fuzzy geometry. The notation P_p for a fuzzy location and P_{α_0} for an α_0-cut of P should not be confused.

Proposition 11. *Let P be a convex fuzzy set and let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ be such that $P(\mathbf{x}_1) = P(\mathbf{x}_2) = \alpha_0$, where $0 < \alpha_0 \leq 1$. Then there exists a set $Y = \{\, y \mid y = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ 0 \leq \lambda \leq 1 \,\}$ such that $P(y) \geq \alpha_0$ for all $y \in Y$.*

Proof. Consider the α_0-cut P_{α_0} (Definition 1). It must be convex, by Proposition 6. Hence Y must be a subset of P_{α_0}, and the assertion follows. □

Proposition 12. *Let P be a convex fuzzy set and let $\alpha_0 = \sup_{\mathbf{x}\in\mathbb{R}^2} P(\mathbf{x})$, where $0 < \alpha_0 \leq 1$. Suppose that there exist $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ such that for every neighbourhood (in the standard topology on $\mathbb{R}^2$) A_1, A_2 of $\mathbf{x}_1, \mathbf{x}_2$, respectively,*

$$\sup_{\mathbf{x}\in A_1} P(\mathbf{x}) = \sup_{\mathbf{x}\in A_2} P(\mathbf{x}) = \alpha_0 .$$

Then there exists a set $Y = \{\, y \mid y = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ 0 \leq \lambda \leq 1 \,\}$ such that for every $y \in Y$ there is a neighbourhood A_y of y for which $\alpha_0 = \sup_{\mathbf{x}\in A_y} P(\mathbf{x})$.

Proof. Suppose that the statement of the proposition is false for some point $y \in Y$. Then there exists ε, $0 \leq \varepsilon < \alpha_0$, such that $\sup_{\mathbf{x}\in A_y} P(\mathbf{x}) < \alpha_0 - \varepsilon$ for every neighbourhood A_y of y. By hypothesis, there exist neighbourhoods A_1, A_2 of $\mathbf{x}_1, \mathbf{x}_2$, respectively, such that $P(\mathbf{z}_1) \geq \alpha_0 - \varepsilon$ and $P(\mathbf{z}_2) \geq \alpha_0 - \varepsilon$ for some $\mathbf{z}_1 \in A_1$ and $\mathbf{z}_2 \in A_2$. Choose λ_0, $0 \leq \lambda_0 \leq 1$, and y_0 belonging to a neighbourhood A_y of y such that $y_0 = \lambda_0\mathbf{z}_1 + (1-\lambda_0)\mathbf{z}_2$. Set $\alpha_1 = \min\{P(\mathbf{z}_1), P(\mathbf{z}_2)\}$ and consider the α_1-cut P_{α_1}. This set is not convex, which implies that P is not convex (compare Proposition 6), contrary to the hypothesis. □

Proposition 13. *Suppose that a convex fuzzy set P has a maximum value, α_0 say. Then the α_0-cut P_{α_0} is a point, or an interval of a line, or a convex subset of $\mathbb{R}^2$.*

Proof. Obvious, since P_{α_0} must be convex. □

Next, a fuzzy set is defined that determines the grade of proximity of any two physical points.

Definition 14. Given two physical points p, q, with fuzzy locations P_p, P_q respectively, the *fuzzy proximity* of p, q, denoted by $\delta(p, q)$, is given by the fuzzy set $P_p \cap P_q$. The two points p, q are said to be *fuzzy proximal* if $\delta(p, q) \neq k_0$; that is, there exists $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\delta(p, q)(\mathbf{x}_0) = \min\{P_p(\mathbf{x}_0), P_q(\mathbf{x}_0)\} > 0$.

Notice that if p, q are fuzzy proximal, then $\sup_{\mathbf{x}\in\mathbb{R}^2} \min\{P_p(\mathbf{x}), P_q(\mathbf{x})\} > 0$. The fuzzy set $\delta(p, q)$ can be thought of as quantifying the vague description "near to". This definition extends naturally to pairs of any (not necessarily finite) number of points: if $\{p_j\}_{j\in J}, \{p_l\}_{l\in L}$ are two sets of physical points, then, with an abuse of notation, their fuzzy proximity is given by

$$\delta(\{p_j\}, \{p_l\}) = \left(\bigcup_{j\in J} P_{p_j}\right) \cap \left(\bigcup_{l\in L} P_{p_l}\right).$$

A fuzzy proximity δ for physical points and their fuzzy locations satisfies conditions analogous to those characterizing a proximity for classical sets, except for a separation condition (see [9]); thus:

Proposition 15. *Let p, q, r be physical points. Then:*

1. $\delta(p, q) = \delta(q, p)$.
2. $\delta(p, q) \neq k_0$ *implies* $P_p \neq k_0$ *and* $P_q \neq k_0$, *where* P_p, P_q *are the fuzzy locations of* p, q *respectively.*
3. $\delta(\{p, q\}, r) \neq k_0$ *if and only if* $\delta(p, r) \neq k_0$ *or* $\delta(q, r) \neq k_0$.

Proof. Statements (1) and (2) are obviously true. To prove statement (3), consider the following. Let P_r be the fuzzy location of r. Suppose that there exists a point $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\min\{\max\{P_p(\mathbf{x}_0), P_q(\mathbf{x}_0)\}, P_r(\mathbf{x}_0)\} > 0$ and $\min\{P_p(\mathbf{x}_0), P_r(\mathbf{x}_0)\} = 0$ and $\min\{P_q(\mathbf{x}_0), P_r(\mathbf{x}_0)\} = 0$. Then $\min\{\max\{P_p(\mathbf{x}_0), P_q(\mathbf{x}_0)\}, P_r(\mathbf{x}_0)\} = 0$, contrary to the hypothesis. Conversely, suppose that there exists a point $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\min\{P_p(\mathbf{x}_0), P_r(\mathbf{x}_0)\} > 0$ or $\min\{P_q(\mathbf{x}_0), P_r(\mathbf{x}_0)\} > 0$. Then $\min\{\max\{P_p(\mathbf{x}_0), P_q(\mathbf{x}_0)\}, P_r(\mathbf{x}_0)\} \geq \min\{P_p(\mathbf{x}_0), P_r(\mathbf{x}_0)\} > 0$, or $\min\{\max\{P_p(\mathbf{x}_0), P_q(\mathbf{x}_0)\}, P_r(\mathbf{x}_0)\} \geq \min\{P_q(\mathbf{x}_0), P_r(\mathbf{x}_0)\} > 0$. □

The form of the fuzzy locations P_p, P_q determines the form of $\delta(p, q)$, as follows.

Proposition 16. *If fuzzy locations P_p, P_q are convex fuzzy sets, then $\delta(p, q)$ is a convex fuzzy set.*

Proof. It is enough to recall that the intersection of two convex classical sets is convex and the result follows from Proposition 6. □

Notice that it is possible to define the fuzzy proximity of an arbitrary, finite number of points, $p_1, p_2, \ldots, p_n$:

$$\delta(p_1, p_2, \ldots, p_n) = \bigcap_{i=1}^{n} P_{p_i},$$

and if all the P_{p_i} are convex fuzzy sets, then $\delta(p_1, p_2, \ldots, p_n)$ is a convex fuzzy set.

Although not developed here, it is easy to see that fuzzy proximities define a basis for an *improper* fuzzy topology on visual space.

Proposition 17. *The family of fuzzy sets formed by fuzzy proximities of the form $\delta(p_1, p_2, \ldots, p_n)$, for every finite integer n, is a basis for an improper fuzzy topology.*

Proof. Let $\mathcal{B}$ be the family of all fuzzy proximities $\delta(p_1, p_2, \ldots, p_n)$, n finite. For all $P_p \in \mathcal{P}$, the family of all fuzzy locations, $\delta(p, p) = P_p$ and hence $\mathcal{P} \subset \mathcal{B}$. Then condition (1) of Proposition 9 is satisfied, since $\sup_{P_p \in \mathcal{P}} P_p(x) = 1$ for all $x \in \mathbb{R}^2$ (Definition 10). Next, given two fuzzy proximities $\delta(p_1, p_2, \ldots, p_m)$, $\delta(p_{m+1}, p_{m+2}, \ldots, p_{m+n})$, their intersection

$$\delta(p_1, p_2, \ldots, p_m) \cap \delta(p_{m+1}, p_{m+2}, \ldots, p_{m+n}) = \delta(p_1, p_2, \ldots, p_{m+n})$$

belongs to $\mathcal{B}$ and thus condition (2) of Proposition 9 holds. □

If fuzzy locations are assumed to be convex, then fuzzy locations and fuzzy proximities are open fuzzy sets of a proper fuzzy topology.

Proposition 18. *The family of all convex fuzzy sets defined in $\mathbb{R}^2$ is a basis for a proper fuzzy topology.*

Proof. Conditions (1) and (2) of Proposition 9 are obviously satisfied because the fuzzy set k_1 is convex and the intersection of convex fuzzy sets is a convex fuzzy set (see Proposition 16). To prove condition (3) it is enough to observe that the fuzzy sets k_c, $c \in [0, 1]$, are convex fuzzy sets. □

4 Elements of a Fuzzy Geometry

In this section, the notion of fuzzy location is extended to the notion of the fuzzy orientation of a pair of points and the fuzzy collinearity of three or more points. It is shown that these notions make it possible to define the fuzzy straightness of a curve and the fuzzy tangency of two curves. First, the notion of the fuzzy betweenness of points is introduced.

Definition 19. Let p, r be two fuzzy proximal physical points. A physical point q is *fuzzy between* p and r if $\delta(p, q) \supset \delta(p, r)$ and $\delta(r, q) \supset \delta(r, p)$.

Consider the set of all possible orientations θ, $0 \leq \theta < 2\pi$, in the plane.

Definition 20. For any pair of physically distinguishable points (p, q), the *fuzzy orientation* of (p, q) is a fuzzy set $O_{p,q} : [0, 2\pi) \to [0, 1]$. If the points are not physically distinguishable, then $O_{p,q} : [0, \pi) \to [0, 1]$.

It is, in principle, possible to derive a fuzzy orientation from the fuzzy location of two distinct physical points p, q. Let A_θ be the set of all pairs $(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in \text{Supp}(P_p)$, $\mathbf{y} \in \text{Supp}(P_q)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, for which the orientation of the line joining $\mathbf{x}$ and $\mathbf{y}$ is θ. The derived fuzzy orientation $O'_{p,q}$ of (p, q) is then defined as the fuzzy set

$$O'_{p,q}(\theta) = \sup_{(\mathbf{x},\mathbf{y}) \in A_\theta} \{\min\{P_p(\mathbf{x}), P_q(\mathbf{y})\}\}, \quad \text{for all } \theta \in [0, 2\pi).$$

Observed orientation estimates need not, however, follow such a rule.

The notion of fuzzy proximity may be extended to fuzzy orientations.

Definition 21. Let $(p, q), (r, s)$ be two pairs of physical points with fuzzy orientations $O_{p,q}, O_{r,s}$ respectively. Their *fuzzy proximity with respect to orientation*, denoted by $\gamma(p, q; r, s)$, is given by the fuzzy set $O_{p,q} \cap O_{r,s}$. The two pairs of points are said to be *fuzzy proximal with respect to orientation* if $\gamma(p, q; r, s) \neq k_0$.

Notice that $\gamma(p, q; r, s) = \gamma(r, s; p, q)$. Fuzzy proximity with respect to orientation makes it possible to define for three physical points their fuzzy collinearity.

Definition 22. Let p, q, r be physical points with fuzzy orientations $O_{p,q}$, $O_{q,r}$, $O_{p,r}$ taken a pair at a time. Then the *fuzzy collinearity* of p, q, r, denoted by $\eta(p, q, r)$, is given by the fuzzy set $O_{p,q} \cap O_{q,r} \cap O_{p,r}$. The points p, q, r are said to be *fuzzy collinear* if $\eta(p, q, r) \neq k_0$.

The definition may be extended to four or more points.

In the following, a physical *curve* in $\mathbb{R}^2$ is considered as the image of a mapping of an interval in $\mathbb{R}$ into $\mathbb{R}^2$ rather than as the mapping itself.

Definition 23. Let c be a physical curve in $\mathbb{R}^2$. The *fuzzy location* P_c of c is the union $\bigcup_{p \in c} P_p$ of the fuzzy locations P_p for all $p \in c$; that is, $P_c(\mathbf{x}) = \sup_{p \in c} P_p(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{R}^2$.

The fuzzy proximity $\delta(p, c)$ of a physical point p and a physical curve c is defined by the extension of Definition 14 as the intersection $P_p \cap P_c$ of their fuzzy locations P_p and P_c. If p and c are fuzzy proximal, then $P_p \cap P_c \neq k_0$; that is, there exists $\mathbf{x}_0 \in \mathbb{R}^2$ such that $\delta(p, c)(\mathbf{x}_0) = \min\{P_p(\mathbf{x}_0), \sup_{q \in c} P_q(\mathbf{x}_0)\} > 0$. It is easy to show that

$$\delta(p, c)(\mathbf{x}) = \min\left\{P_p(\mathbf{x}), \sup_{q \in c} P_q(\mathbf{x})\right\} = \sup_{q \in c}\{\min\{P_p(\mathbf{x}), P_q(\mathbf{x})\}\},$$

for all $\mathbf{x} \in \mathbb{R}^2$. That is, the fuzzy proximity $\delta(p, c)$ of a point and a curve is the union of the fuzzy proximities $\delta(p, q)$ of p and the points q belonging to c.

Proposition 24. *Let p be a physical point and c a physical curve. Then p and c are fuzzy proximal if and only if there exists a point q in c such that p and q are fuzzy proximal.*

Proof. Suppose that there exists no $q \in c$ such that $\delta(p,q) \neq k_0$; then $\delta(p,c)(\mathbf{x}) = \sup_{q \in c}\{\min\{P_p(\mathbf{x}), P_q(\mathbf{x})\}\} = \sup_{q \in c} \delta(p,q)(\mathbf{x}) = 0$, for all $\mathbf{x} \in \mathbf{R}^2$. Conversely, let $\mathbf{x}_0 \in \mathbf{R}^2$ be such that for some $q \in c$, $\delta(p,q)(\mathbf{x}_0) > 0$; then $\delta(p,c)(\mathbf{x}_0) = \sup_{q \in c} \delta(p,q)(\mathbf{x}_0) > 0$. □

Given a physical curve c and two fuzzy proximal physical points p, r, the curve is said to be *fuzzy between* the two points if there exists $q \in c$ such that $\delta(p,q) \supset \delta(p,r)$ and $\delta(r,q) \supset \delta(r,p)$. The fuzzy proximity $\delta(c,h)$ of two physical curves c and h is also defined by the extension of Definition 14; that is, as the intersection $P_c \cap P_h$ of their fuzzy locations P_c and P_h.

Proposition 25. *Two physical curves c, h are fuzzy proximal if and only if there exist points p in c and q in h such that p and q are fuzzy proximal.*

Proof. Omitted, since it is analogous to the proof of Proposition 24.

The definition of the fuzzy collinearity $\eta(p,q,r)$ of three points p, q, r (Definition 22) can be extended to define the fuzzy straightness of a curve.

Definition 26. Let l be a physical curve. The *fuzzy straightness* of l, denoted by $\eta(l)$, is given by the fuzzy set $\bigcap_{p,q,r \in l} \eta(p,q,r)$. The curve l is said to be *fuzzy straight* if $\eta(l) \neq k_0$.

Notice that the fuzzy straightness of a curve is simply the intersection of all fuzzy orientations $O_{p,q}$ for all p, q in the curve.

The definition of the fuzzy proximity with respect to orientation $\gamma(p,q;r,s)$ of two pairs of points $(p,q),(r,s)$ (Definition 21) leads to a definition of fuzzy tangency (a different approach to fuzzy tangency is discussed in [4, 3]).

Definition 27. Let c, h be two physical curves and suppose that there exists a point p that is fuzzy proximal to both c and h. Then c, h are *fuzzy tangent* at p if, for any two points $q \in c$, $r \in h$ that are each fuzzy proximal to but distinct from p, the pairs $(q,p),(r,p)$ are fuzzy proximal with respect to orientation; that is, $\gamma(q,p;r,p) \neq k_0$.

The notion of fuzzy betweenness can be extended to fuzzy orientations. Let $(p,q),(r,s),(u,v)$ be three pairs of distinct points. Then (r,s) is said to be fuzzy between (p,q) and (u,v) with respect to orientation if the fuzzy proximities $\gamma(p,q;r,s) \supset \gamma(p,q;u,v)$ and $\gamma(u,v;r,s) \supset \gamma(u,v;p,q)$.

Proposition 28. *Suppose that two physical curves c, h are fuzzy tangent at a point p, and a third curve g is fuzzy proximal to p. Suppose further that, for any three points q in c, r in g, s in h that are each fuzzy proximal to but distinct from p, (r,p) is fuzzy between (q,p) and (s,p) with respect to orientation. Then g is fuzzy tangent to both c and h at p.*

Proof. The fuzzy proximity with respect to orientation $\gamma(q,p;s,p) \neq k_0$, and both $\gamma(q,p;r,p) \supset \gamma(q,p;s,p)$ and $\gamma(s,p;r,p) \supset \gamma(s,p;q,p)$. □

5 Conclusion

The approach of this study to the geometry of visual space has been formal in that the construction of geometrical properties and relations was not founded on a particular set of empirical data. It is, however, possible to determine by experimental measurement—for a given observer and experimental paradigm—typical instances of fuzzy locations and fuzzy orientations, and typical instances of (in principle) dependent relations and properties such as fuzzy betweenness, fuzzy collinearity, fuzzy straightness, and fuzzy tangency. A possible experimental procedure for making these measurements has been described by Attneave [1]. This procedure could be used to generate examples of fuzzy locations, and extended to the generation of other properties and relations. Whether the results of such measurements can be related to each other according to the present analysis may offer a test of its physical appropriateness.

References

1. Attneave, F. (1955). Perception of place in a circular field, Am. Journal of Psychology 68, pp. 69–82.
2. Chang, C.L. (1968). Fuzzy topological spaces, Journal of Mathematical Analysis and Applications 24, pp. 182–190.
3. Ferraro, M., Foster, D.H. (1993). C^1 fuzzy manifolds, Fuzzy Sets and Systems 54, pp. 99–106.
4. Foster, D.H. (1993). Classical and fuzzy differential methods in shape analysis, this volume, pp. 319–332.
5. Foster, D.H. (1991). Operating on spatial relations. In: Watt, R.J. (ed.), Pattern Recognition by Man and Machine, Macmillan, Basingstoke, Hampshire, pp.50–68.
6. Goguen, J.A. (1967). L-fuzzy sets, Journal of Mathematical Analysis and Applications 18, pp. 145–174.
7. Lowen, R. (1976). Fuzzy topological spaces and fuzzy compactness, Journal of Mathematical Analysis and Applications 56, pp. 621–633.
8. Millman, R.S., Parker, G.D. (1991). Geometry. A Metric Approach with Models (2nd ed.), Springer-Verlag, New York.
9. Naimpally S.A., Warrack, B.D. (1970). Proximity Spaces, Cambridge University Press, Cambridge.
10. Novák, V. (1989). Fuzzy Sets and Their Applications, Adam Hilger, Bristol.
11. Pedrycz, W. (1989). Fuzzy Control and Fuzzy Systems, Research Studies Press, Taunton, Somerset.
12. Rosenfeld, A. (1992). Fuzzy geometry: an overview, IEEE Int. Conf. on Fuzzy Systems, San Diego, CA, pp. 113–117.
13. Russell, B. (1923). Vagueness, Australasian Journal of Psychology and Philosophy 1, pp. 84–92.
14. Zadeh, L.A. (1965). Fuzzy sets, Information and Control 8, pp. 338–353.
15. Zadeh, L.A. (1968). Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23, pp. 421–427.

On the Relationship Between Surface Covariance and Differential Geometry

Jens Berkmann and Terry Caelli *

Computer Science Department, The University of Melbourne Parkville, Vic. 3052, Australia

Abstract. In this paper the application of covariance techniques to surface representations (whether of range or intensity type) of 3-D objects is discussed and is compared to traditional methods using differential geometry. An analogous operator to the classical Weingarten map is defined and it is shown how this operator provides local invariant descriptors without using surface parameterizations or calculus.

Keywords: differential geometry, covariance, Weingarten map, second fundamental form.

1 Introduction

Differential geometry provides one of the more popular forms of surface representation in object recognition and, more recently, for the encoding of images in general [2, 1, 4, 5]. This is because differential geometry provides quantities (in particular Gaussian curvature) which are invariant under rigid motions. Making differential geometry applicable to discrete representations or sampled surfaces is usually done by smoothing with a Gaussian filter or regularizing by fitting at each surface pixel a bi-quadratic surface with respect to a support window. After fitting the surface the derivatives are usually computed from the fitted surface parameters or by using finite-difference operators directly on the filtered surface. The crucial point of all such techniques is their inevitable operation on an erroneous surface. Covariance techniques, on the other hand, provide invariant descriptors in terms of the eigenvalues of the covariance matrix, without using calculus or even the need for a consistent parameterization of the surface, but using methods of topology to orient the surface and infer surface shape. The aim of this paper is to investigate the properties of the covariance approach and to propose a definition of the Weingarten map on the basis of discrete geometry. However, before the covariance method is discussed, differential geometry will briefly be reviewed.

* This project was funded by grants from the Gottlieb Daimler and Karl Benz Foundation and the Australian Research Committee. Requests for reprints should be sent to Terry Caelli.

2 Differential Geometry

We consider the case of a surface in Monge patch representation (view-dependent range map) which is parameterized as:

$$\mathbf{x}(u,v) = u\mathbf{e}_1 + v\mathbf{e}_2 + z(u,v)\mathbf{e}_3 \ . \tag{1}$$

There are two basic mathematical entities that are considered in the classical analysis of surfaces: the first and second fundamental forms.

The first fundamental form I of a surface is defined by the following quadratic form:

$$I(du,dv) = d\mathbf{x}^T \cdot d\mathbf{x} = d\mathbf{u}^T \mathbf{A} d\mathbf{u} = Edu^2 + 2Fdudv + Gdv^2 \ , \tag{2}$$

where the elements of matrix $\mathbf{A}$ are defined as

$$E = \mathbf{x}_u^T \cdot \mathbf{x}_u \ , \qquad F = \mathbf{x}_u^T \cdot \mathbf{x}_v \ , \qquad G = \mathbf{x}_v^T \cdot \mathbf{x}_v \ . \tag{3}$$

The subscripts signify partial derivatives, du and dv are small elements in the u- and the v-direction respectively, $d\mathbf{x}$ denotes the total differential of the vector $\mathbf{x}$ along a chosen direction which is given by $d\mathbf{u} = (du\ dv)$ in the parameter space and $\mathbf{x}^T$ corresponds to the transpose of $\mathbf{x}$. The value I, at a particular point, is invariant to translations and rotations of the surface as well as to parameterization changes. The first fundamental form measures the metric of the surface and does not depend on how the surface is embedded in 3-D space.

The second fundamental form II is, in contrast, an extrinsic property of a surface because of its additional *consideration* of the normal vector. The second fundamental form is defined by the following quadratic form:

$$II(du,dv) = -d\mathbf{x}^T \cdot d\mathbf{n} = d\mathbf{u}^T \mathbf{B} d\mathbf{u} = Ldu^2 + 2Mdudv + Ndv^2 \ , \tag{4}$$

where the elements of matrix $\mathbf{B}$ are defined as

$$L = \mathbf{x}_{uu}^T \cdot \mathbf{n} \ , \qquad M = \mathbf{x}_{uv}^T \cdot \mathbf{n} \ , \qquad N = \mathbf{x}_{vv}^T \cdot \mathbf{n} \ , \tag{5}$$

with

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{|\mathbf{x}_u \times \mathbf{x}_v|} \tag{6}$$

as the unit normal vector.

An alternate formulation for the second fundamental form is as follows. Let P be the point on a surface and Q a point in the neighbourhood of P (Fig. 1). Let $s = \vec{PQ}^T \cdot \mathbf{n}$ be the projection of the vector from P to Q onto the unit normal vector $\mathbf{n}$ at P. So, s is the orthogonal distance from Q to the tangent plane. Now suppose that P and Q are determined by the vectors $\mathbf{x}(u,v)$ and $\mathbf{x}(u+du, v+dv)$ respectively. Applying Taylor's formula to

$$s = (\mathbf{x}(u+du, v+dv) - \mathbf{x}(u,v))^T \cdot \mathbf{n} \tag{7}$$

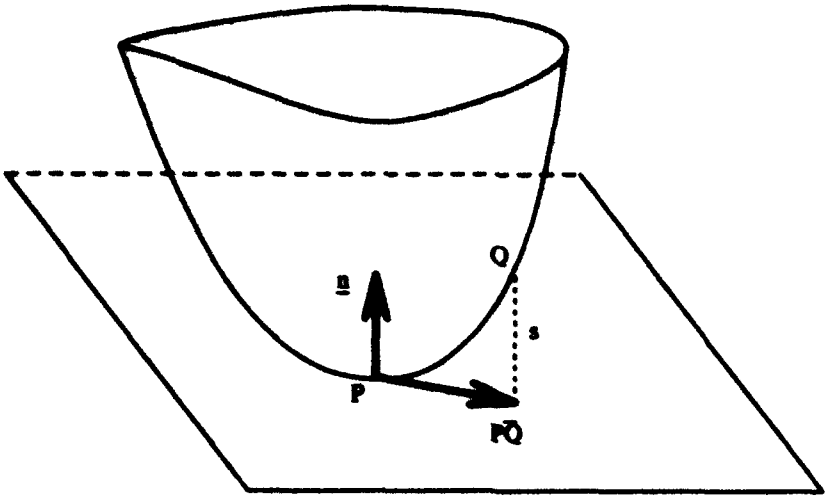


Fig. 1. Definition of the second fundamental form by the projection distance of Q onto the tangent plane at P with normal vector n.

results in

$$\begin{aligned} s &\approx (d\mathbf{x} + \frac{1}{2}d^2\mathbf{x})^T \cdot \mathbf{n} \\ &\approx \frac{1}{2}d^2\mathbf{x}^T \cdot \mathbf{n} = -\frac{1}{2}d\mathbf{x} \cdot d\mathbf{n} \ , \end{aligned} \tag{8}$$

since $d\mathbf{x}$ is parallel to the tangent plane (neglecting higher-order terms). Thus, the second fundamental form II is approximated by simply doubling the orthogonal deviation s from the tangent plane. The function $\frac{1}{2}II$ is also referred to as the osculating paraboloid. The behaviour of this paraboloid describes and characterizes the shape of the surface in a small neighbourhood and the ratio of $II(u,v)/I(u,v)$ is known as the normal curvature at a particular point (Do Carmo [3] and Lipschutz [7]).

Now, let us consider the Weingarten equations Weingarten map. They can be written in the following manner:

$$\begin{pmatrix} \mathbf{n}_u \\ \mathbf{n}_v \end{pmatrix} = -\mathbf{W} \begin{pmatrix} \mathbf{x}_u \\ \mathbf{x}_v \end{pmatrix} \ , \tag{9}$$

where $\mathbf{W} = \mathbf{A}^{-1} \cdot \mathbf{B}$ holds with the matrices $\mathbf{A}$ and $\mathbf{B}$ according to (2) and (4) respectively. These equations define the relationship between the rate of change of the unit normal vector and the corresponding chosen direction of a curve on the tangent plane. Equation (9) defines a mapping between tangent vectors because $d\mathbf{n} = \partial u/\partial t\, \mathbf{n}_u + \partial v/\partial t\, \mathbf{n}_v$ lies on the tangent plane as well as $d\mathbf{x} = \partial u/\partial t\, \mathbf{x}_u + \partial v/\partial t\, \mathbf{x}_v$. The eigenvalues of the matrix W are the principal curvatures (k_1, k_2) and the eigenvectors of W the principal directions $(\mathbf{v}_1, \mathbf{v}_2)$ at a particular point on a surface.

Principal curvatures and principal directions have played an important role in segmentation and feature extraction in the recent focus of object recognition from range data (see, for example, Fan *et al.* [4]). Unfortunately, such descriptors are subject to the limits of their numerical computation. For this reason unified surface geometry without using calculus, but equivalent to the Weingarten map representation, is desirable.

3 The Covariance Method

In the following section a definition of an operator is presented which can be seen as an analogy to the Weingarten map in (9) using covariance techniques.

Liang and Todhunter [6] introduced covariance methods in the calculation of surface normals and principal directions as a technique for orienting surface quadratics. The covariance matrix they used is defined by:

$$\mathbf{C}_I = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{x}_m)\cdot(\mathbf{x}_i - \mathbf{x}_m)^T \ , \tag{10}$$

where $\mathbf{x}_i = (x_i, y_i, z_i)^T$ correspond to the projection plane (x, y) and depth (z) values at position i; the $\mathbf{x}_m = 1/n\sum_{i=1}^{n}\mathbf{x}_i$ correspond to the mean position vector; and n is the total number of pixels used to compute (10).

This method corresponds to a least-squares planar fit of the surface data, since it defines the plane as the *tangent plane* which minimizes the deviation of the data points from the plane under orthogonal projection. It should also be noted that (10) does not depend on a *consistent* labelling of pixel neighbours, as the covariance involves *summation* over all such elements. In this sense the covariance method does not require a consistent surface parameterization. In [6] the tangent plane was originally defined by the eigenvectors corresponding to the two largest eigenvalues of $\mathbf{C}_I$. However, a pair of eigenvectors constitutes a tangent plane (analogous to a tangent plane) only if the order of surface points under orthogonal projection onto this plane is preserved (Fig. 2). If the pixels are ordered in a 3×3 window as follows:

1	2	3
4	5	6
7	8	9

a *wrong* tangent plane may, after projection of the pixels onto this plane, result in the following ordering:

1	2	3
7	8	9
4	5	6

which violates the *order-preserving principle*. This *order-preserving principle* can be seen as a replacement of surface parameterization, since it makes sure that the eigenvectors form analogues to the tangent vectors and the normal vector. It is important to note that such eigenvalues, capturing equivalent information to the first fundamental form, are already invariant to rigid motions of the surface.

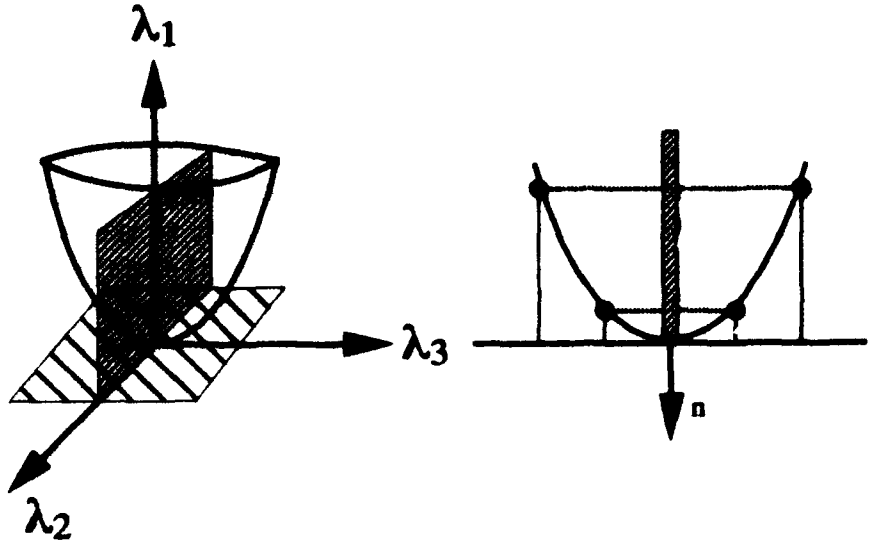


Fig. 2. (left) Surface with possible *tangent planes*. A pair of eigenvectors constitute a correct solution if points of the surface under orthogonal projection onto the plane preserve the order. (right) Cross-sectional view with accepted normal.

We then define, analogous to the Weingarten map about a point $\mathbf{x}_0$ on a surface, a two-dimensional covariance matrix in the following manner:

$$\mathbf{C}_{II} = \frac{1}{n-1}\sum_{i=1}^{n-1}(\mathbf{y}_i - \mathbf{y}_m)\cdot(\mathbf{y}_i - \mathbf{y}_m)^T \ , \tag{11}$$

with the two-dimensional vectors $\mathbf{y}_i$ defined by:

$$\mathbf{y}_i = s_i \cdot [\mathbf{x}_i - \mathbf{x}_0]_p \ , \tag{12}$$

where the $[\ldots]_p$ operator denotes projection onto the tangent plane, $n-1$ is the total number of surface points in the neighbourhood of the central point $\mathbf{x}_0$, and $\mathbf{y}_m = 1/(n-1)\sum_{i=1}^{n-1}\mathbf{y}_i$ corresponds to the mean vector of the data vectors used to compute (11). We project the difference vector which points from the central $\mathbf{x}_0$ to $\mathbf{x}_i$ onto the tangent plane and weight the resulting two-dimensional vector by distance s_i which measures the orthogonal distance from the tangent plane to the point $\mathbf{x}_i$. The analogous definitions for the tangent vectors $\mathbf{t}_1$ and

$\mathbf{t}_2$ as well as for the normal vector $\mathbf{n}$ at point $\mathbf{x}_0$ are obtained by assigning the eigenvectors of $\mathbf{C}_I$ according to (10) as discussed above. (12) may now be written more explicitly:

$$\mathbf{y}_i = \underbrace{\left[(\mathbf{x}_i - \mathbf{x}_0)^T \cdot \mathbf{n}\right]}_{s_i} \cdot \begin{pmatrix} (\mathbf{x}_i - \mathbf{x}_0)^T \cdot \mathbf{t}_1 \\ (\mathbf{x}_i - \mathbf{x}_0)^T \cdot \mathbf{t}_2 \end{pmatrix} . \tag{13}$$

After this we define the quadratic form

$$II_C = \mathbf{v}^T \cdot C_{II} \cdot \mathbf{v} \tag{14}$$

as a *second fundamental form* based on covariance methods with the covariance matrix defined according to (11) and a chosen unit vector $\mathbf{v}$ on the tangent plane, where now the matrix $\mathbf{C}_{II}$ plays an analogous role to the matrix $\mathbf{W}$ in (9), since for each chosen direction $\mathbf{v}$ (unit vector) on the tangent plane the value II_C provides a measure of the *ensemble deviation* of surface points in a small neighbourhood of a particular point. Analogous to classical computations of surface geometry, principal directions (directions of *extremal* curvature) are defined as those directions which are given by the eigenvectors $(\mathbf{v}_1^{II}, \mathbf{v}_2^{II})$ of this covariance matrix. Unfortunately, surface types can no longer be distinguished by the signs of the estimated principal curvatures, that is, the eigenvalues $(\lambda_1^{II}, \lambda_2^{II})$ of the covariance matrix defined above. Also the size of the eigenvalues does not help to classify surface types uniquely. Hence, a classification has to be sought which captures the surface type, independent of the principal curvatures as derived by this method.

This can be done via the relation of surface points to the tangent plane in a small neighbourhood of a particular point. A point is *elliptic* if all points in a certain region around the point lie on only one side of the estimated tangent plane $(\mathbf{t}_1, \mathbf{t}_2)$. In contrast a particular point is defined as *hyperbolic* if the points in a given region around the point lie on both sides of the tangent plane. *Planarity* can be detected by one zero (smaller than an appropriate threshold) eigenvalue of $\mathbf{C}_I$. The neighbourhood of a *parabolic* point has a line in common with the tangent plane. Hence, the detection of such points is achieved by observing the deviation s_i from the tangent plane in a window around the particular point as already computed in (12).

4 Discontinuities

Discontinuities present well-known problems for surface representations using differential geometry, since they violate the assumption of continuity of the surface and its first-order derivatives. Many reseachers have tried to detect such discontinuities by using various techniques so that classical surface computation can be applied to the remaining surface (without discontinuities). However, by using covariance methods discontinuities can be treated as natural parts of discrete surface representations. In order to do so we shall discuss the properties of

the eigenvalues and eigenvectors of the covariance matrix $\mathbf{C}_I$ in a neighbourhood of discontinuities as, for instance, with jumps and creases.

A jump is known as a discontinuity in the surface itself, whereas a crease is known as a discontinuity in the first-order derivatives of the surface. For the sake of simplicity we first assume that a discontinuity appears only in one direction. Hence, without loss of generality, we can choose the x-direction as the critical direction, whereas along the y-axis the surface is assumed to be constant. We will later relax this assumption. Let us first discuss jump-discontinuities.

As in the *continuous* case we define the normal vector as that eigenvector which corresponds to the smallest eigenvalue of $\mathbf{C}_I$. Here, we do not need to consider the proper choice of the eigenvectors controlled by the *order-preserving principle*, since a jump manifests itself in a large first eigenvalue and a corresponding eigenvector which can be interpreted as a tangent vector to an implicit curve. It can easily be shown that the eigenvector corresponding to the middle eigenvalue has no z-component. For this reason the eigenvector corresponding to the smallest eigenvalue can be defined as the normal vector, since ambiguity arises only in the case of an infinitely high jump. Again, such interpretations are derived from the fact that this local eigenvector/eigenvalue representation is concerned with the directions which best describe (in a least-squares sense) the directions of the surface orientation.

The magnitude of the jump discontinuity can be classified by either the steepness of the normal vector or the smallest eigenvalue which increases with jump height but remains bounded even for an infinitely high jump.

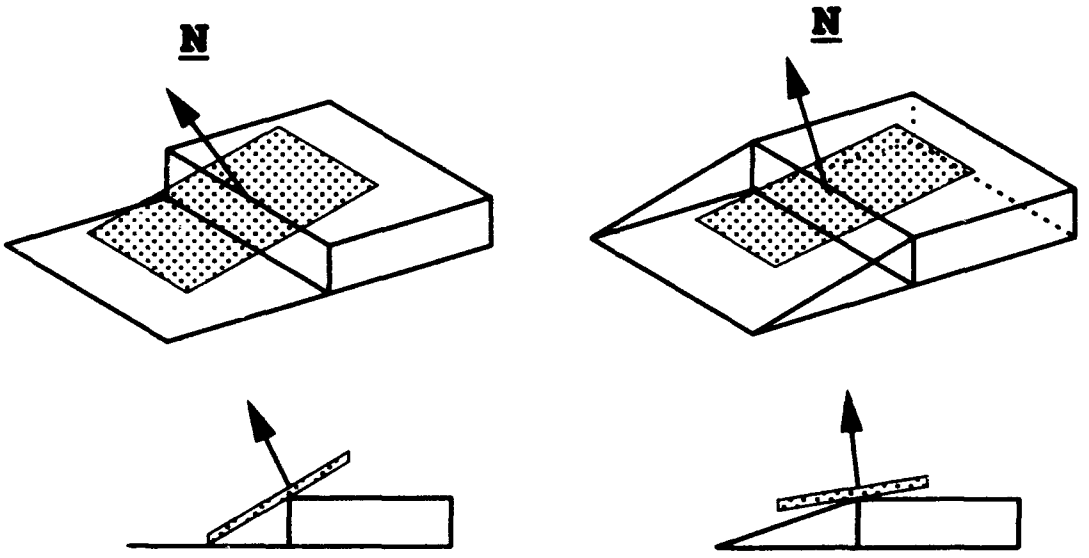


Fig. 3. Appropriate definitions of tangent planes (dotted planes) and normal vectors for jump (left) and crease (right) discontinuities derived from the covariance method with cross-sectional views (below).

Now consider crease-discontinuities. As in the case of jump discontinuities, the covariance method provides natural definitions for the crease's tangent plane and normal vector. The magnitude of the crease-discontinuity may be classified by either the largest or the smallest eigenvalue. Figure 3 illustrates a typical solution of tangent planes and normal vectors for jump- and crease-discontinuities.

If the crease is not symmetric the eigenplane tends to tip in the steeper direction of the crease. The defined normal vector can also be seen as a mean vector between the normal vectors on the left and the right side of the discontinuity. In contrast to jump-discontinuities, a non-symmetric crease may cause a situation where no order-preserving eigenplane can be found. However, this problem is solved by reformulating the *order-preserving principle*. If an eigenplane can be found which preserves the order of surface points under orthogonal projection onto the plane then we define this plane as the tangent plane. If no such tangent plane exists then the eigenvector corresponding to the smallest eigenvalue is chosen as the normal vector. This modification allows us to handle the situations described above.

Here we have dealt with the case of discontinuities in only one direction. However, even for the cases of two-dimensional discontinuities, like jump-corners or combined crease-jump discontinuities, the principles discussed above remain the same, up to the directions of the eigenvectors which will depend on the orientational structure of the discontinuities.

A final note should be made on the detection algorithms for jump- and crease-edges. In the literature, detection algorithms for crease-edges usually require the computation of the normal vector field of the surface. However, it has become clear in this section that with the covariance method a unified detection of jump- *and* crease-edges based on the $\mathbf{C}_I$-eigenvalue spectrum should be feasible.

5 Discussion

The aim of this paper was to extend the covariance technique introduced by Liang and Todhunter [6] and to develop a more formal representation of invariant surface descriptors without using surface parameterizations or calculus. In order to achieve this the covariance method was used as well as the *order-preserving principle*. It is important to note that the eigenvalues of this covariance matrix of the first kind are already invariant to rigid motions whereas differential geometry requires the calculation of second-order derivatives in order to obtain invariant quantities.

Furthermore an analogous operator $\mathbf{C}_{II}$ to the Weingarten map was defined, on the basis of a covariance matrix of weighted data which measure the deviation of points from the tangent plane in the neighbourhood of a particular point. The eigenvalues of this covariance matrix of the second kind are invariant to rigid motion as well, since the computation operates on the tangent plane only. A rigid motion of the entire surface will preserve the orthogonal distances of surface points onto the tangent plane. Table 1 summarizes the covariance representation and compares it with the definitions of differential geometry.

Table 1. Comparison of surface characteristics of differential geometry and discrete geometry using covariance methods.

Classical DG	*Covariance approach*
First fundamental form: $I = d\mathbf{u}^T \cdot \mathbf{A} \cdot d\mathbf{u}$, with $\mathbf{A} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$ $\mathbf{x}_u, \mathbf{x}_v, \mathbf{n} = (\mathbf{x}_u \times \mathbf{x}_v)/\lvert\mathbf{x}_u \times \mathbf{x}_v\rvert$	**Covariance of the first kind:** $I_C = \mathbf{v}^T \cdot \mathbf{C}_I \cdot \mathbf{v}$, with $\mathbf{C}_I = 1/n \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{x}_m)^T \cdot (\mathbf{x}_i - \mathbf{x}_m)$ $\mathbf{t}_1, \mathbf{t}_2, \mathbf{n} \quad \lambda_1^I, \lambda_2^I, \lambda_3^I$
Second fundamental form: $II = \mathbf{v}^T \cdot \mathbf{W} \cdot \mathbf{v}$	**Covariance of the second kind:** $II_C = \mathbf{v}^T \cdot \mathbf{C}_{II} \cdot \mathbf{v}$
Weingarten mapping: $d\mathbf{n} = \mathbf{W} \cdot d\mathbf{x}$ principal data: $\mathbf{v}_1, \mathbf{v}_2, k_1, k_2$	**Analogous Weingarten mapping:** $\Delta\mathbf{n} = \mathbf{C}_{II} \cdot \Delta\mathbf{x}$ principal data: $\mathbf{v}_1^{II}, \mathbf{v}_2^{II}, \lambda_1^{II}, \lambda_2^{II}$

In addition, it has been shown how the covariance method treats discontinuities in a very natural manner. Estimations of the eigenplane and the normal vector can be defined by using the same principles in the *continuous* as well as in the discontinuous case.

It has been dispensed with the notion of parameterization. The tangent and normal vectors were defined as eigenvectors of covariance matrices. Once more, the importance of the *order-preserving principle* is noted which controls the choice of the eigenvectors and that, already the covariance method of the first kind provides three eigenvalues which are invariant to rigid motion.

Finally, since covariance methods do not rely on the definition of a surface in terms of a surface parameterization, the spectrum of the covariance matrix (the full set of eigenvalues) provides us with a type of smooth transformation between lines, surfaces and volumes. One non-zero eigenvalue defines merely linear components while a solution of three equal eigenvalues corresponds to data of uniform density in a volume about a point — analogous to fractile-dimensionality. Further to this, covariance methods provide ideal ways of treating signals embedded in additive Gaussian or white noise and so may provide further integration of geometry with principles of signal processing.

Such covariance representations apply equally well to range or intensity image information where, in the latter case, the surface corresponds to the distribution of light over the projection plane. In this case the representation would be invariant to translations on the projection plane or in intensity — the latter not necessarily being always desired. Further, rotational invariance only has meaning in this latter context when restricted to the image plane (Barth *et al.* [1]).

References

1. Barth, E., Caelli, T., Zetzsche, C. (1993). Image encoding, labelling and reconstruction from differential geometry, Computer Vision, Graphics and Image Processing, in press.
2. Beal, P.J., Jain, R.C. (1986). Invariant surface characteristics for 3-D object recognition in range images, Computer Vision, Graphics and Image Processing 33, pp. 33–80.
3. Carmo Do, M.P. (1976). Differential Geometry of Curves and Surfaces, Prentice Hall, New Jersey.
4. Fan, T., Medioni, G., Nevatia, R. (1989). Recognising 3-D objects using surface descriptions, IEEE Transactions on Pattern Analysis and Machine Intelligence 11, pp. 1140–1156.
5. Jain, A.K., Hoffman, R. (1988). Evidence-based recognition of 3-D objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 10, pp.783–801.
6. Liang, P., Todhunter, J.S. (1990). Representation and recognition of surface shapes in range images, Computer Vision, Graphics and Image Processing 52, pp. 78–109.
7. Lipschutz, M.M. (1969). Differential Geometry, McGraw-Hill, New York.
8. Pennington, A., Caelli, T. (1991). Covariance techniques for invariant descriptions of scene geometry, Tech-Report 92-7, CITRI 723 Swanston St Carlton VIC 3053 Australia.

Image Representation Using Affine Covariant Coordinates *

Jun Zhang

Department of Psychology, Cognition and Perception Programs,
University of Michigan, Ann Arbor, Michigan 48104-2994, USA

Abstract. To achieve affine-invariant image representation and shape recognition, one must rely on a set of affine *co*variant image coordinates. It is proposed that such coordinates be derived from the second derivatives (Hessian) of a grey-level image. The two eigen-directions of the image Hessian are everywhere orthogonal. Connecting corresponding direction at neighbouring points results in a smooth flow field. Proper parameterization by Lie bracket operation gives rise to two orthogonal flow fields that may serve as the coordinate bases for an arbitrary image. From its construction, this coordinating "net" covaries with affine transforms of the visual manifold. Topological deformation of the image shape can be concisely described as Lie group actions on these curvilinear coordinates.

Keywords: shape description, Lie transformation group, image Hessian, Lie bracket, image coordinates, gauge transformation, Gestalt image.

1 Introduction

A central issue in shape representation is that of its invariance. It is common sense that human object recognition is invariant under linear transformations of visual space that may involve translation, scaling, and, to a certain extent, rotation ("affine" transforms technically). In order to derive affine-invariant descriptors of image shape, one must look for a set of affine *co*variant coordinates of the visual manifold (i.e. coordinates that covary with affine space transforms) to counteract the consequence of an affine transform on shape descriptors. In previous studies [3, 1, 4, 6] shape invariance under a global and uniform translation, rotation, or scaling was achieved through a mapping of the visual space onto itself via the action of a corresponding Lie group. The generators of those Lie transformation groups are image-independent, that is, they do not involve the specific image under analysis.

There is yet another kind of shape invariance, that is, under moderate but arbitrary deformation of the visual space where shapes are defined. Similarity

* The author thanks the McDonnell-Pew Center for Cognitive Neuroscience at San Diego and Dr. Terrence J. Sejnowski for support.

judgements can be easily made on an object undergoing various degrees and manners of distortion. However, different image shapes may tolerate different patterns of distortion before visual recognition finally breaks down. In other words, shape invariance under this local and point-wise distortion is image-specific. To characterize such distortion (and therefore achieve invariance), the set of desired coordinates, which serve to re-partition the visual space, ought to be derived from a grey-level image. The image-dependent coordination of the space will not only enable invariant shape description under *global* affine transforms, but also allow convenient expressions for *local* symmetry at individual locations associated with relatively "harmless" deformation.

The intention here is to derive such coordinates for an arbitrary grey-level image. Noting that the two-dimensional visual manifold is the natural support of shape perception (and visual perception in general), it is important to understand first why the said manifold should be "coordinated" in one way as opposed to another. One immediate reason is that a good, image-driven coordination of visual manifold facilitates (or even enables) perceptual processing. It has previously been suggested [8] that visual perception has the mathematical structure of fibre bundles. The sensory representation of image attributes is described as constructing vector fields defined on the visual space (base manifold). Image segmentation is achieved by identifying intrinsically constant portions of the sensory vector fields or cross-section of the fibre bundle under a given connection on the base manifold. The connection for the tangent bundle of the visual manifold (i.e. the space under which image motion is processed) has been derived. The "good" coordinates or geodesics are given by the first-order image gradients. The metric tensor consistent with this interpretation was shown to equal the square of the second-order gradients (the Hessian) of a grey-level image:

$$g_{\mu\nu} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} , \tag{1}$$

where f_{xx}, f_{yy}, and f_{xy} are second-order spatial derivatives of an image function or grey-level intensity distribution $f(x, y)$. The metric tensor is, quite naturally, symmetric (with respect to its lower indices) and semi positive-definite (non-negative). The associated Riemann curvature tensor is identically zero, indicating that it is indeed possible to globally define image coordinates on the two-dimensional base manifold where image shape is to be defined. The task then is to find these image coordinates based on the above-mentioned geometrical framework of visual perception.

2 Establishing Image-based Coordinates

Let us start by diagonalizing the image Hessian and, according to (1), the metric tensor of the visual manifold (the latter is called "perceptual" metric to distinguish it from the trivial, Euclidean metric). The characteristic directions or eigenvectors are (with corresponding eige λ_1, λ_2)

$$\begin{cases} \mathbf{n}_1 = \cos\phi\,\mathbf{e}_1 \quad \mathbf{e}_2 \\ \mathbf{n}_2 = -\sin\phi\,\mathbf{e}_1 + \cos\phi\,\mathbf{e}_2 \end{cases} \tag{2}$$

where ϕ is the angle to be rotated with respect to a pre-chosen Cartesian coordinate basis $\mathbf{e}_1$, $\mathbf{e}_2$ such that

$$\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} . \tag{3}$$

The second derivatives (Hessian) of an image function $f(x, y)$ can be explicitly related to λ_1, λ_2, and ϕ, after some rearranging:

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} \lambda_1 \cos^2\phi + \lambda_2 \sin^2\phi & (\lambda_1 - \lambda_2)\sin\phi\cos\phi \\ (\lambda_1 - \lambda_2)\sin\phi\cos\phi & \lambda_1 \sin^2\phi + \lambda_2 \cos^2\phi \end{pmatrix} . \tag{4}$$

The directional derivatives (Lie derivatives) along $\mathbf{n}_1$, $\mathbf{n}_2$ are, respectively,

$$\frac{d}{dl_1} = \cos\phi \frac{\partial}{\partial x} + \sin\phi \frac{\partial}{\partial y} , \qquad \frac{d}{dl_2} = -\sin\phi \frac{\partial}{\partial x} + \cos\phi \frac{\partial}{\partial y} . \tag{5}$$

Since

$$\frac{\partial f_{xx}}{\partial y} = \frac{\partial f_{xy}}{\partial x} , \tag{6}$$

it follows that

$$\begin{aligned} &\left(\sin\phi \frac{d}{dl_1} + \cos\phi \frac{d}{dl_2}\right)(\lambda_1 \cos^2\phi + \lambda_2 \sin^2\phi) \\ &= \left(\cos\phi \frac{d}{dl_1} - \sin\phi \frac{d}{dl_2}\right)((\lambda_1 - \lambda_2)\sin\phi\cos\phi) ; \end{aligned} \tag{7}$$

Since

$$\frac{\partial f_{xy}}{\partial y} = \frac{\partial f_{yy}}{\partial x} , \tag{8}$$

it follows that

$$\begin{aligned} &\left(\sin\phi \frac{d}{dl_1} + \cos\phi \frac{d}{dl_2}\right)((\lambda_1 - \lambda_2)\sin\phi\cos\phi) \\ &= \left(\cos\phi \frac{d}{dl_1} - \sin\phi \frac{d}{dl_2}\right)(\lambda_1 \sin^2\phi + \lambda_2 \cos^2\phi) . \end{aligned} \tag{9}$$

Simplifying (7) and (9) yields

$$\frac{d\lambda_2}{dl_1}\sin\phi + \frac{d\lambda_1}{dl_2}\cos\phi + \frac{d\phi}{dl_1}(\lambda_2 - \lambda_1)\cos\phi + \frac{d\phi}{dl_2}(\lambda_2 - \lambda_1)\sin\phi = 0 , \tag{10}$$

$$\frac{d\lambda_2}{dl_1}\cos\phi - \frac{d\lambda_1}{dl_2}\sin\phi - \frac{d\phi}{dl_1}(\lambda_2 - \lambda_1)\sin\phi + \frac{d\phi}{dl_2}(\lambda_2 - \lambda_1)\cos\phi = 0 , \tag{11}$$

or, written compactly ($(\cdot,\cdot)^T$ denotes matrix transposition),

$$(\lambda_2 - \lambda_1)\begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}\left(\frac{d\phi}{dl_1}, \frac{d\phi}{dl_2}\right)^T + \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}\left(\frac{d\lambda_1}{dl_2}, \frac{d\lambda_2}{dl_1}\right)^T = \begin{pmatrix} 0 \\ 0 \end{pmatrix} . \tag{12}$$

Therefore

$$\frac{d\phi}{dl_1} = \frac{1}{\lambda_1 - \lambda_2}\frac{d\lambda_1}{dl_2}\,, \qquad \frac{d\phi}{dl_2} = \frac{1}{\lambda_1 - \lambda_2}\frac{d\lambda_2}{dl_1}\,. \tag{13}$$

Note that, although the three functions λ_1, λ_2, and ϕ are algebraically independent at any image point, they are analytically related to each other at a common neighbourhood, as indicated by (13). This is because the Hessian of any image function satisfies (6) and (8).

Now, at each image location, there is a pair of orthogonal unit vectors (tiny "needles") which represent eigen-directions of the image Hessian. Assuming they are continuous vector fields (i.e., for sufficiently smooth images, presumably after filtering), the corresponding eigenvectors at neighbouring locations may be connected to form two orthogonal flow fields, which can in turn act on the two-dimensional visual manifold. Either flow field will "fill up" a surface patch; together, they mesh into a new coordinating "net" for the underlying manifold. However, the eigenvector fields (5) do not themselves form coordinate bases. To see this, calculate their Lie bracket which expresses the commutativity of the two flow field actions (c.f. [5, pp. 42-49]):

$$\begin{aligned}\left[\frac{d}{dl_1}, \frac{d}{dl_2}\right] &= \frac{d}{dl_1}\left(\frac{d}{dl_2}\right) - \frac{d}{dl_2}\left(\frac{d}{dl_1}\right) \\ &= \left(\cos\phi\frac{\partial}{\partial x} + \sin\phi\frac{\partial}{\partial y}\right)\left(-\sin\phi\frac{\partial}{\partial x} + \cos\phi\frac{\partial}{\partial y}\right) \\ &\quad - \left(-\sin\phi\frac{\partial}{\partial x} + \cos\phi\frac{\partial}{\partial y}\right)\left(\cos\phi\frac{\partial}{\partial x} + \sin\phi\frac{\partial}{\partial y}\right)\,. \end{aligned} \tag{14}$$

Evaluated and expressed in terms of d/dl_1 and d/dl_2, the Lie bracket becomes

$$\left[\frac{d}{dl_1}, \frac{d}{dl_2}\right] = -\left(\frac{d\phi}{dl_1}\frac{d}{dl_1} + \frac{d\phi}{dl_2}\frac{d}{dl_2}\right) \neq 0\,. \tag{15}$$

This means that the two directional fields, though mutually orthogonal, have not been properly parameterized to form a coordinate system. Let

$$\frac{d}{du} = \Lambda_1\frac{d}{dl_1}\,, \qquad \frac{d}{dv} = \Lambda_2\frac{d}{dl_2} \tag{16}$$

be the coordinate system along the same characteristic directions of the image, with (u, v) the orthogonal, properly parameterized coordinates and the two functions Λ_1, Λ_2 to be determined. Recalculate their Lie bracket:

$$\begin{aligned}\left[\frac{d}{du}, \frac{d}{dv}\right] &= \frac{d}{du}\frac{d}{dv} - \frac{d}{dv}\frac{d}{du} = \frac{d}{du}\left(\Lambda_2\frac{d}{dl_2}\right) - \frac{d}{dv}\left(\Lambda_1\frac{d}{dl_1}\right) \\ &= \Lambda_1\Lambda_2\left(\frac{d}{dl_1}\frac{d}{dl_2} - \frac{d}{dl_2}\frac{d}{dl_1}\right) + \frac{d\Lambda_2}{du}\frac{d}{dl_2} - \frac{d\Lambda_1}{dv}\frac{d}{dl_1} \\ &= \left(-\Lambda_2\frac{d\phi}{du} - \frac{d\Lambda_1}{dv}\right)\frac{d}{dl_1} + \left(-\Lambda_1\frac{d\phi}{dv} + \frac{d\Lambda_2}{du}\right)\frac{d}{dl_2}\,, \end{aligned} \tag{17}$$

where (15) is used in the last step. For (u, v) to be coordinates, (17) is required to be identically zero (from now on we write $\partial/\partial(\cdot)$ instead of $d/d(\cdot)$ to indicate that the directional derivatives of the u, v variables are, in addition, coordinate or "partial" derivatives):

$$\frac{\partial \phi}{\partial u} = -\frac{1}{\Lambda_2}\frac{\partial \Lambda_1}{\partial v}\,, \qquad \frac{\partial \phi}{\partial v} = \frac{1}{\Lambda_1}\frac{\partial \Lambda_2}{\partial u}\,. \tag{18}$$

This, along with (13), gives rise to the following useful identities:

$$\frac{\partial \phi}{\partial u} = -\frac{1}{\lambda_2 \Lambda_2}\frac{\partial(\lambda_1 \Lambda_1)}{\partial v}\,, \qquad \frac{\partial \phi}{\partial v} = \frac{1}{\lambda_1 \Lambda_1}\frac{\partial(\lambda_2 \Lambda_2)}{\partial u}\,; \tag{19}$$

$$\frac{\partial(\lambda_1 \Lambda_1)}{\partial v} = \lambda_2 \frac{\partial \Lambda_1}{\partial v}\,, \qquad \frac{\partial(\lambda_2 \Lambda_2)}{\partial u} = \lambda_1 \frac{\partial \Lambda_2}{\partial u}\,; \tag{20}$$

$$\frac{\partial \log \Lambda_1}{\partial v} = -\frac{1}{\lambda_1 - \lambda_2}\frac{\partial \lambda_1}{\partial v}\,, \qquad \frac{\partial \log \Lambda_2}{\partial u} = \frac{1}{\lambda_1 - \lambda_2}\frac{\partial \lambda_2}{\partial u}\,. \tag{21}$$

The last equation (21) specifies, at least in theory, the two unknown functions Λ_1 and Λ_2 up to a freedom to be discussed in the next section.

The integrability conditions for ϕ (i.e. the exchangeability of its mixed second derivatives) associated with (18) and (19) are as follows:

$$\frac{\partial}{\partial u}\left(\frac{1}{\Lambda_1}\frac{\partial \Lambda_2}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\Lambda_2}\frac{\partial \Lambda_1}{\partial v}\right) = 0\,, \tag{22}$$

$$\frac{\partial}{\partial u}\left(\frac{1}{\lambda_1 \Lambda_1}\frac{\partial(\lambda_2 \Lambda_2)}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\lambda_2 \Lambda_2}\frac{\partial(\lambda_1 \Lambda_1)}{\partial v}\right) = 0\,. \tag{23}$$

These equations may yield formal solutions under certain circumstances, but they do not impose further constraints on the two unknowns Λ_1, Λ_2.

3 Gauge Freedom in Image Coordinates

So far, we have derived, for an arbitrary image, the set of orthogonal curves $u = u_0$ and $v = v_0$ which correspond to directional flows of the image Hessian. These curves are dependent and only dependent on the image function $f(x, y)$; hence they are specific to a given image. This curvilinear coordinate net (u, v), together with the critical points of the image Hessian (either umbilic points $\lambda_1 = \lambda_2$, or degenerate points $\lambda_1 \lambda_2 = 0$), form the "signature" of a grey-level image. This characteristic flow "portrait" is called *image coordinates*.

The perceptual metric (1), being the square of the image Hessian, is diagonal under this new, image-dependent, coordinate system (u, v):

$$g_{\mu\nu} = \begin{pmatrix} (\lambda_1 \Lambda_1)^2 & \\ & (\lambda_2 \Lambda_2)^2 \end{pmatrix}\,, \tag{24}$$

so that the line element (under the perceptual metric) can be written as

$$ds^2 = (\lambda_1 \Lambda_1)^2\, du^2 + (\lambda_2 \Lambda_2)^2\, dv^2\,. \tag{25}$$

On the other hand, the line element under the physical metric ($d\bar{s}^2 = dx^2 + dy^2$, denoted by an overhead bar) is now expressible using (u, v) coordinates as

$$d\bar{s}^2 = \Lambda_1^2\, du^2 + \Lambda_2^2\, dv^2 \,, \tag{26}$$

with the metric tensor

$$\bar{g}_{\mu\nu} = \begin{pmatrix} \Lambda_1^2 & \\ & \Lambda_2^2 \end{pmatrix} . \tag{27}$$

From (27) and (24), one can see that the curvilinear image coordinates (u, v) are orthogonal under both the physical and the perceptual metric (their respective metric is diagonalized). In fact, these two families of coordinating curves are unique in that they are orthogonal everywhere and in any sense (physical or perceptual). As a comparison, the image-independent coordinates (x, y), which render trivial the physical metric (identity matrix), are not best suited to express the perceptual metric (resulting in the complicated Hessian matrix). Likewise, though the geodesic coordinates $(X, Y) = (f_x, f_y)$ trivialize the perceptual metric as in [8], they are not orthogonal under the image-independent physical metric.

As good image coordinates as (u, v) are, the metric tensor (either perceptual or physical) is not trivialized, but merely diagonalized. Furthermore, they are only partially determined, which will now be discussed. Note that there is a basic freedom in specifying any coordinate system, that is, the freedom of arbitrarily (but individually) stretching the two coordinating lines. If $\partial/\partial u$, $\partial/\partial v$ are the coordinate bases, then $A(u)\,\partial/\partial u$, $B(v)\,\partial/\partial v$ are also coordinates for arbitrary functions $A(u)$ and $B(v)$; it is easy to verify that their Lie bracket vanishes identically. In the present case, this freedom is reflected in the solutions of $\Lambda_1(u, v)$ and $\Lambda_2(u, v)$ from (18). To see this, let

$$\frac{\partial}{\partial u} \to A(u)\,\frac{\partial}{\partial u} \,, \qquad \frac{\partial}{\partial v} \to B(v)\,\frac{\partial}{\partial v} \,, \tag{28}$$

then

$$\Lambda_1(u, v) \to A(u)\,\Lambda_1(u, v) \,, \qquad \Lambda_2(u, v) \to B(v)\,\Lambda_2(u, v) \tag{29}$$

also satisfy (18), the equation to re-parameterize the flow fields. The directional derivatives (5) remain unaffected (see (16)). Equations (28) and (29) together form what can be called a *gauge transformation* of the image coordinates.

To fix a particular gauge, the following approach may be adopted. The Laplacian of any function in orthogonal coordinates under metric tensor $G_{\mu\nu}$ is given by (e.g., [2, p. 41])

$$\Delta = \frac{1}{\sqrt{G_{11}G_{22}}} \left(\frac{\partial}{\partial u} \left(\sqrt{G_{11}G_{22}}\, G^{11} \frac{\partial}{\partial u} \right) + \frac{\partial}{\partial v} \left(\sqrt{G_{11}G_{22}}\, G^{22} \frac{\partial}{\partial v} \right) \right) . \tag{30}$$

Applying the perceptual metric (24) yields:

$$\Delta\phi(u, v) = \frac{1}{(\lambda_2\Lambda_2)(\lambda_1\Lambda_1)} \left\{ \frac{\partial}{\partial u} \left(\frac{\lambda_2\Lambda_2}{\lambda_1\Lambda_1} \frac{\partial\phi}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{\lambda_1\Lambda_1}{\lambda_2\Lambda_2} \frac{\partial\phi}{\partial v} \right) \right\}$$

$$= \frac{1}{(\lambda_2\Lambda_2)(\lambda_1\Lambda_1)}\left\{\frac{\partial}{\partial u}\left(-\frac{1}{\lambda_1\Lambda_1}\frac{\partial(\lambda_1\Lambda_1)}{\partial v}\right)+\frac{\partial}{\partial v}\left(\frac{1}{\lambda_2\Lambda_2}\frac{\partial(\lambda_2\Lambda_2)}{\partial v}\right)\right\}$$

$$= \frac{1}{(\lambda_2\Lambda_2)(\lambda_1\Lambda_1)}\frac{\partial^2}{\partial u\partial v}\left(\log\frac{\lambda_2\Lambda_2}{\lambda_1\Lambda_1}\right) . \tag{31}$$

It can easily be verified that $\Delta\phi$ is a gauge-invariant quantity. It is suggested that a class of image functions ("good" or Gestalt images) exist such that

$$\Delta\phi = 0 . \tag{32}$$

This makes

$$\frac{\partial^2}{\partial u\partial v}\left(\log\frac{\lambda_2\Lambda_2}{\lambda_1\Lambda_1}\right) = 0 , \tag{33}$$

or

$$\log\left(\frac{\lambda_2\Lambda_2}{\lambda_1\Lambda_1}\right) = \log A(u) - \log B(v) . \tag{34}$$

The arbitrary functions $A(u)$, $B(v)$ may always be absorbed into Λ_1, Λ_2 respectively (due to the basic gauge freedom), giving rise to the perceptual gauge for Gestalt images:

$$\lambda_1\Lambda_1 = \lambda_2\Lambda_2 \ (= \Omega) . \tag{35}$$

From (19), it can been seen that $\phi(u,v)$ and $\log\Omega(u,v)$ satisfy Cauchy-Riemann equations:

$$\frac{\partial\phi}{\partial u} = -\frac{\log\Omega}{\partial v} , \qquad \frac{\partial\phi}{\partial v} = \frac{\log\Omega}{\partial u} . \tag{36}$$

The line element of the perceptual metric now becomes

$$ds^2 = \Omega^2\,(du^2 + dv^2) . \tag{37}$$

This is to say that, when the perceptual gauge (35) is satisfied, the perceptual metric is merely an isothermal mapping under image coordinates (u, v).

To find the relation between image coordinates (u,v) and geodesic coordinates $(X,Y) = (f_x, f_y)$ (as in [8]), apply the chain rule of differentiation and observe:

$$\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)^T = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}\left(\frac{\partial}{\partial X}, \frac{\partial}{\partial Y}\right)^T . \tag{38}$$

An application of (3), (5), and (16) yields

$$\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial v}\right)^T = \begin{pmatrix} \lambda_1\Lambda_1 & \\ & \lambda_2\Lambda_2 \end{pmatrix}\begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}\left(\frac{\partial}{\partial X}, \frac{\partial}{\partial Y}\right)^T . \tag{39}$$

Under the perceptual gauge (35), equation (39) can be recast after introducing complex variables $w = u + iv$, $Z = X + iY$:

$$dw = (\Omega\exp(i\phi))^{-1}dZ = \exp(-(\log\Omega + i\phi))\,dZ . \tag{40}$$

Since $\log\Omega$ and ϕ form a Cauchy-Riemann pair (see (36)), the above expression indicates that the image coordinates (u,v) and the geodesic coordinates (X,Y) are conformally related.

A similar derivation may be carried out using the physical metric $\bar{g}_{\mu\nu}$. The analogous equation for (31) is

$$\bar{\Delta}\phi(u,v) = \frac{1}{\Lambda_1\Lambda_2}\frac{\partial^2}{\partial u \partial v}\left(\log\frac{\Lambda_2}{\Lambda_1}\right) , \tag{41}$$

which leads to the following physical gauge:

$$\Lambda_1 = \Lambda_2 \ (= \Lambda) \tag{42}$$

and the corresponding Cauchy-Riemann pair:

$$\frac{\partial\phi}{\partial u} = -\frac{\log\Lambda}{\partial v} , \qquad \frac{\partial\phi}{\partial v} = \frac{\log\Lambda}{\partial u} . \tag{43}$$

Analogous to (40), the new coordinates (u,v) are now conformally related to the Cartesian coordinates (x,y) through (after introducing $z = x + iy$):

$$dw = (\Lambda\exp(i\phi))^{-1}dz = \exp(-(\log\Lambda + i\phi))\,dz . \tag{44}$$

The conformal relationships (40) under the perceptual gauge or (44) under the physical gauge indicate that ϕ — and hence (u,v) — will be completely specified given the boundaries and critical points of the mapping. These include locations where the image Hessian is umbilic ($\lambda_1 = \lambda_2$) or degenerate ($\lambda_1\lambda_2 = 0$).

4 Manipulations Using Image Coordinates

The curvilinear image coordinates (u,v) involve only the second derivatives with respect to space. Hence, they are covariant under an affine transformation (or mapping) of the two-dimensional visual space. They are "centred" on the visual image — in fact, the two sets of curves $u = u_0$ or $v = v_0$ represent "line-drawings" of a grey-level image. We now intend to relate their curvature, a geometrical descriptor, to the original image function. The intrinsic (geodesic) curvatures of the orthogonal coordinating curves $u = u_0$ and $v = v_0$ are (e.g., [7, p.130]):

$$k_1 = -\frac{1}{\sqrt{G_{11}G_{22}}}\frac{\partial\sqrt{G_{11}}}{\partial v} \quad \text{along } u = u_0 , \tag{45}$$

$$k_2 = \frac{1}{\sqrt{G_{11}G_{22}}}\frac{\partial\sqrt{G_{22}}}{\partial u} \quad \text{along } v = v_0 . \tag{46}$$

Under the perceptual metric $g_{\mu\nu}$, they are

$$k_1 = \frac{1}{\lambda_1\Lambda_1}\frac{\partial\phi}{\partial u} , \qquad k_2 = \frac{1}{\lambda_2\Lambda_2}\frac{\partial\phi}{\partial v} . \tag{47}$$

Likewise, under the physical metric $\bar{g}_{\mu\nu}$,

$$\bar{k}_1 = \frac{1}{\Lambda_1}\frac{\partial\phi}{\partial u} , \qquad \bar{k}_2 = \frac{1}{\Lambda_2}\frac{\partial\phi}{\partial v} . \tag{48}$$

Therefore

$$\bar{k}_1 = \lambda_1 k_1 \ , \qquad \bar{k}_2 = \lambda_2 k_2 \ . \tag{49}$$

Topological mappings of the visual manifold that preserve the neighbourhood relationship, including translation, rotation, scaling, and rubber-sheet deformation, can now be described as Lie dragging of the image "line-drawings". In particular, the two vector fields $A(u)\partial/\partial u$ and $B(v)\partial/\partial v$ themselves form the commutative bases for Lie transformation groups operating on an image or its associated sensory fields. They capture local symmetry of an image and express all sensible rubber-sheet deformations (see Sect. 1).

The image coordinates can be used to compute the intrinsic (Gaussian) curvature of a space via the well-known Gauss equation (e.g., [7, p.113])

$$K = -\frac{1}{2\sqrt{G_{11}G_{22}}}\left\{\frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G_{11}G_{22}}}\frac{\partial G_{11}}{\partial v}\right) + \frac{\partial}{\partial u}\left(\frac{1}{\sqrt{G_{11}G_{22}}}\frac{\partial G_{22}}{\partial u}\right)\right\} \ . \tag{50}$$

For the perceptual metric (24), a straightforward calculation using (23) yields

$$K = 0 \ . \tag{51}$$

This is consistent with the fact that the Riemann curvature tensor of the perceptual space vanishes identically [8]. There is no intrinsic curviness for the perceptual space. Since the Gaussian curvature is an intrinsic geometrical quantity, any three-dimensional embedding of the two-dimensional perceptual geometry, if possible, would have to be a developable surface. By the way, a similar calculation demonstrates, quite trivially, that the Gaussian curvature $\bar{K}$ under the physical metric (27) is also identically zero.

5 Patterns with Circular Symmetry: an Example

As a concrete example of this approach, let us calculate the image coordinates of a circularly symmetric pattern $f(x,y) = F(r)$, with $r \equiv \sqrt{x^2+y^2}$. The second derivatives (image Hessian) are

$$f_{xx} = \frac{x^2}{r^2}F'' + \frac{y^2}{r^3}F' \ , \quad f_{xy} = \frac{xy}{r^2}F'' - \frac{xy}{r^3}F' \ , \quad f_{yy} = \frac{y^2}{r^2}F'' + \frac{x^2}{r^3}F' \ , \tag{52}$$

with eigenvalues easily found to be

$$\lambda_1 = F''(r) \ , \qquad \lambda_2 = F'(r)/r \ . \tag{53}$$

Their corresponding eigenvectors simply point along the radial and angular directions respectively,

$$\mathbf{n}_1 = [x/r, y/r]^T \ , \qquad \mathbf{n}_2 = [y/r, -x/r]^T \ . \tag{54}$$

Denoting $\theta \equiv \arctan(y/x)$, the two directional derivatives in (5) become

$$\frac{d}{dl_1} = \frac{\partial}{\partial r} \ , \qquad \frac{d}{dl_2} = \frac{1}{r}\frac{\partial}{\partial \theta} \ . \tag{55}$$

Equation (13) can be verified as being satisfied. To solve for the unknown, parameterizing functions Λ_1, Λ_2, apply (16), (55) and (21):

$$\frac{\partial \log \Lambda_1}{\partial \theta} = -\frac{1}{\lambda_1 - \lambda_2}\frac{\partial \lambda_1}{\partial \theta} = 0 , \tag{56}$$

$$\frac{\partial \log \Lambda_2}{\partial r} = \frac{1}{\lambda_1 - \lambda_2}\frac{\partial \lambda_2}{\partial r} = \frac{1}{r} . \tag{57}$$

The solutions are, along with arbitrary functions $A(r)$ and $B(\theta)$,

$$\Lambda_1 = A(r) , \qquad \Lambda_2 = rB(\theta) , \tag{58}$$

so that

$$\frac{\partial}{\partial u} = A(r)\frac{\partial}{\partial r} , \qquad \frac{\partial}{\partial v} = B(\theta)\frac{\partial}{\partial \theta} . \tag{59}$$

Therefore, the variables r and θ (or arbitrary functions of either) are indeed image coordinates for circularly symmetric images, though they were previously introduced merely as shorthand notations of given functions of (x, y). In this case, both $\Delta\phi$ and $\tilde{\Delta}\phi$ equal zero (see (33) and (41)); thus patterns of circular symmetry are good "Gestalt" figures. It can be shown that $(u, v) = (\log F', \theta)$ and $(u, v) = (\log r, \theta)$ under perceptual and physical gauges respectively.

References

1. Ferraro, M., Caelli, T.M. (1988). Relationship between integral transform invariances and Lie group theory, J. Opt. Soc. Am. A 5, pp. 738–742.
2. Göckeler, M., Schücker T. (1987). Differential Geometry, Gauge Theories, and Gravity, Cambridge University Press, Cambridge, UK.
3. Hoffman, W.C. (1966). The Lie algebra of visual perception, J. Math. Psychol. 3, pp. 65–98.
4. Pintsov, D.A. (1989). Invariant pattern recognition, symmetry, and Radon transforms, J. Opt. Soc. Am. A 6, pp. 1544–1554.
5. Schutz, B.F. (1980). Geometrical Methods of Mathematical Physics, Cambridge University Press, Cambridge, UK.
6. Segman, J. (1992). Fourier cross correlation and invariance transformations for an optimal recognition of functions deformed by affine groups, J. Opt. Soc. Am. A 9, pp. 895–902.
7. Struik, D.J. (1950). Lectures on Classical Differential Geometry, Addison-Wesley, Massachusettes. Republished in 1988 by Dover, New York.
8. Zhang, J., Wu, S. (1990). Structure of visual perception, Proc. Natl. Acad. Sci. USA. 87, pp. 7819–7823.

Equivariant Dynamical Systems: a Formal Model for the Generation of Arbitrary Shapes

William C. Hoffman

Professor Emeritus, P.O. Box 2005, Sierra Vista, AZ 85636, USA

Abstract. The nature of the visual system is discussed. It achieves "constancy" and shape recognition by means of an exact map. The ideals of this mapping are invariants of G_V, the Lie transformation group of the constancies and of shape recognition, thus generating "perception by exception". The neuropsychological correlates, both neuronal and psychological, of the Lie transformation group are discussed at length in terms of the Bishop-Coombs-Henry model of the basic neocortical circuit and are shown to constitute a hyperbolic dynamical system. The local and global topologies of the latter are analysed. Shape recognition via annulment by the Lie derivatives of the constancies is discussed. Shape generation by means of the Lie transformation group of the constancies, using for this purpose the exponential map and "dragging the flow" along the group orbits, is then illustrated.

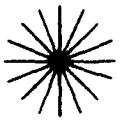
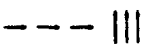
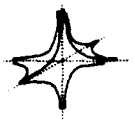
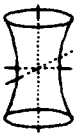
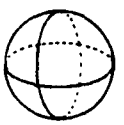
Keywords: shape generation, perceptual neuropsychology, geometric psychology, Lie tranformation groups, invariants, psychological constancies, pattern recognition.

1 Introduction: Structure and Function of the Visual Pathway

To record shape has been far easier than to understand it. C.S. Sherrington
Man on His Nature, Cambridge University Press, 1940.
Geometry is a magic that works. Rene Thom

The visual system processes shape stimuli by first of all preprocessing them in the visual cortex (Area 17) to remove the distortions imposed by the viewing conditions. This is the role of the visual constancies: size, shape, motion, binocularity, and colour. An electrode or strychnine pad placed upon the exposed visual cortex evokes such images as stars, sparks, whirling spirals, and circles, which are exactly the path-curves (*orbits*) of the constancy transformation groups listed in Table 1. An electrode or pad similarly placed upon the psychovisual cortex

Table 1. Lie transformation groups corresponding to the constancies. $\tau = c't$ where c' is the peak cortical signal velocity in cortical plane-time (x, y, t), the planar projection of the visual field of view. Constancies in parentheses are predictions of the theory that are borne out. $\mathcal{L}_x = \frac{\partial}{\partial x}$, $\mathcal{L}_y = \frac{\partial}{\partial y}$, $\mathcal{L}_\tau = \frac{\partial}{\partial \tau}$; $\partial_{var} = \frac{\partial}{\partial(var)}$.

Constancy	Lie tranformation group	Orbit patterns	Lie derivatives
Size constancy	Dilation group (perspectivities)		$\mathcal{L}_s = x\partial_x + y\partial_y$ $\mathcal{L}_{s1} = \tau\partial_\tau + x\partial_x$ $\mathcal{L}_{s2} = \tau\partial_\tau + y\partial_y$
Shape constancy	Affine (unimodular) group SL(2)		
Right-left, up-down location in field of view	Horizontal & vertical translation groups		$\mathcal{L}_x = \partial_x$, $\mathcal{L}_y = \partial_y$
(Form memory)	Time translations	Time continuum or clock "ticks" (discrete group)	$\mathcal{L}_\tau = \partial_\tau$
Orientation and obliquity	Rotation group SO(2)		$\mathcal{L}_O = -y\partial_x + x\partial_y$
Afferent binocular function	Pseudo-Euclidean (hyperbolic) rotations		$\mathcal{L}_b = y\partial_x + x\partial_y$
(Efferent binocular function)	Pseudo-Euclidean rotations in space-time		$\mathcal{L}_B = x\partial_x - y\partial_y$ $\mathcal{L}_{B1} = \tau\partial_\tau - x\partial_x$ $\mathcal{L}_{B2} = \tau\partial_\tau - y\partial_y$
Motion-invariant perception	Generalized Lorentz group of order 2		$\mathcal{L}_m = \mathcal{L}_M$ $\mathcal{L}_{m1} = \tau\partial_x + x\partial_\tau$ $\mathcal{L}_{m2} = \tau\partial_y + y\partial_\tau$
	Group of rotations SO(3)		$\mathcal{L}_M = -\mathcal{L}_O$ $\mathcal{L}_{M1} = x\partial_\tau - \tau\partial_x$ $\mathcal{L}_{M2} = y\partial_\tau - \tau\partial_y$

(Areas 18 and 19) produces ordinary mental images such as those we ordinarily see about us, *but without any apparent size or location in space.* It is clear that the primary role of the visual cortex is to preprocess higher visual forms and pass them on to the higher visual areas stripped of their distortion-induced redundancies. As von Fieandt [29] has put it, without the constancies we would always be moving through a surrealistic world of perpetually deforming rubbery objects. The penalty for memory storage is obvious.

McKay's Complementary (actually transverse) After Images (CAI) are precisely those induced by the orbits of the constancies, including certain new ones such as the hyperbolas corresponding to binocular functions. Furthermore, only the orbits of the constancies induce such CAI. McKay [24] attributed the site of such CAI to the visual cortex.

These same patterns — and their local linear combinations — are the bases for much of optical art and so apparently correspond to something rather deep and intrinsic in human perception. For example, the golden ratio of shape length to width follows readily from the logarithmic spiral pattern generated by a linear combination of the Lie derivatives for size constancy and the rotation-invariance component of shape constancy, $\mathcal{L}_s$ and $\mathcal{L}_0$ in Table 1.

2 The Topology of Visual Perception

2.1 Global Aspects

Not every psychologist will accept the figure-ground relation, which in effect states that figure-emerges-from-ground is a basic visual phenomenon, even though this seems to be the first sort of visual phenomenon noted by those persons blind from birth who have had their sight surgically restored [30]. But all psychologists will apparently accept the presence of visual contours. We therefore take such visual contours as the *states of the visual system.* It follows that, mathematically speaking, the "cortical retina" V is a path-connected manifold. (Colonnier's [5] "cortical retina", constituting the cortical correlate of the retina of the eye, will be denoted by V. The retina itself will be denoted by M.) The argument for the retina and "cortical retina" being manifolds has been given at length in [20] and will not be repeated here.

According to the visual constancies (size, shape, colour, motion, etc.), recognition of visual objects defined by visual contours is invariant under the distortions imposed by viewing conditions. This makes the "cortical retina" into a so-called *orbifold*, that is, a space with the local structure of the group orbits of a finite transformation group. The appropriate group here is the Lie group of the constancies G_V, acting on V:

$$G_V \times V \to V \to V/G_V \ , \tag{1}$$

where G_V is the direct product of the conformal group $CO(1,3)$ and the General Linear Group $GL(4,R)$ [17].

The retinal image on M projects along the visual pathway to the "cortical retina" V. The visual system is thus a *fibre structure* (M,p,V) consisting of the

two spaces V and M and a continuous surjection $p : M \rightarrow V$ that corresponds to the retinotopic projection. This mapping further has a cross-section corresponding to the inverse mapping that is given locally by $p^{-1} : V \rightarrow M$. This cross-section is a lifting from V to M of the identity map of the Lie group of the constancies G_V, that is, an efferent projection from the "cortical retina" to the midbrain region.

The temporal variation in the successive distortions of any given visual contour constitutes a *homotopy*, that is, a continuous deformation of any one distorted contour into successive ones. Since every homotopy (distorted image) upon M is lifted in this way onto V, the fibre structure (M, p, V) has the *covering homotopy property*. This makes the fibre structure into a *fibration*, in fact a Hurewicz fibration [6, p. 393]. This fact has important consequences, not only for the coherent lifting process from retina to "cortical retina" basic to shape perception, but also for cognitive phenomena, which may be regarded symbolically as fibrations in the sense of Kan [8, p. 65] acting upon the category of simplicial objects [19]. This is the way that *meanings* are attached to the *shapes* that we perceive. Attaching meaning to perceived form is the essence of cognition.

2.2 Local Structure

Thus far the structure and function of the visual system at the macroscopic, psychological scale have been considered. No less important is the structure and function at the microscopic, neuronal scale. According to the Neuron Doctrine, which asserts that the neuron is the fundamental structural, functional, and trophic unit of the nervous system, the two must be consistent, and the local structure must somehow generate the global, psychological structure. The many facets of Lie transformation groups make them and their associated manifolds ideal for this purpose. Where there are Lie transformation groups, fibre bundles, dynamical systems, functorial maps, and invariant structures cannot be far behind.

The Lie group $G_V \times V \rightarrow V$ is determined locally by a vectorfield $\mathbf{X}$ acting upon V. This makes the fibre bundle described above into a *vector bundle*. The infinitesimal transformations of this local structure are generated by the *Lie derivatives* that are determined by the group G_V. From a theorem of Vilms [28] we know that associated with the tangent space TV of a smooth vector bundle such as $(M, p, V; G_V)$ there are two kinds of vector bundle structures:

$$p_* : TM \rightarrow TV \ , \tag{2}$$

where p_* is the tangent map of p, and the tangent bundle $\tau : TM \rightarrow M$ itself. Here M is the visual field of view; p is, as noted earlier, the retinotopic projection from retina to "cortical retina"; V is the cortical manifold in subjective space-time; TV is the cortical vectorfield found electrohistologically by Hubel and Wiesel [22] and others; and τ denotes dissection of the field of view by tangent elements at the microscopic scale. By duality, the cotangent bundle T^*M induces representations as differential forms of the visual contours themselves in the cortical microstructure.

In short, the retinotopic map induces a cortical tangent bundle with Hubel and Wiesel orientation responses as vectorfield-generated cross-sections. At the same time appropriate contact transformations [13, 16, 20] generate visual contours as elements of the orbitspace V/G_V. Hence perceived shapes $f \in \mathcal{F}$ constitute equivariant imbeddings in the visual manifold V:

$$fg = gf \Rightarrow f = g^{-1}fg \ , \quad f \in \mathcal{F} \ , \quad g \in G_V \ . \tag{3}$$

If $\mathbf{X}$ is a vectorfield belonging to V/G_V, and if $f \in \mathcal{F}$ is a shape invariant under the Lie derivative of G_V, then

$$\mathcal{L}_{\mathbf{X}} f = 0 \ , \tag{4}$$

and all such $\mathbf{X}$ and f generate the *ideal* of the Lie group. This is the basis for the exact mapping cited above in the abstract:

$$\text{visual object} \xrightarrow{u} \text{perceived form} \xrightarrow{v} \text{shape recognition} \ .$$

The relationship between the local scale, given by TV and/or T^*V, and the scale of the full image is given by the *exponential map*:

$$TV \xrightarrow{exp} V \to V/G_V \ , \tag{5}$$

of which more later in connection with the generation of images.

As will shortly appear, cell morphologies of the visual cortex are such as to generate a *hyperbolic dynamical system*. The neocortical circuit model of Bishop *et al.* [1], together with the hyperbolic character of the neuronal network, strongly suggests a *normal hyperbolic flow* [10], that is, one wherein the the mapping of Tf normal to the neighbourhood U is hyperbolic and dominates the tangent behaviour.

More precisely, *normal hyperbolicity* is defined as follows. Suppose U is a smooth compact submanifold of V and $f(U)$ a diffeomorphism of V leaving U invariant, such as a constancy transformation. Then f is *normally hyperbolic* at U iff the tangent bundle of V, restricted to U, splits into three continuous subbundles

$$T_U V = N^u \oplus TU \oplus N^s \tag{6}$$

that are invariant under the tangent operation Tf, and further are such that

(i) Tf expands the unstable manifold N^u more rapidly than Tf expands TU;

(ii) Tf contracts the stable manifold N^s more rapidly than Tf contracts TU.

Normal hyperbolicity appears to describe well the branching of the neuronal arborescence from the cell body of the neuron (the *soma*). We will now establish the significance of this structure for the cortical neuronal net.

3 The Visual System is a Hyperbolic Dynamical System

In earlier papers [12, 14] it was postulated that cortical neurons constitute local phase portraits that are characteristic of the Lie derivatives of the constancies and their prolongations. In these papers it was shown that the intrinsic morphology of the dendritic arborescence in the neighbourhood of the perikaryon itself agrees in essentials with the appropriate local phase portraits, thus constituting a *Lie group germ*. This postulate antedated the Bishop-Coombs-Henry [1] model for the basic neocortical circuit and also Hubel and Wiesel's [22] findings on the electrophysiological nature of cortical cells, which appear to provide further support for the postulate. We therefore take this opportunity to update the theory and demonstrate the further correspondences to the postulate.

Hubel and Wiesel [22] found that the incoming information from the visual pathway is rearranged so that, first of all, most neurons in the visual cortex respond to specifically oriented line segments and, secondly, information originating from the two eyes converges upon single cells. These two functions are embodied in cortical structure and function wherein cells with common neuropsychological properties are grouped together in columns that traverse the cortical layers transversely, and ocular dominance columns respond to the same eye. This layered, columnar system of neurons is superimposed upon the well-known topographic representation of the visual field.

The foregoing description of functional cortical microstructure has the essential nature of a physiological embodiment of the general conformal group of transformations CO(1,3) acting upon Minkowski space, as is required by motion constancy [3]. The binocular representation also corresponds well to the known hyperbolic geometry of binocular visual space [23, 16]. The successive small shifts in orientation during traversal of adjacent hypercolumns connotes a cortical vectorfield. These columns are tubular neighbourhoods in the cortical fibre bundle described above, with a principal connection induced by the contact structure [20].

Bishop-Coombs-Henry's [1] model for the basic neural circuits of the visual cortex as follows is shown in Fig. 1.

Like Hubel and Wiesel, Bishop-Coombs-Henry [1] remark upon the presence of a slight overlap of successive afferents, together with a columnar structure of the cortex for processing these afferents. In mathematical terms this slight overlap, together with the recurrent neuronal morphology within the laminas of the cortical cytoarchitecture (Fig 2) represents progression through the exponential map (5) that generates contours. In other words, basic perceptual contours are generated by interacting neurons of the same morphological type which represent successive terms in the exponential map series. This is one type of action that could generate such visual contours.

According to "Geometric Psychology" (also termed the L.T.G./N.P.: the "Lie transformation group theory of neuropsychology"), the first-stage cortical processing of an afferent volley represents the action of the transformation group $G_V \times V \rightarrow V$, where V is the cortical representation of the visual field of view and the biochemical/psychophysical control processes reside in the parameter group

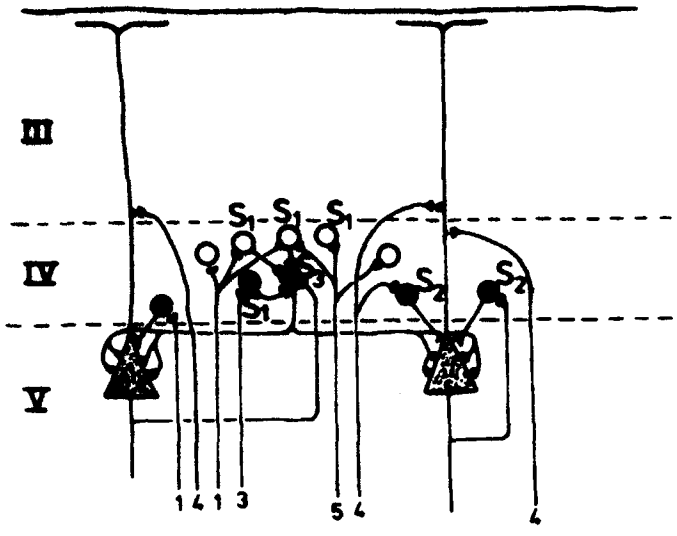


Fig. 1. Diagram of the Bishop-Coombs-Henry [1] model for the basic neural circuits in the organization of the receptive fields of simple cells in the striate cortex (area 17). The specific afferents are numbered to correspond to their lateral geniculate cells of origin. The stippled cells are pyramidal cells and the rest are various kinds of stellate cells. S_1, S_2, and S_4: Golgi Type II (short axon) stellate cells; S_3: basket cells. Open cell bodies and synaptic boutons: inhibitory neurons and synapses.

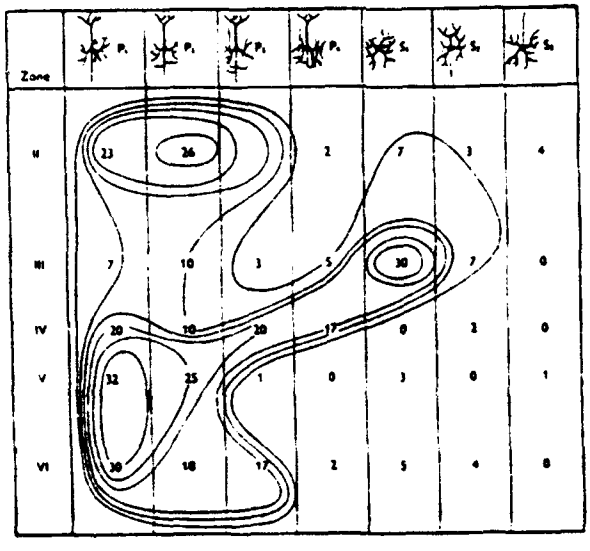


Fig. 2. Contour diagram for densities of pyramidal neurons (P_1, P_2, P_3, and P_4) and stellate cells (S_1, S_2, S_3) in the human visual cortex (after Sholl [27]). According to Colonnier [5] the pyramidal cells basically exhibit cylindrical symmetry; the stellate cells, radial or spheroidal symmetry. The numbers indicate the density of that type of neuron per cubic centimeter of cortical tissue.

G_V . The induced tangent map $p_* : TM \to TV$ thus defines a cortical tangent bundle which contains the local cortical vectorfields for the Hubel-and-Wiesel "orientation response." A local vectorfield $\mathbf{X}$ is thus defined on V, and associated with it, by a standard theorem on the characteristics of a partial differential equation, is a corresponding Pfaffian system [7]. The solution of the latter determines the orbit family embodied in the ensemble of neuronal processes. In turn the Pfaffian system induces a dynamical system. (See (7) - (9).)

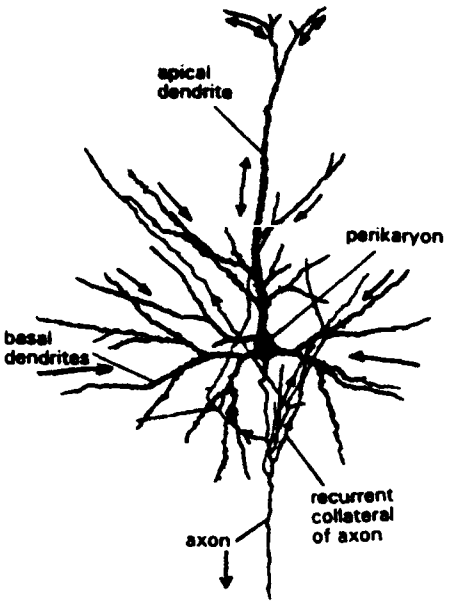


Fig. 3. Golgi preparation of a pyramidal neuron from the sensorimotor cortex of cat.

Consider the archetypal pyramidal neuron of Fig. 3. The classical view is that afferent volleys synapse upon the dendrites as well as the soma and are conducted electrochemically as excitatory post-synaptic potentials or inhibitory post-synaptic potentials toward the soma, as indicated by the arrows in the figure on the dendrites and basal dendrites. A traditional exception has been the apical dendrite, which is capable of antidromic flow that can conduct such post-synaptic potentials away from the cell body as well as toward it. A more recent view [26] is that other dendrites may also on occasion be capable of antidromic flow.

After the inward flow of an afferent volley has been sufficiently "integrated" within the cell in a graded, electronic way, an outgoing discharge, which can be efferent or corticortical, occurs along the cell's axon. This discharge has been likened to that of a firecracker fuse. It furnishes the stimulus to the next station (or stations) in the neuronal network. In addition to this "divergence" flow, a neuron may also exhibit pacemaker, recurrent, or spontaneous discharges.

The neural circuit configuration depicted in Fig. 1 is strongly suggestive of the dynamical system shown in Fig. 4. Let us therefore consider the evidence

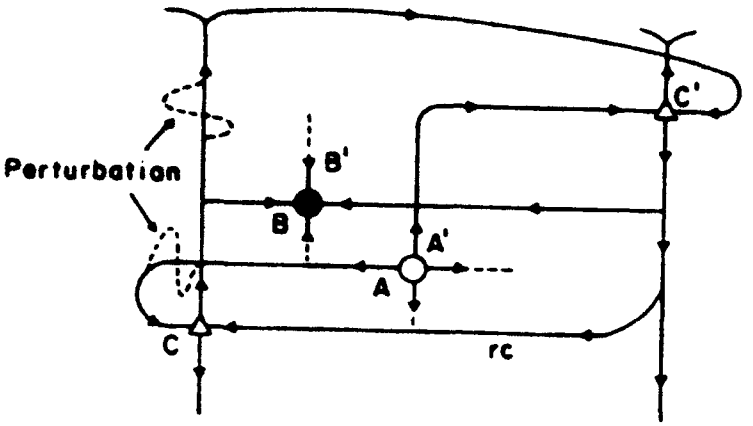


Fig. 4. Modified "preparation for an omega explosion" in a balanced hyperbolic dynamical system (after Nitecki, [25]). "Omega" here refers to the family of orbits that are associated with the system, through which the flow occurs after the "omega explosion." The latter apparently constitutes the cortical counterpart of the firecracker-fuse like axonal discharge. rc = recurrent collateral. Conventions as in Fig. 1.

from the L.T.G./N.P. that the configurations of Figs. 1 and 4 do in fact, apart from translations and rotations, embody a hyperbolic dynamical system.

Table 2 lists relative versus absolute invariants for the Lie group of the visual constancies. Absolute invariants are those that correspond to pattern recognition; they are the ones that are annulled by the action of the corresponding Lie derivative. Relative invariants, on the other hand, are derived functions generated by the action of the Lie derivative that are handed on to be recognized as contact differential forms at some subsequent state of visual processing. In other words, after a Lie derivative acts upon a shape that is not an absolute invariant, the result is transmitted down the perceptual pathway in *standard position*, free from distortions, to be recognized as an absolute invariant at some later stage.

In Table 2 it is worthy of note that $\mathcal{L}_s$, which governs size constancy, precedes all the other Lie derivatives in the list of relative invariants. Except for $\mathcal{L}_b$, which governs afferent binocular function and so would be distinguished cortically from monocular afferents, the Lie derivatives $\mathcal{L}_O$, $\mathcal{L}_m$, $\mathcal{L}_M$ all have some sort of circularly symmetric orbit structure resembling the approximately spheroidal pericellular nests of the basket cells. $\mathcal{L}_s$ itself has the orbit structure of a stellate cell (column (2) of Table 1). $\mathcal{L}_b$, $\mathcal{L}_B$, $\mathcal{L}_{m1}$, and $\mathcal{L}_{m2}$ have hyperbolic orbit structures resembling those of the archetypal pyramidal cell of Fig. 3. It is emphasized that in this progression of an afferent volley $\mathcal{L}_s$ (stellate cell morphology) precedes $\mathcal{L}_b$ and $\mathcal{L}_B$ (pyramidal cell morphology) as well as the translational Lie derivatives $\mathcal{L}_x$ and $\mathcal{L}_y$ (laminar grids of nerve fibers in layers I, III-IVa, and V) in accord with the progression depicted in Figs. 1 and 4 for the Bishop-Coombs-Henry model for the cortical neuronal circuit.

Table 2. Relative invariants corresponding to absolute invariants and their associated Lie derivatives.

(1) Constancy	(2) Absolute invariants	(3) Lie derivatives and sequences giving (2) as relative invariant	(4) Lie derivative(s) for which (2) is an absolute invariant
Shape constancy			
Location			
Horizontal	$(y,t) = const.$	$\mathcal{L}_s$; $\mathcal{L}_y$, $\mathcal{L}_t$; $\mathcal{L}_B$	$\mathcal{L}_x \rightarrow 0$
Vertical	$(x,t) = const.$	$\mathcal{L}_s$; $\mathcal{L}_x$, $\mathcal{L}_t$; $\mathcal{L}_B$	$\mathcal{L}_y \rightarrow 0$
Temporal	$(x,y) = const.$	$\mathcal{L}_s$; $\mathcal{L}_x$, $\mathcal{L}_y$; $\mathcal{L}_B$	$\mathcal{L}_t \rightarrow 0$
Orientation	$x^2 + y^2 = const.$	$\mathcal{L}_s$	$\mathcal{L}_O, \mathcal{L}_m, \mathcal{L}_M \rightarrow 0$
Afferent binocularity	$x^2 - y^2 = const.$	$\mathcal{L}_s$; $\mathcal{L}_B$	$\mathcal{L}_b \rightarrow 0$
Efferent binocularity	$xyt = const.$	$\mathcal{L}_s$; $\mathcal{L}_x$, $\mathcal{L}_y$; $\mathcal{L}_t$	$\mathcal{L}_B \rightarrow 0$
Size constancy	$y/x = const.$	$\mathcal{L}_O$, $\mathcal{L}_m$; $\mathcal{L}_b$; $\mathcal{L}_{s1}$, $\mathcal{L}_{s2}$; $\mathcal{L}_M$; $\mathcal{L}_B$; $\mathcal{L}_t$	$\mathcal{L}_s \rightarrow 0$
	$t/x = const.$	$\mathcal{L}_{m1}$; $\mathcal{L}_s$, $\mathcal{L}_{s2}$; $\mathcal{L}_{M1}$; $\mathcal{L}_B$; $\mathcal{L}_y$	$\mathcal{L}_{s1} \rightarrow 0$
	$t/y = const.$	$\mathcal{L}_{m2}$; $\mathcal{L}_s$, $\mathcal{L}_{s1}$; $\mathcal{L}_{M2}$; $\mathcal{L}_B$; $\mathcal{L}_x$	$\mathcal{L}_{s2} \rightarrow 0$
Invariant movement perception	$\tau^2 - x^2 - y^2 = const.$	$\mathcal{L}_s$; $\mathcal{L}_B$	$\mathcal{L}_m, \mathcal{L}_O \rightarrow 0$
Cyclopean (egocentred) perception	$\tau^2 + x^2 + y^2 = const.$	$\mathcal{L}_s$	$\mathcal{L}_M, \mathcal{L}_O, \mathcal{L}_m \rightarrow 0$

We now take up the question of whether the dynamical system associated with the Lie derivatives of Table 1 is hyperbolic in the technical mathematical sense. A typical Lie derivative from Table 1 is of the form

$$\mathcal{L}_{\mathbf{X}} = X_1(\xi,\eta)\frac{\partial}{\partial\xi} + X_2(\xi,\eta)\frac{\partial}{\partial\eta} \ , \tag{7}$$

where ξ and η denote any two of the variables x, y, and t. The corresponding Pfaffian system, whose solution gives the orbital structure of the Lie group, is

$$\frac{d\xi}{X_1(\xi,\eta)} = \frac{d\eta}{X_2(\xi,\eta)} \ , \tag{8}$$

and the associated dynamical system has the form

$$\frac{d}{da}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}X_1(\xi,\eta)\\ X_2(\xi,\eta)\end{pmatrix} \ , \tag{9}$$

where a is the parameter of the one-parameter Lie group that is provided by the psychobiological functioning of the visual-teleceptor system.

The coefficients X_1 and X_2 in the Lie derivatives of G_V are linear functions, and the associated dynamical systems take the form of a linear system

$$\frac{d}{da}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = A\begin{pmatrix}\xi\\ \eta\end{pmatrix} \ ,$$

where A is a 2×2 matrix consisting of zeros and ones. A singularity at the origin $O \in R^3$ of such a system is a *sink* if the eigenvalues of A have only negative real parts and a *source* if the eigenvalues have only positive real parts. A more general situation is that of *hyperbolic flow* [11] wherein all the eigenvalues of A need only have non-zero real parts.

We now take up in turn the dynamical systems and eigenvalues associated with each of the Lie derivatives in Table 1. For $\mathcal{L}_x = \partial/\partial x$, the A matrix is

$$A = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$$

which corresponds to a Pfaffian system of the form $dx/1 = dy/0$. The eigenvalues are 1 and 0, and similarly for the other translation operators $\mathcal{L}_y$ and $\mathcal{L}_t$. The translation operators are thus non-hyperbolic, which is as it should be, given the nature of the neuronal net, wherein horizontal and vertical translations are not associated cytoarchitecturally with particular neurons but instead apparently correspond to the plexuses of horizontal fibres that run laterally in layers I, III, IV, and V. Any hyperbolic dynamical system must therefore embody G_V minus at least its translation group. As we shall shortly see, the rotation subgroup must also be subtracted out. Neither exhibit the singularities required to have the neuronal soma as neuropsychological correlates.

Recall that for the Lie derivatives determining rotations, $\mathcal{L}_0 = -y\partial/\partial x + x\partial/\partial y$, and for those closely related, $\mathcal{L}_m = \mathcal{L}_M = -\mathcal{L}_0$, the neuroanatomical

correlate resides in the pericellular nests that surround the stellate and pyramidal cells. For $\mathcal{L}_0$ the matrix A takes the form

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

for which the eigenvalues are complex: $\pm i$, and there is no non-zero real part. Thus none of the Lie derivatives for G_V which lack a central singularity, namely translations and rotations, have a hyperbolic character.

However, all the other operators in G_V do have a hyperbolic character, which is as it should be, for it is the morphologies of the latter that are recognizable in Fig. 2. For the Lie derivative governing efferent binocular function $\mathcal{L}_B = (\mathcal{L}_B, \mathcal{L}_{B1}, \mathcal{L}_{B2})$, the A matrices have the respective forms

$$A_B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \quad A_{B1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} , \quad A_{B2} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} .$$

Each matrix has the eigenvalues $\lambda = \pm 1, \mathrm{Re}\,\lambda \neq 0$, so that the dynamical systems associated with $\mathcal{L}_B$ are *hyperbolic*. Given their similarity of form, we are led to identify the corresponding local phase portraits with the morphology of pyramidal neurons [14].

The time-varying operators of the 2-dimensional Lorentz group, $\mathcal{L}_{m1} = \tau\partial/\partial x + x\partial/\partial\tau$ and $\mathcal{L}_{m2} = \tau\partial/\partial y + y\partial/\partial\tau$, are also hyperbolic, the eigenvalues of the associated dynamical systems again being ± 1, and the same applies to the Lie derivative governing efferent binocular function $\mathcal{L}_b = y\partial/\partial x + x\partial/\partial y$.

The Lie derivatives for size constancy, $\mathcal{L}_s = (\mathcal{L}_s, \mathcal{L}_{s1}, \mathcal{L}_{s2})$, are also hyperbolic, the corresponding matrices all being of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

with eigenvalues 1, 1 and non-vanishing real parts. The corresponding neuroanatomical correlate, namely, the stellate cells, thus act as the sources for the pyramidal cell inputs in the other cortical layers in accord with the neuroanatomical realities.

To sum up, the neuron types in the Bishop-Coombs-Henry model for the basic neocortical circuit shown in Figs. 1 and 4 are of *hyperbolic* type. Their characteristic morphology agrees with that of the dynamical systems associated with the isotropy subgroups of G_V. The other neuron morphologies — pericellular nests, basket cells, and the horizontal rectangular grids of nerve fibers in layers I, IIIA, IVa and IVb, and Vc — are *non-hyperbolic*, again in accord with the L.T.G./N.P. Thus, as is consistent with the cytoarchitecture and the psychological correlates in Table 1, the respective neural systems act as separate entities that perform different neuropsychological functions. Here the main interest lies in the hyperbolic system involved in the Bishop-Coombs-Henry model for the basic neocortical circuit. For the other aspects, see [14, 16, 18, 20, 21].

Two properties of such hyperbolic flows are important for our purposes. First of all, there is Hartman's Theorem ([25, p. 80]) to the effect that in the neighbourhood of a hyperbolic point (the cell soma) the mapping is topologically

conjugate to the derivative at the hyperbolic fixed point. It follows that the local action of pyramidal and stellate cells is equivalent to differentiation (lateral inhibition), which once again establishes the presence of the induced tangent bundle mapping (2) $p_* : TM \to TV$.

A second important property of hyperbolic flows for present purposes is that such flows are direct sums of contraction and expansion mappings, as in (6). More generally, the Bishop-Coombs-Henry neocortical circuit, together with the hyperbolic nature of the other operators in G_V, is suggestive of the normal hyperbolic flow described in (6), wherein the tangent bundle $T_U V$ splits locally into a local tangent bundle TU and stable and wandering sets. In the context of cortical cytoarchitecture, this means that the axons of pyramidal cells wander more widely than do those of the intrinsic cells and those of the basket cells consisting of pericellular nests.

The **Fundamental Theorem of Normally Invariant Manifolds** [10] is as follows:

Let f be r-normally hyperbolic at U. Through U pass stable and unstable manifolds, invariant under f and tangent at U to $TU \oplus N^s$ and N^u, which are of class C^r. The stable manifold is invariantly fibered by C^r-submanifolds tangent at U to the subspaces N^s; similarly for the unstable manifold and N^u. These structures are unique and permanent under small perturbations of f. Similar results hold for flows.

This theorem describes well the instantaneous state of the neuronal net. In short, the following identifications are postulated — in accord with the the basic neocortical circuit of Bishop, Coombs, and Henry [1] and the known characteristics of a hyperbolic dynamical system — for the stellate and pyramidal cells of the visual cortex:

1. During the afferent phase the stellate cells of the visual cortex constitute sinks (N^s);
2. In their efferent phase the stellate cells become sources (N^u with $\mathrm{Re}\lambda > 0$).
3. In the afferent phase of pyramidal cell function the pyramidal cells constitute saddles owing to the possibility of antidromic flow in apical dendrites and recurrent collateral axons ($N^u \oplus N^s$).
4. In their efferent phase the pyramidal cells become sources (N^u) since axon discharges correspond to translation along a single dimension, that of the canonical coordinate established via the cell morphology.

(A Lie group is in *canonical form* when it is defined by the Lie derivative $\mathcal{L} = \partial/\partial \mathbf{y}$ for translations. The variable $\mathbf{y}$ which reduces it to this form is a *canonical variable*. It is a theorem [4] that every Lie group can be reduced to canonical form.)

A further word about the fourth identification in the above list. In general there is only one axon emergent from the soma that conducts away the efferent neuronal discharge. This feature is in marked contrast to the rest of the neuronal morphology which exhibits complicated branching in the neuronal arborescence. In accord with earlier work [13, 16], the neuronal output along an axon is regarded

as constituting a canonical coordinate for a one-parameter Lie group. Such a canonical coordinate can be either an absolute invariant of the Lie subgroup or else a differential invariant, first or higher, that has been established by flow through the neuronal morphology, the latter being regarded as the embodiment of a local phase portrait for the Lie derivative.

We therefore advance the following.

Principle. *The several neuronal morphologies constitute local phase portraits for the Pfaffian systems corresponding to the Lie derivatives of the constancies and their prolongations. Except for translations in the field of view and rotations, these Pfaffian systems correspond to hyperbolic dynamical systems. Axonal discharges thus correspond to canonical coordinates generated by flows through these phase portraits.*

4 Lie's Fundamental Theorem

Lie's Fundamental Theorem [4, p. 218] may be stated as follows:

Every Lie group involving r essential parameters has r linearly independent infinitesimal generators $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_r$, in terms of which every infinitesimal transformation of the group can be expressed as a linear combination

$$\mathcal{L} \equiv a_1\mathcal{L}_1 + a_2\mathcal{L}_2 + \ldots a_r\mathcal{L}_r \ . \tag{10}$$

Moreover, every transformation of this form, for all choices of the parameters $a_1, a_2, \ldots a_r$, belongs to the Lie group. $\mathcal{L}$ is made the infinitesimal generator of a one-parameter Lie group in this way.

In the neuropsychological context, each of the $\mathcal{L}_i$ corresponds to a constancy group transformation (see Table 1), and the a_i are psychophysical parameters communicated to the visual system by visual and teleceptor physiology and factored by the various neurotransmitter pathways. In the case of the visual system r appears to have the value 17 (Table 3).

Consider a visual form $f \in \mathcal{F}$. It has already been said that the non-specific recruiting response and the stimulus itself interact in such a way that the combination is cancelled by the action of the Lie derivative. This is achieved as follows. Let $g(V)$ be the non-specific input from the reticular activating formation. The Lie derivative (10) applied to the composition $g \circ f$ yields

$$\mathcal{L}(g \circ f) = \mathcal{L}g(f) = \frac{dg}{df}\mathcal{L}f = 0 \ ,$$

since, for f an invariant, $\mathcal{L}f = 0$. On the other hand, where the contour f is not present, $\mathcal{L}g \neq 0$, since g is like "background noise." In the above expression $\mathcal{L}$ will in general be one of the $\mathcal{L}_i$ rather than the general $\mathcal{L}$ in (10). The same argument then applies: $\mathcal{L}_i(f \circ g) \neq 0$ since f is not an invariant of the group of $\mathcal{L}_i$, and $\mathcal{L}_i(f \circ g)$ is passed on for further processing.

Table 3. Multiplication table for the Lie algebra LV of mature visual perception.

$[\mathcal{L}_p,\mathcal{L}_\sigma]$	$\mathcal{L}_x$	$\mathcal{L}_y$	$\mathcal{L}_t$	$\mathcal{L}_B$	$\mathcal{L}_{B1}$	$\mathcal{L}_{B2}$	$\mathcal{L}_m$	$\mathcal{L}_{m1}$	$\mathcal{L}_{m2}$	$\mathcal{L}_M$	$\mathcal{L}_{M1}$	$\mathcal{L}_{M2}$	$\mathcal{L}_b$	$\mathcal{L}_s$	$\mathcal{L}_{s1}$	$\mathcal{L}_{s2}$	$\mathcal{L}_O$
$\mathcal{L}_x$	0	0	0	$\mathcal{L}_x$	$-\mathcal{L}_x$	0	$-\mathcal{L}_y$	$\mathcal{L}_\tau$	0	$-\mathcal{L}_y$	$\mathcal{L}_\tau$	0	$\mathcal{L}_y$	$\mathcal{L}_x$	$\mathcal{L}_x$	0	$\mathcal{L}_y$
$\mathcal{L}_y$	0	0	0	$-\mathcal{L}_y$	0	$-\mathcal{L}_y$	$\mathcal{L}_x$	0	$\mathcal{L}_\tau$	$\mathcal{L}_x$	0	$\mathcal{L}_\tau$	$\mathcal{L}_x$	$\mathcal{L}_y$	0	$\mathcal{L}_y$	$-\mathcal{L}_x$
$\mathcal{L}_t$	0	0	0	0	$\mathcal{L}_t$	$\mathcal{L}_t$	0	$c'\mathcal{L}_x$	$c'\mathcal{L}_y$	0	$-c'\mathcal{L}_x$	$-c'\mathcal{L}_y$	0	0	$\mathcal{L}_t$	$\mathcal{L}_t$	0
$\mathcal{L}_B$	$-\mathcal{L}_x$	$\mathcal{L}_y$	0	0	0	0	$-2\mathcal{L}_b$	$\mathcal{L}_{M1}$	$-\mathcal{L}_{M2}$	$-2\mathcal{L}_b$	$\mathcal{L}_{m1}$	$-\mathcal{L}_{m2}$	$-2\mathcal{L}_m$	0	0	0	$2\mathcal{L}_b$
$\mathcal{L}_{B1}$	$\mathcal{L}_x$	0	$-\mathcal{L}_t$	0	0	0	$\mathcal{L}_b$	$-2\mathcal{L}_{M1}$	$-\mathcal{L}_{M2}$	$\mathcal{L}_b$	$-2\mathcal{L}_{m1}$	$\mathcal{L}_{m2}$	$-\mathcal{L}_M$	0	0	0	$-\mathcal{L}_b$
$\mathcal{L}_{B2}$	0	$\mathcal{L}_y$	$-\mathcal{L}_t$	0	0	0	$-\mathcal{L}_b$	$-\mathcal{L}_{M1}$	$-2\mathcal{L}_{M2}$	$-\mathcal{L}_b$	$-\mathcal{L}_{m1}$	$-2\mathcal{L}_{m2}$	$-\mathcal{L}_M$	0	0	0	$\mathcal{L}_b$
$\mathcal{L}_m$	$\mathcal{L}_y$	$-\mathcal{L}_x$	0	$2\mathcal{L}_b$	$-\mathcal{L}_b$	$\mathcal{L}_b$	0	$\mathcal{L}_{m2}$	$-\mathcal{L}_{m1}$	0	$\mathcal{L}_{M2}$	$-\mathcal{L}_{M1}$	$-2\mathcal{L}_B$	0	$\mathcal{L}_b$	$-\mathcal{L}_b$	0
$\mathcal{L}_{m1}$	$-\mathcal{L}_\tau$	0	$-c\mathcal{L}_x$	$-\mathcal{L}_{M1}$	$2\mathcal{L}_{M1}$	$\mathcal{L}_{M1}$	$-\mathcal{L}_{m2}$	0	$-\mathcal{L}_m$	$-\mathcal{L}_{m2}$	$2\mathcal{L}_{B1}$	$-\mathcal{L}_b$	$-\mathcal{L}_{M2}$	$-\mathcal{L}_{M1}$	0	$\mathcal{L}_{M1}$	$\mathcal{L}_{m2}$
$\mathcal{L}_{m2}$	0	$-\mathcal{L}_\tau$	$-c\mathcal{L}_y$	$\mathcal{L}_{M2}$	$\mathcal{L}_{M2}$	$2\mathcal{L}_{M2}$	$\mathcal{L}_{m1}$	$\mathcal{L}_m$	0	$\mathcal{L}_{m1}$	$-\mathcal{L}_b$	$2\mathcal{L}_{B2}$	$-\mathcal{L}_{M1}$	$-\mathcal{L}_{M2}$	$\mathcal{L}_{M2}$	0	$-\mathcal{L}_{m1}$
$\mathcal{L}_M$	$\mathcal{L}_y$	$-\mathcal{L}_x$	0	$2\mathcal{L}_b$	$-\mathcal{L}_b$	$\mathcal{L}_b$	0	$\mathcal{L}_{m2}$	$-\mathcal{L}_{m1}$	0	$\mathcal{L}_{M2}$	$-\mathcal{L}_{M1}$	$-2\mathcal{L}_B$	0	$\mathcal{L}_b$	$-\mathcal{L}_b$	0
$\mathcal{L}_{M1}$	$-\mathcal{L}_\tau$	0	$c'\mathcal{L}_x$	$-\mathcal{L}_{m1}$	$2\mathcal{L}_{m1}$	$\mathcal{L}_{m1}$	$-\mathcal{L}_{M2}$	$-2\mathcal{L}_{B1}$	$\mathcal{L}_b$	$-\mathcal{L}_{M2}$	0	$\mathcal{L}_M$	$-\mathcal{L}_{m2}$	$-\mathcal{L}_{m1}$	0	$\mathcal{L}_{m1}$	$\mathcal{L}_{M2}$
$\mathcal{L}_{M2}$	0	$-\mathcal{L}_\tau$	$c'\mathcal{L}_y$	$\mathcal{L}_{m2}$	$\mathcal{L}_{m2}$	$2\mathcal{L}_{m2}$	$\mathcal{L}_{M1}$	$\mathcal{L}_b$	$-2\mathcal{L}_{B2}$	$\mathcal{L}_{M1}$	$-\mathcal{L}_M$	0	$-\mathcal{L}_{M1}$	$-\mathcal{L}_{m2}$	$\mathcal{L}_{m2}$	0	$-\mathcal{L}_M$
$\mathcal{L}_b$	$-\mathcal{L}_y$	$-\mathcal{L}_x$	0	$2\mathcal{L}_M$	$-\mathcal{L}_M$	$\mathcal{L}_M$	$2\mathcal{L}_B$	$\mathcal{L}_{M2}$	$\mathcal{L}_{M1}$	$2\mathcal{L}_B$	$\mathcal{L}_{m2}$	$\mathcal{L}_{m1}$	0	0	$\mathcal{L}_m$	$-\mathcal{L}_m$	$-2\mathcal{L}_B$
$\mathcal{L}_s$	$-\mathcal{L}_x$	$-\mathcal{L}_y$	0	0	0	0	0	$\mathcal{L}_{M1}$	$\mathcal{L}_{M2}$	0	$\mathcal{L}_{m1}$	$\mathcal{L}_{m2}$	0	0	0	0	0
$\mathcal{L}_{s1}$	$-\mathcal{L}_x$	0	$-\mathcal{L}_t$	0	0	0	$-\mathcal{L}_b$	0	$-\mathcal{L}_{M2}$	$-\mathcal{L}_b$	0	$-\mathcal{L}_{m2}$	$-\mathcal{L}_m$	0	0	0	$\mathcal{L}_b$
$\mathcal{L}_{s2}$	0	$-\mathcal{L}_y$	$-\mathcal{L}_t$	0	0	0	$\mathcal{L}_b$	$-\mathcal{L}_{M1}$	0	$\mathcal{L}_b$	$-\mathcal{L}_{m1}$	0	$\mathcal{L}_m$	0	0	0	$-\mathcal{L}_b$
$\mathcal{L}_O$	$-\mathcal{L}_y$	$\mathcal{L}_x$	0	$-2\mathcal{L}_b$	$\mathcal{L}_b$	$-\mathcal{L}_b$	0	$-\mathcal{L}_{m2}$	$\mathcal{L}_{m1}$	0	$-\mathcal{L}_{M2}$	$\mathcal{L}_{M1}$	$2\mathcal{L}_B$	0	$-\mathcal{L}_b$	$\mathcal{L}_b$	0

$\mathcal{L}_B = x\partial_x - y\partial_y$, $\mathcal{L}_{B1} = t\partial_t - x\partial_x$, $\mathcal{L}_{B2} = t\partial_t - y\partial_y$; $\mathcal{L}_m = y\partial_x - x\partial_y$, $\mathcal{L}_{m1} = \tau\partial_x + x\partial_\tau$, $\mathcal{L}_{m2} = \tau\partial_y + y\partial_\tau$;
$\mathcal{L}_b = y\partial_x + x\partial_y$, $\mathcal{L}_O = -\mathcal{L}_m$; $\mathcal{L}_M = \mathcal{L}_m$, $\mathcal{L}_{M1} = x\partial_\tau - \tau\partial_x$, $\mathcal{L}_{M2} = y\partial_\tau - \tau\partial_y$;
$\mathcal{L}_s = x\partial_x + y\partial_y$, $\mathcal{L}_{s1} = t\partial_t + x\partial_x$, $\mathcal{L}_{s2} = t\partial_t + y\partial_y$; $\mathcal{L}_x = \frac{\partial}{\partial x}$, $\mathcal{L}_y = \frac{\partial}{\partial y}$, $\mathcal{L}_t = \frac{\partial}{\partial t}$, $\mathcal{L}_\tau = \frac{\partial}{\partial \tau}$;
$\partial_{var} = \frac{\partial}{\partial(var)}$, $\tau = c't$.

5 Application: the Generation of Shape

Thus far the focus has been on form memory and the basis for visual pattern recognition, wherein image *recognition* occurs through cancellation by the action of a Lie derivative upon the shape contours. We now turn attention to the *generation* of shape. For this purpose there are two methods: the exponential map and "dragging the flow" along the group orbits. It is worthy of note in this connection, however, that the rotation group operation is enough to specify a plane curve in terms of its natural equation in the intrinsic coordinate arclength s. It is a theorem that any plane curve can be specified in terms of the natural equation $\kappa = \kappa(s)$, where κ is the curvature,

$$\kappa = \frac{y''}{(1 + (\frac{dy}{dx})^2)^{\frac{3}{2}}} .$$

Prolongation of the Lie derivative $\mathcal{L}_0$ for the rotation group leads to the intrinsic equation

$$\kappa = F(u, u_1) = F(x^2 + y^2, \frac{y - xy'}{x + yy'}) ,$$

where u is the absolute invariant, and u_1 the first differential invariant of $\mathcal{L}_0$.

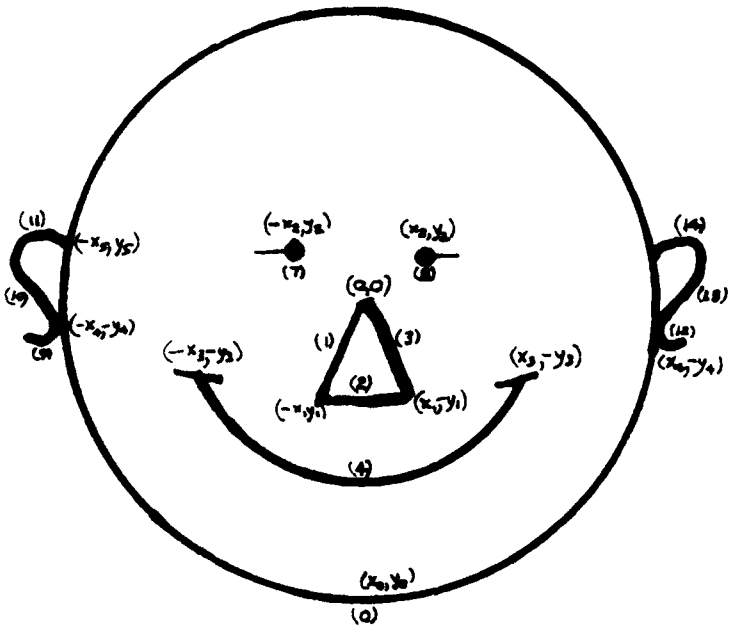


Fig. 5. The familiar "smiling face" augmented by a triangular nose.

Consider the "smiling face" (Fig. 5), familiar from many posters and advertisements, modified here by the addition of a triangular nose. The curves and

arcs in Fig. 5 have been numbered (1) to (14). We now show how the exponential map (5) may be used to generate these various curves. Curve (0) is a circle and hence an orbit of the rotation group of the plane. Let (x, y) be any point on this curve. The exponential map

$$e^{a\mathcal{L}_0}(x,y) = \sum_{n=0}^{\infty} \frac{a^n}{n!}\mathcal{L}_0^n = (x\cos a - y\sin a, x\sin a + y\cos a)$$

then gives the parametric equation of a circle, where the parameter a runs from 0 to 2π. The circular eyes (7) and (8) are generated in the same way, applying the exponential map to the points $(x + x_2, y - y_2)$ (right eye) and $(x - x_2, y - y_2)$ (left eye). The parameter a again traces out the full range $(0, 2\pi)$. The succession of terms in the exponential expansion corresponds to the role of the cytoarchitecture. Each term marks another element of the progression through a cortical layer of particular neuron types.

The mouth (4) is an arc of a circle and so is generated by the same sort of exponential map, where the parameter a now runs from $-a_1$ to a_1.

Consider now the triangular nose. Triangles are conformal invariants, and so the nose is, as it should be, an invariant of the special projective group [2]. Consider first the line segment (1): $y_1 x + x_1 y = 0$. The exponential map of $\alpha\mathcal{L}_x + \beta\mathcal{L}_y$ applied to the initial point on this line

$$\begin{aligned} e^{a(\alpha\mathcal{L}_x+\beta\mathcal{L}_y)}(x,y) &= (1 + a((\alpha\mathcal{L}_x + \beta\mathcal{L}_y)) + \frac{a^2}{2!}(\alpha\mathcal{L}_x + \beta\mathcal{L}_y)^2 + \ldots)(x,y) \\ &= (x + a\alpha, y + a\beta) \end{aligned}$$

takes $(x, y) = (0, 0)$ to $(a\alpha, a\beta) = (-x_1, -y_1)$. It follows that $a = -x_1/\alpha = -y_1/\beta$. Choosing $\alpha = 1$ implies $\beta = y_1/x_1$, and $a = -x_1$. Thus

$$e^{a(\mathcal{L}_x+\frac{y_1}{x_1}\mathcal{L}_y)}(x,y) = (x - x_1, y - y_1) \ ,$$

and, for $(x, y) = (0, 0)$, the exponential map yields the terminal point $(-x_1, -y_1)$ of curve (1).

To illustrate the way that the Lie derivative "drags the flow" along the path-curves we use a drawing (Fig. 6), "Preliminary draft of the portrait of Sir Osiris, Paris, 1990," contributed by Pierre Székely to *Symmetry: Culture and Science.*

The only curve that differs from the straight line segments and arcs of circles that were treated previously by means of the exponential map is Sir Osiris' elliptically shaped nose. We therefore illustrate "dragging the flow" on an elliptical orbit of the Lie group of $\mathcal{L}_e$, where

$$\mathcal{L}_e = -\alpha y\frac{\partial}{\partial x} + \beta x\frac{\partial}{\partial y} \ .$$

The next step in "dragging the flow" involves going to the corresponding Pfaffian system and finding the invariant curves [4]:

$$\frac{dx_1}{-\alpha y_1} = \frac{dy_1}{\beta x_1} = \frac{da}{1} \ , \tag{11}$$

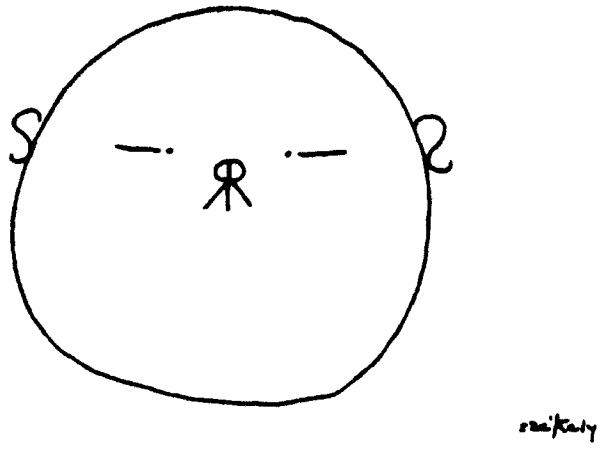


Fig. 6. Preliminary draft of the portrait of Sir Osiris, Paris, 1990, by Pierre Székely. (after Symmetry: Culture and Science, 1990, p. 109).

the solution of which is the ellipse $\phi(x,y) = \beta {x_1}^2 + \alpha {y_1}^2 = \beta x^2 + \alpha y^2 = c$, where c = constant, as an invariant of the group. The next step is to solve this equation for one variable in terms of the other:

$$x_1 = \sqrt{\frac{c}{\beta} - \frac{\alpha}{\beta} y_1^2} \ . \tag{12}$$

Substituting this back into (11) and integrating yields

$$\sqrt{\frac{\beta}{\alpha}} \ \arcsin \frac{y_1}{\sqrt{\frac{c}{\alpha}}} = a$$

whence, recalling (12), we have the parametric equations of the elliptical orbit in terms of the parameter a and indexed by the constant c:

$$x_1 = \sqrt{\frac{c}{\beta}} \ \cos(\sqrt{\frac{\alpha}{\beta}} a) \ , \quad y_1 = \sqrt{\frac{c}{\alpha}} \ \sin(\sqrt{\frac{\alpha}{\beta}} a) \ .$$

6 Application to Affine Transformations

The group of affine transformations is the direct product of the General Linear Group and the group of translations,

$$\aleph \mathbf{x} = A\mathbf{x} + \mathbf{c} \ .$$

[7, p. 43] gives the following formula for the vectorfield that generates the affine group:

$$X_i = c_i + \sum_{j \neq i} \gamma_{ij} x_j \ , \quad (i,j) = 1, 2, \ldots, n \ .$$

In the planar case affine transformations are thus the result of rotations, dilations, and translations, the Lie derivatives for which are given in Table 1. Either the exponential map or the "dragging the flow" method can be used to determine the associated path-curves.

7 Conclusion

Reasons have been reviewed for thinking that the Lie group of the constancies and its prolongations describe the phenomena of visual perception, and the connection with standard models of the neuronal cytoarchitecture has been shown. The visual pathway involves preprocessing at the visual cortex level to remove the distortions of shape imposed by viewing conditions. This is followed at the level of the psychovisual cortex by image recognition of actual shape in standard position free from distortions. It has also been shown that the neuronal soma correspond to the isotropy subgroup of the Lie group of the constancies after subtracting translations and rotations, and so to a hyperbolic dynamical system. The topology of the latter was discussed in terms of the Hirsch-Pugh-Shub theorem on normally invariant manifolds.

Methods of image generation were also demonstrated. Used for this purpose were the exponential map of a Lie group and the process of "dragging the flow" along orbits by means of the dynamical / Pfaffian system associated with the Lie derivatives of a Lie group. The generation of higher differential invariants for higher forms was not covered but may be found in [13, 15, 18]. REDUCE is the natural software [9] for computation of such differential-geometric, partial-differential-equation entities as those that have been discussed in this paper.

References

1. Bishop, P.O., Coombs, J.S., Henry, G.H. (1971). Interaction effects of visual contours on the discharge frequency of simple striate neurons, J. of Physiology 219, pp. 659–687.
2. Bluman, G.W., Cole, J.D. (1974). Similarity Methods for Differential Equations, Springer-Verlag, New York.
3. Caelli, T., Hoffman, W.C., Lindman, H. (1978). Subjective Lorentz transformations and the perception of motion, Journal of the Optical Society of America 68, pp. 402–411.
4. Cohen, A. (1931). An Introduction to the Lie Theory of One-Parameter Groups, Hafner, New York.
5. Colonnier, M. (1964). The tangential organization of the visual cortex, Journal of Anatomy (London) 98, pp. 327–344.
6. Dugundji, J. (1965). Topology. Allyn & Bacon, Boston.
7. Eisenhart, L.P. (1961). Continuous Groups of Transformations, Dover, New York.
8. Gabriel, P., Zisman, M. (1967). Calculus of Fractions and Homotopy Theory, Springer-Verlag, Berlin.
9. Gragert, P.K.H., Kersten, P.H.M., Martini, R. (1983). Symbolic computations in applied differential geometry, Acta Applicandae Mathematicae 1, pp. 43–77.

10. Hirsch, M.W., Pugh, C.C., Shub, M. (1977). Invariant Manifolds, Springer Verlag, New York.
11. Hirsch, M.W., Smale, S. (1974). Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, NewYork.
12. Hoffman, W.C. (1968). The neuron as a Lie group germ and a Lie product, Quarterly of Applied Mathematics 25, pp. 423–440.
13. Hoffman, W.C. (1970). Higher visual perception as prolongation of the basic Lie transformation group, Mathematical Biosciences 6, pp. 437–471.
14. Hoffman, W.C. (1971). Memory grows, Kybernetik 8, pp. 151–157.
15. Hoffman, W.C. (1977). An informal, historical description (with bibliography) of the L.T.G./N.P., Cahiers de Psychologie 20, pp. 135–174.
16. Hoffman, W.C. (1978). The Lie transformation group approach to visual neuropsychology. In: Leeuwenberg, E. L. J., Buffart, H. (eds.), Formal Theories of Visual Perception, Halsted Press of John Wiley, Chichester, pp. 27–66.
17. Hoffman, W. C. (1980). Subjective geometry and Geometric Psychology, Mathematical Modelling 1, pp. 349–367.
18. Hoffman, W.C. (1984). Figural synthesis by vectorfields: Geometric Neuropsychology. In: Dodwell, P.C., Caelli, T. (eds.), Figural Synthesis. Erlbaum, Hillsdale.
19. Hoffman, W.C. (1985). Some reasons why algebraic topology is important in neuropsychology: perceptual and cognitive systems as fibrations, International Journal of Man-Machine Studies 22, pp. 613–650.
20. Hoffman, W.C. (1989). The visual cortex is a contact bundle, Applied Mathematics and Computation 32, pp. 137–167.
21. Hoffman, W.C. (1990). The conformal group CO(1,3) as basis for both "Nature" and perception cum self-reference. In: Manikopoulos, C.N. (ed.), Proc. of the 8th International Congress of Cybernetics and Systems, Vol. I, New Jersey Institute of Technology Press, Newark.
22 Hubel, D. H., Wiesel, T.N. (1977). Ferrier Lecture: Functional architecture of macaque monkey visual cortex, Proc. of the Royal Society of London 198B, pp. 1–59.
23. Luneburg, R.K. (1950). The metric of binocular visual space, Journal of the Optical Society of America 40, pp. 627–642.
24. McKay, D.M. (1960). In: Rosenblith, W.A. (ed.), Sensory Communication. John Wiley, New York.
25. Nitecki, Z. (1971). Differentiable Dynamics, M.I.T. Press, Cambridge.
26. Shepherd, G.M. (1974). The Synaptic Organization of the Brain, Oxford University Press, New York.
27. Sholl, D.A. (1956). The Organization of the Cerebral Cortex, Methuen, London.
28. Vilms, J. (1967). Connections on tangent bundles, Journal of Differential Geometry 1, pp. 235–243.
29. von Fieandt, K. (1966). The World of Perception, Dorsey Press, Homewood, Illinois.
30. Von Senden, M. (1960). Space and Sight: The Perception of Space and Shape in the Congenitally Blind Before and After Operation, Free Press, Glencoe, Illinois.

Neural Processing of Overlapping Shapes *

André J. Noest

Utrecht Biophysics Institute, Department of Medical and Physiological Physics, Utrecht University, Princetonplein 5, NL-3584-CC Utrecht, The Netherlands

Abstract. Visual information from physically distinct sources often becomes overlaid or finely interspersed in the very process of image formation. For example, one may think of shadows overlaid on surface patterns, or of the multitude of tree branches and leaves that occur in images of a forest. Analyzing such images leads naturally to multi-valued fields of local features. This paper proposes a general model structure for recovering shape characteristics from such data. It uses a "blurred relation" representation to group and segment the data in a way that agrees well with psychophysical and neurophysiological evidence. Some core examples of the behaviour of the model are worked out analytically.

Keywords: shape, neural networks, transparency, visual coherence, multi-valued fields, grouping, splitting, interpolation, segmentation.

1 Introduction

Living and artificial systems alike must exploit the space-time structure of their environment. However, the extraction of useful geometric information from sensory inputs is often a non-trivial problem. For vision, it is the rule rather than the exception that even simple geometric structure in the world is partially lost or non-trivially transformed in the very process of image formation. Consider, for example, an environment of long grass and bushes, or the canopy of a forest. Objects of interest are usually seen only through many small gaps in the grass or foliage. The visual structure of the object has thus become spatially interleaved with that of the occluders. How is the resulting image to be segmented, and how can the scattered data from multiple sources be grouped in a sensible way? Often the situation will be worse still, because the foliage will sway in the wind. The occlusion pattern then fluctuates wildly, and the sunlight piercing the canopy casts dynamic patterns of light and shade on the scene below. The dangerous consequences of failing to separate such overlaid or intertwined patterns must have influenced the evolution of most animals. Accordingly, my proposal for handling

* This work was sponsored by SNN, the Netherlands Foundation for Neural Network Research.

these problems attempts to incorporate some relevant ideas and results from psychophysics and neurobiology. Thus, one aim is to develop a general model of the neural networks that tackle these problems in living systems. Independently of its biological veridicality, the model may be of use to computer-vision systems.

1.1 The Basic Problem: Grouping and Splitting

The task is to group local measurements of visual attributes into coherent, distinct object representations suitable for computing quantities that characterize the object positions and shapes. In particular, the question is how to do this when data from multiple objects has become spatially interleaved or pointwise superimposed. After introducing an appropriate representation, very simple and neurally plausible operations will be shown to produce a "grouping and splitting" process that brings together spatially dispersed data which probably originated from one object(-patch), while separating intertwined data from probably distinct objects. I also propose ways of extracting from this representation some shape characteristics which have hitherto been discussed only in a more convential setting.

The proposed model differs fundamentally from that of most presently popular and well-studied models that exploit assumed spatial coherence in attribute fields. Many of these are based on regularization theory [8]. Their goal is to reconstruct say velocity or depth as a function of space, using certain smoothness constraints. However, most regularization models are equivalent to spatial smoothing or spline-fitting, which implies a conflict between interpolation and the preservation of crisp boundaries. Embellishments such as "breakable" splines (or strongly related nonlinear diffusion) can preserve discontinuities [1], but only at the cost of introducing convergence and uniqueness problems. The fundamental problem of all models using function-based representations is the inability to let multiple attribute values coexist locally. Thus, unlike human vision, these models cannot handle "transparency", that is overlaid or finely interleaved data from multiple objects.

2 "Blurred Relations" Models

The data representation which is central to the proposed way of handling multi-valued data naturally can be considered as an elaboration of the notion of "channel-coding" [9] in psychophysics. The basic idea is to let each position carry not some estimate of the local attribute value, but a distribution over the set of all possible values. There are parallels with "value-cell" or "generalized Hough"-coding from computer science, but these methods usually lack the required geometric structure

Conceptually, one can distinguish at least three stages of processing: First, local measurements are taken. It is very important to have a proper choice of the type of measurement and the format in which they are represented. Second, local data will be "grouped" to exploit the physically plausible coherence within

and incoherence between objects. Third, quantities characterizing the shape of contours and surfaces are to be computed. In reality, the latter two stages may be strongly intertwined.

2.1 Local Measurement or Pre-processing

The key characteristic of the detectors in the first stage is that they have local, spatially overlapping "receptive fields", and that their responses are "tuned" smoothly for at least one attribute dimension, for example local orientation, velocity, binocular disparity, etc. Much of the early processing in living visual systems is of this nature. For example, many visual cortex neurons compute low-order directional derivatives of the (multi-scale blurred) image $L(x)$ [5]. Thus, local orientation tuning is introduced very early on. More generally, linear or non-linear combinations of the derivatives at nearby positions in space-time, or nearly corresponding positions in the two eyes, then produce a collection of functionals $M_{v,x}(L)$, parametrized by the position x and attribute value v for which their sensitivity to L is maximal. For our present purpose, the input L may be assumed fixed, so the measurements can be denoted simply by $M(v,x)$.

2.2 The Geometric Structure of Blurred Relations

The central idea behind the proposed models is to represent the data by a "blurred relation" between attribute values $v \in V$ and positions $x \in X$, instead of by a function $v(x) : X \to V$. The notion of a blurred relation generalizes the classical notion of a (crisp) relation in set theory. There, a relation $\mathcal{R}(a,b)$ between $a \in A$ and $b \in B$ is identified with a subset of the Cartesian product $A \times B$. In our model, the factors in the Cartesian product are the set X of position indices x and the set V of attribute indices v. The first generalization is to replace the usual $\{0,1\}$-valued subset indicator function by a real-valued one; in our case, the measurement values $M(v,x) \in \mathbb{R}^+$. A natural notion of "complement" no longer exists, unlike in the "fuzzy set" version of relations, where the indicator takes values in the closed unit interval $[0,1]$. The motivation for this asymmetry is the easy visual detectability of peaks, but not dips, in the distributions of velocity- or disparity-attributes of a stimulus. Formally though, one could probably also construct a fuzzy set analogue of our proposed model structure.

The next vital steps in constructing the model introduce the possibility of blurring and differentiating $M(v,x)$ in a way that makes sense with respect to the coherence in the data. All the grouping and shape extraction operations to be proposed rely on these notions. The required structure is a smooth geometry on the set of detector indices (v,x), with at least the appropriate local and global topology, as well as the notion of a connection of at least the affine type, embodied in covariant derivative operators.

In artificial vision, this would be done by setting up appropriate wiring or a data-structure, calibrating the sensor geometry and responses, and defining blurring and derivative operators. In living visual systems, the structure has to

arise autonomously. This is not the place to explore exactly how this happens in nature, but it is worth noting a few aspects that are relevant to the choice of an appropriate mathematical structure for the model.

First of all, the generalized receptive fields of the detectors, indexed by (v, x), should pave the intended product space $V \times X$ in an overlapping manner in order to establish the required topology on what would otherwise just be a product of unstructured index sets. The natural measure of overlap is the correlation between "nearby" $M(v, x)$, and this quantity is likely to modify the neural wiring. The modular structure of visual cortex [11] suggests that the x and v ordering are more or less decoupled initially. Within a "hypercolumn" module, all receptive fields are overlapping in space, but all orientations, etc. are represented. The spatial extent of the image is then covered by a large collection of such modules.

It may thus be assumed that the x-ordering defines a 2-manifold X, and that at each point $p \in X$, the v-ordering creates a manifold V_p, with all V_p assumed diffeomorphic to a typical V. Examples of V would be (topologically at least) $\mathbf{S}^1$ for orientation, or the open 2-disk $\mathbf{D}^2$ for motion and —perhaps— for binocular disparity. The latter attribute may in fact be 1-dimensional, since human sensitivity to vertical disparity seems to be weak.

The function $M(v, p) : V_p \rightarrow \mathbb{R}^+$ at point p then is an element of some function space $\mathcal{M}_p$, with all $\mathcal{M}_p$ diffeomorphic to a typical $\mathcal{M}$. For the systems modelled here, the dimension of $\mathcal{M}$ exceeds that of V considerably. Taking the example of orientations, V is obviously 1-dimensional, but the dimension of $\mathcal{M}$ can be estimated as somewhere in the range 10-15, based on measured tuning-widths of orientation detectors [11].

At this point, one has a fibre bundle with base space X and typical fibre $\mathcal{M}$. Thus, *locally* the intended structure exists, but it is unclear how a self-organizing system could ensure that the *global* structure is as intended (that is "trivial"). Of course, taking X to be, say, a 2-disk $\mathbf{D}^2$ guarantees that the bundle is globally trivializable. Yet, any realistic process would take a very long time to actually equalize the fibers globally using only local signal correlations. Moreover, the retinal blind spot introduces a fundamental problem. With annular X, globally non-trivial bundle structure is possible, even likely. For example, orientation may then have a non-zero winding number on paths around the hole. No mechanism using only local correlations and local wiring modifications could avoid getting trapped in such structures.

Such topological defects become evident only under large displacements of a physically constant visual pattern. This is precisely what occurs in saccadic eye-movement. If —as seems likely— insufficient genetic information is available for specifying the required topology, then the neural wiring will have to self-organize using the saccadic image transformations as an error-check. Perceived constancy of the world across saccades (as far as this occurs) implies that the fibre bundle is trivialized to a global Cartesian product $\mathcal{M} \times X$. This also fixes $V \times X$ as the global structure of the set of (v, x), and calibrates $M(v, x)$ across X.

What is still missing is the ability to compare $M(v, x)$ across V. For the attribute types mentioned, each value v is associated with a vector in (the tangent of) X, so the gauging over V is constrained by that over X. In any event, the

system may manipulate objects while viewing them, and require the percept to be transformed accordingly. It has to rotate objects to compare across orientations, change viewing-distance to do so for disparity, and scan the eyes across objects to compare across velocities.

The system may even transcend the affine type of connections which suffice for defining derivatives and blurring. To realize this, the system must require triviality of the holonomy for self-induced image transformations which correspond to parallel transport, that is, its percepts should remain constant after traversing any such loop. This constrains the connections to be flat, so V and X are Euclidean, and one may transform the (v, x) coordinates to form a Cartesian system. The assumption that the human visual system uses a flat geometric structure cannot be spectacularly wrong since we do not notice non-trivial holonomies in our daily life. Yet, it is of great interest to explore the matter via serious experiments.

For the time being, I shall simply assume a Euclidean structure for the blurred relations models to be analyzed below.

3 Grouping and Splitting; Some Illustrative Examples

Blurring $M_\sigma(v, x) = G_\sigma \star M(v, x)$ with a Gaussian G_σ already realises a very useful operation on the data. Loosely speaking, the result is a grouping based on proximity in position as well as attribute value. Some blurring is inherent already in the finite spatial and attribute resolution of the detection stage. Of course, the blurring may be anisotropic, inhomogeneous, or both. As an extreme example, blurring only in x, not in v, leads to a representation similar to a "sliding" histogram. Actually, the assumption used here is that the blurring can be made isotropic and homogeneous (with $\sigma = 1$, say) by a suitable smooth transformation of (v, x). The index σ will then be dropped.

The properties of the distribution $M(v, x)$ have been analyzed for some simple but important examples, which would pose very difficult problems for more standard models. For simplicity only, the input is often taken to have structure only in one x and one v dimension. Also, we ignore any boundaries of X, and the possibly periodic global structure of V (for example for orientations).

3.1 Smooth, Opaque or Transparent Attribute Fields

An image characterized by an ("idealized") single-valued and smooth attribute value $s(x)$ will be represented as the blurred graph of $s(x)$:

$$M(v, x) = \delta[v - s(x)] \star e^{-(v^2+x^2)/2} \approx \exp\{-\frac{[v - s(x)]^2}{2[1 + (\partial_x s(x))^2]}\} \ .$$

The approximation above requires $\partial_{xx} s(x) \ll 1$. Evidently, some v-resolution is lost when $s(x)$ has a large gradient, but it will be shown below that discontinuities are handled quite sensibly.

Transparency (that is multi-valued $s(x)$) is represented just as naturally. The only new feature is that $M(v,x)$ becomes multi-modal in v. The branches interact negligibly as long as their distance Δ is well above unity. At $\Delta = 1$, two parallel branches fuse into one (slightly "thick") branch.

Missing data are represented trivially as $M(v,x) = 0$ within the gap. More generally, the intensity scale of $M(v,x_0)$ reflects the strength of evidence for the attributes at x_0. Models based on function reconstruction often confuse gaps with $v = 0$ patches, which can lead to large biases, unless extraneous information is used to limit the damage.

3.2 Filling-in, Extrapolation and Capture

Gaps can occur in image data because of partial occlusion or missing local detectors (blind spot, retinal blood vessels). In any case, it may be necessary to interpolate somehow across a gap ("filling-in"). It seems advisable to perform such operations on the level of local attributes, rather than luminance values. Indeed, this seems to happen in humans.

For example, assume a gap at $|x| < \lambda/2$ in data $s(x) = 0$, say. The ensuing representation becomes

$$M(v,x) = e^{-v^2/2}\{1 + \frac{1}{2}[\mathrm{erf}(\frac{x-\lambda/2}{\sqrt{2}}) - \mathrm{erf}(\frac{x+\lambda/2}{\sqrt{2}})]\} \ .$$

Small gaps, say with $\lambda < 1$, are essentially filled in. At the midpoint of the gap, $M(0,0) = 1 - \mathrm{erf}[\lambda/(2\sqrt{2})] = 1 - \lambda/\sqrt{2\pi} + O(\lambda^3)$. Thus, for any reasonable choice of threshold, gaps are only encoded in units smaller than the gap diameter.

On the standard (=unit) scale, one has essentially the case of an "open boundary" for all $\lambda \gg 1$. The unit-size "halo" of M activation which extends into a gap can be revealed perceptually if v-noise is added in the gap. Those noise components that lie in the halo are then "captured", that is they are seen as if they were attached to the edge of the object.

3.3 Jump Discontinuity: Smoothing versus Segmentation

The stimulus is now characterized by $s(x) = \Delta\ \mathrm{sgn}(x)$, and its representation becomes:

$$M(v,x) = F(x)\, e^{-(v-\Delta)^2/2} + F(-x)\, e^{-(v+\Delta)^2/2} \ ,$$

where $F(x) = (\sqrt{2\pi})^{-1} \int_{-x}^{\infty} \exp[-z^2/2]\mathrm{d}z = \frac{1}{2}[1+\mathrm{erf}(x/\sqrt{2})]$.

There are two regimes, separated by the condition $\Delta = 1$, in which $M(v,x)$ near the jump has a qualitatively different shape. Analytically, the distinction is evident in the behaviour of $M(v,x)$ near the origin ($|x| \ll 1$ and $|v| \ll \Delta$):

$$M(v,x) = e^{-\Delta^2/2}\{1 + (\Delta^2 - 1)\frac{v^2}{2} + \sqrt{\frac{2}{\pi}}\Delta x v + O(x^n v^{3-n})\} \ .$$

The origin is always a saddle point ($\Delta \neq 0$) with the midline $v = 0$ as one of its two crossing contours where $M(v,x) = M(0,0) = e^{-\Delta^2/2}$. The important point to note is that the other contour emerges from the saddle as

$$x = \sqrt{\frac{\pi}{2}}(\Delta^{-1} - \Delta)\frac{v}{2} \ .$$

Accordingly, in the regime $\Delta < 1$, $M(v,x)$ is *unimodal* in v at any x, and the position of the mode $\hat{v}$ is a smooth function of x. Perceptual confusion with a smooth-step stimulus should occur. For $\Delta \ll 1$, one finds $\hat{v}(x) = \Delta\ \mathrm{erf}(x/\sqrt{2})$, the same as in function-fitting models that apply Gaussian smoothing to samples $v(x)$.

In the regime $\Delta > 1$, the M-contours that pass near the origin are folded back along x. Then $M(v,x)$ becomes *bimodal* in v near the jump, as if the stimulus were locally *transparent*. The modes $\hat{v}$ remain very close to their asymptotes $\pm\Delta$ right up to the jump, unless one is in the critical regime $\Delta = 1 + \epsilon$. Away from the jump, one branch quickly dominates the other. It may also be shown that the position of the jump can be detected with hyperacuity (threshold < blurscale) for $\Delta > 1$ jumps, whereas this ability degrades with jumpsize in the $\Delta < 1$ regime.

Very similar behaviour emerges if the jump is not truly discontinuous, but has a width $\lambda < 1$, or when an equally narrow, empty, or noise-filled gap occurs in the data at the jump location.

3.4 Interleaved Patterns: Merging into Transparent Planes

Recall the situation in which a complicated, patchy occluder is in front of a continuous background pattern. The contributions of the two sources become spatially interleaved in the image. The goal is to separate and re-group these contributions appropriately. Here, I shall study the simplest possible example which contains the essence of the problem: Two "layers", each with a constant but distinct attribute value, which are multiplexed spatially in a periodic stripe pattern. Thus, the stimulus has square-wave modulated attribute values: $s(x) = \Delta\mathrm{sgn}[\sin(2\pi x/\lambda)]$.

We already know the behaviour of isolated jumps ($\lambda \gg 1$), but new phenomena occur in the regime $\lambda < 1$. For any Δ, the representation can be written in the form:

$$M(v,x) = \frac{1}{2}[1 + M_\lambda(x)]\, e^{-(v-\Delta)^2/2} + \frac{1}{2}[1 - M_\lambda(x)]\, e^{-(v+\Delta)^2/2}$$

with a small $M_\lambda(x) \approx \frac{4}{\pi}\exp[-2\pi^2/\lambda^2]\sin(2\pi x/\lambda)$. Again, one can distinguish two subregimes, depending on Δ.

For $\Delta < 1$, our model is again equivalent to smoothing of $v(x)$. Then $M(v,x)$ is unimodal in v, with modal value $\hat{v}(x) \approx \Delta M_\lambda(x)$. In addition to this residual $\hat{v}$-modulation, the effective v-width of the $M(x,v)$ pattern increases to $\approx \sqrt{1+\Delta^2}$.

For $\Delta > 1$, the $M(v,x)$ distribution will split in the v-direction, whereas the high-v and low-v "segments" must still be merged spatially since $\lambda < 1$.

As a result, two transparent planes are formed. Indeed, one finds analytically that $M(v,x)$ is bimodal in v at any x, with modes at $\hat{v} = \pm\Delta(1-\epsilon)$. The quickly vanishing error $\epsilon \approx \exp[-2\Delta^2]$ is due to residual overlap of the blurred representation of the planes.

4 Extracting Geometric Information

Many aspects of the 3-D geometry of the world and of an observer's path through it are reflected in the 2-D structure of visual attribute fields of the kind we have been discussing. Questions of how to use this structure for segmenting scenes and extracting relevant information about paths and shapes have so far been posed in the classical setting of single-valued, non-interleaved fields. Within this setting, much progress has been made [6] in understanding how the differential structure of smooth optic flow (or binocular disparity) fields [4] can be exploited. The robustness of natural visual systems in dealing with superimposed or interleaved attribute fields invites an extension of such studies to the present setting. Below, I shall sketch how this can be done.

The basic approach is to extend the blurred derivative methods [5] normally applied to $L(x)$ to $M(v,x)$. To prevent an explosive growth in the number of quantities to be computed, it may be necessary to trade off the resolution (in x and/or v) against the size of the repertoire of operators. Living visual systems indeed show a progressive loss of spatial resolution in higher stages of processing.

Before the differential structure of $M(v,x)$ can be used, one has to remove its dependency on spatially varying contrast. Part of this problem is tackled by "gain control" at the L level, but the remaining attribute-specific contrast has to be handled at the $M(v,x)$ level. It is reasonable to assume that one can, at least approximately, factor the (noise-free) response as $M(v,x) = \gamma(x)J(v,x)$, where $J(v,x)$ then encodes the V-attribute structure independently of the V-specific contrast function $\gamma(x)$. Ideally, only $J(v,x)$ must be used in extracting object-specific information, but one should also avoid using data from regions where $\gamma(x)$ is smaller than the noise level of the initial stages. One obvious way to proceed is to estimate $\gamma(x)$ from the data, and use it as a gain-control signal. This is equivalent to the "divisive" inhibition [10] found throughout the visual system. The simplest formulation of such a scheme is $N(v,x) = M(v,x)/[\epsilon + \int M(v,x)\mathrm{d}v]$, where ϵ is a small bias which prevents undue amplification of the noise where $\gamma(x)$ is small. A possible variation is the use of a spatially more blurred version of $M(v,x)$ in the integral.

4.1 Extracting Boundary Location and Shape

Segmenting an image representation into smooth patches with well-localized boundaries is an obviously useful step towards object extraction. The natural notion of grouping and splitting of the data which emerges within the present model considerably simplifies the task of locating patch boundaries. Moreover, the representation allows easy access to local boundary shape, which is an important input to object shape mechanisms since object contour shape constrains the

possible 3-D shape of the surface near the rim [2]. For example, contour convexity implies surface convexity, and contour concavity implies surface hyperbolicity near the rim [3].

The simplest useful boundary operator is $B(v,x) = \nabla_x N(v,x)$, the projection of the true gradient ∇N on the position space X. For the usual 2-dimensional X, one needs to compute at each position just a pair of directional derivatives $(\partial_{x^a}, \partial_{x^s})$ in transversal (preferably orthogonal) directions. Nature seems to prefer a somewhat richer sampling of directions.

Maximal sensitivity of $|B(v,x)|$ occurs for reasonably large gaps ($\lambda > 1$) or jumps ($\Delta > 1$), and for those cases it is virtually independent of the structure of the data in the adjoining patch(es). For the (1+1)-dimensional toy-model of a jump $s(x) = \Delta \mathrm{sgn}(x)$, one gets $\sqrt{2\pi}B(v,x) = \exp[-(x^2 + (v-\Delta)^2)/2] - \exp[-(x^2 + (v+\Delta)^2)/2]$. Thus, sensitivity decays for jumps (or gaps) smaller than the blurscale. In addition, a smoothly varying $s(x)$ causes some $B(v,x)$ responses, but these are much smaller than the responses on jumps or gaps.

Note the consistency of $B(v,x)$ with the representational structure of the model. At each point p, one has a (vector-valued) distribution representing the strength $|B(v,p)|$ and the inward normal direction $B/|B|$ of the "boundariness" for each attribute value. The advantages of this scheme are that each of several overlayed attribute fields may have independent boundaries, and that the boundary of a patch of any attribute structure always "belongs to" a well-defined patch, namely the one that extends from the boundary into the direction of $B(v,x)$. The simplest example occurs in the representation of a jump discontinuity ($\Delta > 1$): Two distinct, but spatially superimposed boundaries are signalled by $B(v,x)$, each of which "points at" the half-space that it bounds.

The local curvature of a boundary can be obtained by comparing pairs of unit normal vectors along its length, or by the method proposed by [7] for curves in the L domain. The latter method therefore only uses the information in $|B(v,x)|$. In the present case, the former scheme may be preferred because it requires only one vector subtraction instead of three more orders of differentiation. The unequivocal assignment of boundaries to their own patches is particularly useful here, since the surface shape constraint given by contour curvature can now be made to propagate unilaterally from a boundary onto its proper patch.

4.2 Extracting Within-patch Shape Characteristics

In the classical setting of single-valued attribute fields as generated by smooth, opaque object surfaces, the first-order differential structure of, for example, disparity or motion vector fields contains information about the relative attitudes and motions of local surface elements with respect to the observer [4]. Likewise, the second- (plus first-) order structure determines local surface shape up to a depth scaling [6].

Not much is known about how complete and accurate the use of this information is, nor what the relevant neural mechanisms are for extracting it. Necessarily then, these last paragraphs are somewhat speculative.

One cannot expect to be able to extract much relevant structure from data that are very diffuse in V. This suggests computing operators such as $D_0(v,x) = \nabla_v^2 N(v,x)$. In fact, a more physiologically plausible and functionally useful variant would be $D^+ = \max(0, D)$. This is most sensitive to a narrow v-mode $\hat{v}$ which is (locally) constant along x. The natural extension then is to compute similar functionals $D_A(v,x)$, which differ from $D_0(v,x)$ merely by a rotation with respect to the (v,x) coordinates, such that $D_A(v,x)$ is maximally sensitive to narrow ridges with $\hat{v} = v_0 + Ax$. The response distribution over an appropriate range of A corresponds to measuring a field of first-order *contact elements* to the possibly multi-valued attribute field. The second-order structure could be extracted either by pairs of first-order responses at small spatial offsets, or directly by means of second-order contact elements.

Near boundaries, all such measures will show artifacts. Yet, this failure can be cured nicely by suppressing the $D_A(v,x)$ outputs for all A by means of the output of the nearby boundary signals $|B(v,x')|$ that "point" from x' to x. As mentioned previously, these boundary signals can take over the role of constraining the surface shape near the rim.

References

1. Blake, A., Zisserman, A. (1987). Visual Reconstruction, MIT Press, Cambridge (MA).
2. Koenderink, J.J. (1984). What does the occluding contour tell us about solid shape?, Perception, 13, pp. 321–330.
3. Koenderink, J.J. (1990). Solid Shape, MIT Press, Cambridge (MA).
4. Koenderink, J.J., van Doorn, A.J. (1975). Invariant properties of the motion parallax field due to movement of rigid bodies relative to an observer, Optica Acta 22, pp. 773–791.
5. Koenderink, J.J., van Doorn, A.J. (1987). Representation of local geometry in the visual system, Biol. Cybern. 55, pp. 367–375.
6. Koenderink, J.J., van Doorn, A.J. (1992). Second-order optic flow, J. Opt. Soc. Am. A 9, pp. 530–538.
7. Koenderink,J.J., Richards, W. (1988). Two-dimensional curvature operators, J. Opt. Soc. Am. A 5, pp. 1136–1141.
8. Poggio,T., Torre, V., Koch, C. (1985). Computational vision and regularization theory, Nature 317, pp. 314–319.
9. Snippe, H., Koenderink, J.J. (1992). Discrimination thresholds for channel-coded systems, Biol. Cybern. 66, pp. 543–551.
10. Snowden, R.J., Treue, S., Erickson, R.G., Anderson, R.A. (1991). The response of area MT and V1 neurons to transparent motion, J. Neurophys. 11, pp. 1768–2785.
11. Spillman, L., Werner, J.S. (1990). Visual Perception: The Neurophysiological Foundations, Academic Press, San Diego.

Contour Texture and Frame Curves for the Recognition of Non-Rigid Objects

J. Brian Subirana-Vilanova

Artificial Intelligence Laboratory, Massachusetts Institute of Technology
545 Technology Square, Cambridge, MA 02139, USA
Email: brian@ai.mit.edu

Abstract. An oak leaf can visually be easily distinguished from an elm leaf; yet the same oak leaf cannot be distinguished from other oak leaves unless a detailed inspection is performed. This paper presents a filter-based scheme for the recognition of non-rigid objects, such as leaves, proposing a two-level representation based on two novel notions, frame curve and contour texture. Examples of contour textures include many complex object boundaries such as leaves, clouds, forests, complex tools, and city skylines.

Contour texture is defined so that it is similar for leaves of the same type and different across leaf types. The nature of the contour texture of a curve and its relation to two-dimensional (2-D) texture is discussed, which it is contended should be thought of as a separate concept. Several applications are suggested and results of an implemented filter-based scheme are given.

Keywords: non-rigid objects, contour texture, frame curve, texture, filter, shape description, human perception.

1 Introduction

The visual recognition of non-rigid objects has received very little attention in the past (see [21] for a review). This paper examines so-called contour textures, a certain type of non-rigid objects. A two-level shape description for contour textures is proposed and an implemented filter-based scheme is suggested.

Oak leaves are readily distinguishable from other types of leaves (see Fig. 1). The ability to distinguish leaves cannot be attributable to an exact shape property since the leaf contours change significantly from one leaf to another. Instead, another property, more *statistical* in nature, must be used. This property can be called *contour texture*; this term will be used to refer to the contour property as well as the object itself. Contour texture has received very little attention in the past and is the subject of this paper.

Contour texture is interesting because there are many non-rigid or complex objects with distinctive contour textures (such as clouds, trees, hair, and mountains) and because images without contour texture appear less vivid and are

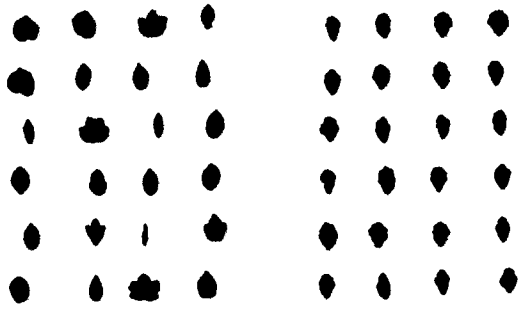


Fig. 1. Which of these leaves are oak leaves? Some objects are defined by the contour textures of their boundaries.

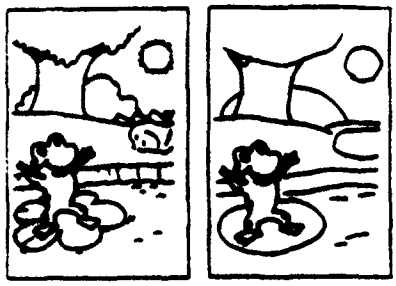


Fig. 2. These two images are identical with the exception that one of them has been drawn by removing the contour texture of its curves. The image without contour texture appears less vivid and there is more ambiguity in identifying the origin of the different contours in the image. In other words, significant information is lost when the contour texture of a contour is replaced by its frame curve.

harder (or even impossible) to recognize (see cartoons in Fig. 2). Rigid-object recognition schemes do not handle contour textures because they rely on "exact" shape properties. Contour textures (and other non-rigid objects) cannot be recognized by matching a pictorial version of the shape since the shape changes from one instance to another. Contour texture is very common in classification problems; however, it may also be used in the recognition of rigid objects as an indexing measure, especially if the shapes are complex such as skylines of towns.

Fig. 3. This figure illustrates some of the applications of contour texture. (left) Contours can be grouped based on their contour texture. (right) The contour texture can be used as a powerful indexing measure in large databases of objects.

In addition, contour texture may help perceptual organization and indexing schemes (see Fig. 3). In this paper a filter-based model is proposed that can be

used for contour texture recognition, segmentation, and indexing.

Not all objects can be described just by their contour textures. In fact, leaves are a good example of this. Botanists have divided leaves using several attributes but only one is based on contour texture (they use the term "leaf margin"). This attribute generates several classes such as dentate, denticulate, incised or serrulate. Botanists also use another attribute (called "leaf shape") which is complementary to contour texture. Leaf shape categories include oval, ovate, cordate or falcate (see [17] for a complete list). The distinction between contour texture and shape is particularly important for deciding what type of representation to use, a question which will be addressed in this paper.

Of particular interest is the search for a useful and computable shape representation for contour textures. The findings presented in this paper argue in favour of a two-level representation for contour textures such that one level, which is called the frame curve, embodies the "overall shape" of the contour and the other, the contour texture, embodies more detailed information about the boundary's shape. These two levels correspond closely to the two above attributes used to describe leaves by botanists.

The notion of contour texture prompts many other questions as well: Can we give a precise definition of contour texture? What is the relation between two-dimensional texture and contour texture? Is there a computationally-efficient scheme for computing contour texture? There are several factors that determine the contour texture of a curve, for example, the number and shape of its protrusions, but other factors influence the contour texture of a shape? Does shape influence contour texture?

The rest of the paper addresses these questions and is organized as follows: In the next section we discuss the definition of contour texture and its relation to 2-D texture. In the following two sections the relation that contour texture has with scale and inside/outside relations are discussed respectively. Lastly, in Sect. 5 an implemented filter-based scheme for contour texture is presented.

2 Contour Texture and Frame Curves

2-D texture has received considerable attention, both in the computational and psychological literature. However, there is no unique definition of it. Roughly speaking, 2-D texture is a statistical measure of a *two-dimensional region* based on local properties. Such properties typically include orientation and number of terminations of the constituent elements (also known as textons).

In this paper it is proposed that contour texture, a related but *different* concept, plays an important role in human perception. Contour texture can be defined as a statistical measure of a *curve* based on local properties (see Fig. 4). Such a curve is called the *frame curve*. The notion of a frame curve as presented here is closely related to the one presented in [19]. A frame curve is defined there as *a virtual curve in the image which lies in "the centre" of the figure's boundary.* In the context of this paper the whole contour texture is the figure, defined in [19] as the collection of image structures supporting visual analysis of a scene.

Fig. 4. (left) An image with a vertical two-dimensional texture discontinuity. The discontinuity is defined by the average orientation of the segments near a point in the image. Such orientation is different for the two regions surrounding the central vertical line. (right) The tilted segments in this image define a horizontal line. A contour texture discontinuity in such a line is perceived in the middle of it. The discontinuity is defined also by the average orientation of the segments surrounding a point. One of the differences between contour texture and two-dimensional texture is that the statistics are computed over a curve in one case and on a two-dimensional region in the other. Other differences and similarities are discussed in the text.

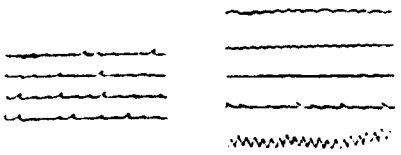


Fig. 5. (left) Different curves with similar contour texture. The frame curve in all these cases is a horizontal line. (right) Some curves with different contour textures.

A frame curve can also be used as a "non-circular" topological obstruction to extend size functions [25] or to compute a part-description of a shape (see [19]).

Figure 5 shows some contours with different contour textures, all of which have "invisible" horizontal lines as frame curves. The contours were drawn by an implemented contour texture generator which takes as input a sample drawing of a *COntour Texture ELement* or *Cotel* (akin to texton and protrusion) and produces as output a concatenation of one or more of these cotels subject to certain random transformations.

3 Inside/Outside and Convexity

There are several factors that determine contour texture. In this section it is argued that the side of the contour perceived as inside influences contour texture perception. Consider the examples in Fig. 6. The left and right stars in the third row have very similar outlines since one is a reversed version of the other. By "reversed" we mean that a mirror image of one of the two contours around the frame curve yields the other. They are partially smoothed versions of the centre star but each of them looks very different from the others; in fact, the left one is more similar to the centre star [19] despite the fact that both have the same number of smoothed corners. Subirana-Vilanova and Richards [19] made this observation and proposed that it is due to a bias which makes the outside of the

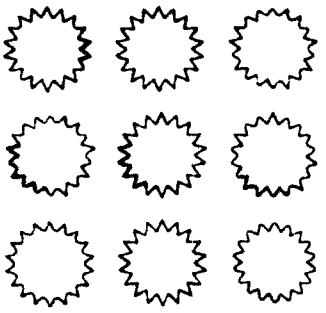


Fig. 6. (top row) Use the middle pattern as reference. Most see the left pattern as more similar to the reference. This could be because it has a smaller number of modified corners (with respect to the centre) than the right one, and therefore, a pictorial match is better. (second row) In this case, the left and right stars look equally similar to the centre one. This seems natural if we consider that both have a similar number of corners smoothed. (third row) Most see the left pattern as more similar despite the fact that both, left and right, have the same number of smoothed corners with respect to the centre star. Therefore, in order to explain these observations, one cannot base an argument on just the number of smoothed corners. The positions of the smoothed corners need be taken into account, i.e. preferences are not based on just pictorial matches. Rather, here the convexities on the *outside* of the patterns seem to drive our similarity judgement. (These figures were taken from [19].)

shapes more "salient". In the context of this paper, the findings of [19] imply that the contour texture of a shape depends on which side is perceived as inside.

4 The Role of Scale and Contour Complexity in the Distinction of Shape and Contour Texture

As mentioned in the introduction, the notion of contour texture is meant to be used in the differentiation of shapes belonging to different perceptual categories (e.g. an oak vs an elm leaf) and not to distinguish shapes belonging to similar perceptual categories (e.g. two oak leaves). This raises the following questions: Are two types of representations (shape and contour texture) necessary? When are two objects in the same category? When is a contour texture description appropriate? An answer to these questions will be given later in the paper when an implemented contour texture scheme designed to determine contour similarity based on contour texture will be presented.

In this section it is argued that the difference between shape and contour texture is relevant to computer vision (regardless of implementation details) and, in particular, that it is important to find schemes which automatically determine whether a shape is a contour texture or not. For contour textures it is also important to embody both representations (shape *and* contour texture) for every image contour. One of the strongest arguments was presented at the beginning of this paper: some shapes cannot be distinguished by exact shape properties (while others can).

Fig. 7. This figure provides evidence that for simple objects like the top one (left), the matching across scales is done pictorially (see second row, left). For more complex shapes, on the other hand, such as the one on the right, the matching is performed by maintaining the contour texture description (see lower row). See text for details.

Three other psychological observations that support this difference will now be presented.

First, studies with pigeons have shown that they can discriminate elements with different contour textures but have problems when the objects have similar contour textures [8, 4]. This suggests that different schemes may be needed for the recognition of shape and contour texture.

Second, consider the object in Fig. 7 top left. Below the object, there are two transformations of it: the left one is a pictorial enlargement, and the right one is an enlargement in which the protrusions have been replaced by a repetition of the contour (preserving the contour texture). The shape on the left appears more similar to the one on the right [19]. It is contended that this is true in general if the shapes have a small number of protrusions (i.e. their "complexity" is low). In these cases, contour texture does not seem to have an important role in their recognition. However, when the shapes are more complex (see Fig. 7, three shapes on the right), the similarity is not based on an exact pictorial matching. Instead, the enlarged shape with the same contour texture is seen as more similar [21]. For "complex" shapes, the visual system tends to abstract the contour texture from the shape and the "enlargement" of such a property is done at a symbolic level. In addition to supporting the distinction between contour texture and shape (first question above), this observation suggests that *complexity and scale play a role in determining what type of description (shape or* contour texture*) should be used in each case: simple shapes are fully represented and complex ones are represented just by abstract contour texture descriptors.* [7], [14], and [11] each performed similar experiments on a two-dimensional texture version of the problem. [7] also presents some one-dimensional contour-texture-like examples (using the notions presented here) which support the role of omplexity described above. Note that contour texture may not play an important role in the recognition of simple shapes but may be used as an indexing property of the shapes.

The third study which agrees with the distinction made between contour texture and shape is that of Rock, Halper, and Clayton [16]. They showed subjects a complex figure and later showed them two figures which had the same

overall shape and contour texture (using the terms defined here), but only one of which was exactly the same. The subjects had to find which was the previously seen shape. They found that subjects performed only a little better than random. This suggests, again, that they were just remembering the overall shape and an abstract description of the contour texture of the boundary's shape. When subjects were presented with non-complex versions of the same shapes, the distinctions were based on the exact shapes themselves, which agrees with the model given above.

5 A Filter-Based Scheme

The definitions of contour texture and two-dimensional texture, given in Sect. 2, point out some of the relationships between them: both notions are based on statistics of local properties, but they differ in the extent of such statistics — a curve for contour texture and a surface for two-dimensional texture. In fact, most existing schemes for two-dimensional textures can be applied, after some modifications, to contour texture. Some of the problems that have to be solved in doing so are the computation of frame curves and inside/outside relations.

Many theories of two-dimensional texture exist, but just a few will be mentioned. Preattentive texture discrimination has also been attributed to differences in nth-order statistics of stimulus features such as orientation, size, and brightness [10, 9, 2, 27]. Other theories have been proposed, especially ones that deal with repetitive textures (textures in which textons are similar and on a regular pattern), such as Fourier Transform based models [1] and histogramming of displacement vectors [23]. All of these theories tend to work well on a restricted set of textures but have been proved to be unable to predict human texture perception sufficiently accurately in all of its spectrum. In addition, it is unclear how these schemes could compute frame curves or inside/outside relations, specially in the presence of fragmented and noisy contours.

Another popular approach has been to base texture discrimination on the outputs of a set of linear filters applied to the image (see [24, 6, 15, 13, 3, 22]). These approaches differ among themselves on the set of selected filters and/or on the *required* post-processing operations. A purely linear scheme cannot be used (see for example [13]), justifying the need for non-linear post-processing operations. Malik and Perona [13] compare the discriminability in humans to the maximum gradient of the post-processed output of the filters they use and find a remarkable match among them. The approach is appealing also because of its simplicity and scope and because it is conceivable that it may be implemented by cortical cells.

Some work exists on curve discrimination which could be applied to contour texture discrimination, however previous approaches are designed to process fully connected curves [26, 5, 12, 25]. The present model, instead, works directly on images and does not require that the contour be fully connected. The ability to process the contour directly on the image enables the scheme to naturally extend to fragmented curves and to curves without a single boundary (e.g. a contour composed of two adjacent curves).

This scheme segments and recognizes the curves based on their contour texture and consists of the following steps:

1. Find the *frame curves* of the contour to be processed.
2. Decide which is the inner side of the contour and colour (label) it.
3. Filter the image I with a set of oriented and unoriented filters F_i at different scales, which yields $I * F_i^+$ and $I * F_i^-$, the negative and positive responses to the filters.
4. Perform nonlinear operations on the outputs obtained, such as spreading the maxima and performing lateral inhibition (in the current implementation).
5. Normalize the orientation of the directional filters to the orientation of the frame curve's tangent.

Contour texture discontinuities can be defined as places of maximum gradient (along the direction of the frame curve) in the obtained responses, and recognition can be done by matching such responses. Steps 3, 4, and 5 have been implemented on the Connection Machine and tried successfully on a variety of segmentation examples (see Fig. 8). Filters were used similar to those which

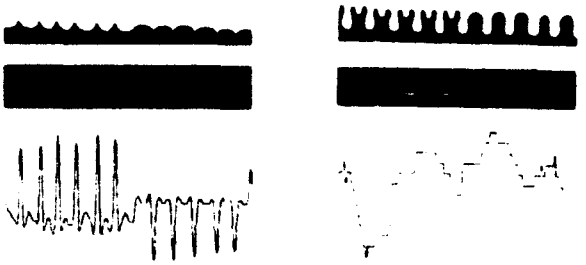


Fig. 8. (left) Coloured contour. (second row) Normalized output of the selected filter. (third row) Cross-section of the post-processed filter output along the *frame curve*, horizontal in this image. (right) same as above for another contour. However, in this case step 4 was omitted.

Malik and Perona [13] used in the context of 2-D texture; it is unclear though whether only even-symmetric filters are needed as they proposed.

5.1 Computing Frame Curves

Finding the frame curve is the hardest part and is not fully implemented. A possible solution involves smoothing [18] but has problems with the often-occurring complex or not fully connected curves.

However, frame curves tend to lie in the ridges of one of the filter's responses. This suggests that frame curves can be computed by a ridge detector that can locate long, noisy, smooth ridges of variable width in the filter's output. One such approach was presented recently [20]. Note that computing ridges is different from finding discontinuities in the filter's response, which is what would be used to compute two-dimensional texture discontinuities in the schemes mentioned above.

5.2 Colouring

Step 2, colouring, is needed to account for the dependence of contour texture on the side perceived as inside, as discussed above (see Fig. 8). Colouring may also be useful in increasing the response of the filters. Colouring runs into problems if the contour is not fully connected or if the inner side of the contour is hard to determine. Possible alternatives include using the frame curve as a basis to spread and stop the colouring, and enlarging the contour's width to increase the filter's response.

6 Conclusion

Contour texture has received very little attention in the past but it is suggested that it plays an important role in visual perception and, in particular, in the shape recognition of some *non-rigid or complex objects* and possibly in *grouping, attention, indexing, and shape-from-contour*. It is also proposed that complex contours (i.e. non-smooth or disconnected) be represented by abstract contour texture descriptors, while simple ones be represented by the detailed location of the contour's points.

A filter-based approach to contour texture is simple and yields useful results in a large number of cases. Different techniques for colouring have been described and it is recommended that a *ridge detector* be used to find *frame curves* in the output of the *filters* of the present model.

References

1. Bajcsy, R. (1973). Computer description of textured surfaces, Proc. IJCAI, pp. 572–579.
2. Beck, J. (1982). Textural segmentation. In: Beck, J. (ed.), Organization and Representation in Perception, Erlbaum, Hillsdale NJ, Chapter 15.
3. Bovik, Clark, Geisler (1990). Multichannel texture analysis using localized spatial filters, IEEE Trans. on Pattern Analysis and Machine Intelligence 12, pp. 56–65.
4. Cerella, J. (1982). Mechanisms of concept formation in the pigeon. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (eds.), Analysis of visual behavior, MIT Press, Cambridge, MA, pp. 241–260.
5. Dudek, G. (1993). Shape description and classification using the interrelationship of structures at multiple scales, this volume, pp. 473–482.
6. Fogel, I., Sagi, D. (1989). Gabor filters as texture discriminators, Biol. Cybern. 61, pp. 103–113.
7. Goldmeier, E. (1972). Similarity in visually perceived forms, Psychological Issues 8, pp. 29–65 (originally published 1936).
8. Herrnstein, R.J. (1984). Objects, categories, and discriminative stimulus. In: Roitblat, H.L., Bever, T.G., Terrace, H.S. (eds), Animal Cognition: Proc. Frank Guggenheim Conf. Lawrence Erlbaum Associates, Hillsdale NJ, pp. 233-261.
9. Julesz, B. (1986). Texton gradients: the texton theory revisited, Biol. Cybern. 54, pp. 245–251.

10. Julesz, B., Bergen, J.R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures, Bell Syst. Tech. J. 62, pp. 1619–1645.
11. Kimchi, R., Palmer, S.E. (1982). From and Texture in Hierarchically Constructed Patterns, J. Exp. Psych.: Human Perception and Performance 8 (4), pp. 521–535.
12. Maeder, A.J. (1993). Polygonal harmonic shape characterization, this volume, pp. 463–472.
13. Malik, J., Perona, P. (1989). A computational model of texture perception, Report No. UCB-CSD 89-491, Computer science division (EECS), University of California, Berkeley, CA.
14. Palmer, S. (1982). Symmetry, transformation, and the structure of perceptual systems. In: Beck, J. (ed.), Organization and Representation in Perception, Erlbaum, Hillsdale, NJ.
15. Rentschler, I., Hubner, M., Caelli, T. (1988). On the discrimination of compound gabor signals on textures, Vision Research 28 (2), pp. 279–291.
16. Rock, I., Halper, F., Clayton, T. (1972). The perception and recognition of complex figures, Cognitive Psychology 3, pp. 655–673.
17. Smith, J.P. (1977). Vascular Plant Families, Mad River Press Inc., Eureka, California.
18. Subirana-Vilanova, J.B. (1991). On contour texture. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Ann Arbor, MI, pp. 753–754.
19. Subirana-Vilanova, J.B , Richards, W. (1991). Perceptual organization, figure-ground, attention and saliency, A.I. Memo No. 1218, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.
20. Subirana-Vilanova, J.B., Sung, K.K. (1992). Perceptual organization without edges, Proc. Image Understanding Workshop, Morgan and Kaufman, pp. 289–298.
21. Subirana-Vilanova, J.B. (1993). Machine Perception of Non-Rigid Objects. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, to appear.
22. Thau, R.S. (1990). Illuminant precompensation for texture discrimination using filters. In: Proc. Image Understanding Workshop, Pittsburgh, Pennsylvania, Morgan Kaufman Publishers Inc., San Mateo, CA, pp. 179–184.
23. Tomita, F., Shirai, Y., Tsuji, S. (1982). Description of textures by a structural analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence, 4 (2), pp. 183–191.
24. Turner, M. (1986). Texture discrimination by gabor functions, Biol. Cybern. 55, pp. 71–82.
25. Uras, C., Verri, A. (1993). Studying shape through size functions, this volume, pp. 81–90.
26. Van Otterloo, P.J. (1991). A contour-oriented approach to shape analysis, Prentice Hall, UK.
27. Voorhees, H., Poggio, T. (1988). Computing texture boundaries from images, Nature 333 (6171), pp. 364–367.

Conic Primitives for Projectively Invariant Representation of Planar Curves*

Stefan Carlsson

Computational Vision and Active Perception Laboratory
Department of Numerical Analysis and Computing Science
Royal Institute of Technology, S - 100 44 Stockholm, Sweden

Abstract. An algorithm is presented for computing a decomposition of planar shapes into convex subparts represented by ellipses. The method is invariant to projective transformations of the shape, and thus the conic primitives can be used for matching and definition of invariants in the same way as points and lines. The method works for arbitrary planar shapes admitting at least four distinct tangents and it is based on finding ellipses with four points of contact to the given shape. The cross-ratio computed from the four points on the ellipse can then be used as a projectively invariant index. It is shown that a given shape has a unique parameter-free decomposition into ellipses with unit cross-ratio.

Keywords: shape representation, projective invariance, conics, shape decomposition.

1 Introduction

The desire to achieve efficient viewpoint independent object recognition has recently led to an increased interest in projective invariance for object representation [1, 6, 7, 11, 15]. Projectively invariant descriptors of objects can be computed from relations between points, lines and conics that are coplanar on object surfaces in 3-D.

For arbitrary curved objects, projectively invariant point and line descriptions are more complex. In this case invariant points and lines can be extracted from inflexions or bitangents (e.g. [8, 11]). Related to this is the use of combined algebraic and differential invariant descriptors [14].

These methods of point and line descriptions have limitations for arbitrary curved shapes. They cannot for example be used for convex shapes. For complex shapes the number of invariants grow very rapidly with the number of points and lines used in the representation which leads to problems when the invariants are used for indexing.

* Acknowledgements: I would like to thank Lars Svensson of the Royal Institute of Technology for illuminating discussions on invariants. This work was part of Esprit Basic Research Action 6448, VIVA, with support from Swedish NUTEK.

It is therefore desirable to look for more complex primitives for projectively invariant representation of planar shape. A most natural extension of the use of lines is to use homogeneous polynomials. A homogeneous polynomial is transformed projectively into a homogeneous polynomial with the same degree. The parameters of the homogeneous polynomial can therefore be used in the same way as the coordinates of points and lines as projectively invariant shape descriptors.

The problem lies in the association of the homogeneous polynomial curve with the given shape. This association must commute with the transformation. For a restricted class of affine transformations, methods of invariant associations of homogeneous polynomials to point sets were developed in [3]. This was extended to the general affine case in [7]. The restriction to point sets and affine transformations is a limitation of this method. It requires the affine matching of the point sets used for association and, since shape data are in general given as continuous curves, this means that the point sets have to be extracted in an affine invariant manner from the curves.

In the projective case and with continuous curves instead of points the association has to be based on projectively invariant properties. Two curves are said to be in contact of order n if they coincide at a certain point and their derivatives up to order $n-1$ are the same. Contact is a property that is invariant over projective transformations. For a given shape we can consider the class of homogeneous polynomials and a certain number of contact points with a specified order of contact. A member of this class will then project to a member of the corresponding class given by the projective transform of the shape.

The method that will be presented, outlined in [5], is based on using ellipses, which are a subset of second-order homogeneous polynomials. For a given shape we will study the class of ellipses with four contact points with the shape. The order of contact is two, that is, we consider ellipses where the tangents of the four contact points coincide with the tangents of the given shape. As will be explained in the next section, the choice of four contact points and second-order contact provides a method of identifying single members of this family of ellipses over projective transforms using the cross-ratio of four points on a conic. It is thus posible to extract a finite set of ellipses from a shape in two projectively corresponding images in such a way that the ellipses in the two frames are in projective correspondence.

The proposed method of shape representation has similarities with the medial axis transform [2, 4, 12] and can in certain respects be seen as a generalization. Using ellipses as primitives, which are convex, recalls shape decomposition methods [9, 13] . It will be seen that the ellipses extracted with this method, will in general correspond to a perceptual decomposition of the object into its convex subparts.

2 Projective Invariance of Contact Point Ellipses

2.1 Cross-ratio of Points and Lines on a Conic

To discuss the invariance properties of ellipses with various contact points we start with the cross-ratio, the fundamental invariant for points and lines in the plane. The cross-ratio can be expressed as a ratio involving determinants. Given three column vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ in $\mathbb{R}^3$ we will use the bracket notation for the determinant of the 3×3 matrix formed by these vectors:

$$[\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3] \;=\; \det(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\ . \tag{1}$$

For 5 points in the plane, with homogeneous coordinates $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_e$, the cross-ratio is defined as:

$$\frac{[\mathbf{x}_a\ \mathbf{x}_b\ \mathbf{x}_e]\,[\mathbf{x}_c\ \mathbf{x}_d\ \mathbf{x}_e]}{[\mathbf{x}_a\ \mathbf{x}_c\ \mathbf{x}_e]\,[\mathbf{x}_b\ \mathbf{x}_d\ \mathbf{x}_e]} \;=\; \sigma\ . \tag{2}$$

The invariance of the cross-ratio over projective transformations $\mathbf{x}_i' = T\,\mathbf{x}_i$, where T is a nonsingular 3×3 matrix, follows easily from the rule

$$[\mathbf{x}_i'\ \mathbf{x}_j'\ \mathbf{x}_k'] \;=\; [T\mathbf{x}_i\ T\mathbf{x}_j\ T\mathbf{x}_k] \;=\; [T]\,[\mathbf{x}_i\ \mathbf{x}_j\ \mathbf{x}_k]\ . \tag{3}$$

Inserting this expression on the left-hand side in (2) we see that all determinants $[T]$ will cancel. If we consider the fifth point as a variable $\mathbf{x}$ and denote the cross-ratio $\sigma = \lambda_1/\lambda_2$ we have

$$\lambda_1[\mathbf{x}_a\ \mathbf{x}_c\ \mathbf{x}]\,[\mathbf{x}_b\ \mathbf{x}_d\ \mathbf{x}] \;-\; \lambda_2[\mathbf{x}_a\ \mathbf{x}_b\ \mathbf{x}]\,[\mathbf{x}_c\ \mathbf{x}_d\ \mathbf{x}] \;=\; 0\ . \tag{4}$$

This is a second-order polynomial in $\mathbf{x}$ representing a conic through the points $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d$. The cross-ratio $\sigma = \lambda_1/\lambda_2$ is then the cross-ratio of the four points on the conic. Varying the cross-ratio we get a pencil of conics through the four points. This pencil of conics can be expressed using the homogeneous symmetric matrices $P(\sigma), P_1, P_2$ as:

$$\mathbf{x}^T P(\sigma)\mathbf{x} \;=\; \mathbf{x}^T(\lambda_1 P_1(\mathbf{x}_a \ldots \mathbf{x}_d) \;+\; \lambda_2 P_2(\mathbf{x}_a \ldots \mathbf{x}_d))\mathbf{x} \;=\; 0\ . \tag{5}$$

Due to the duality between points and lines, a conic can be expressed as a quadratic form in line coordinates $\mathbf{u}$. The pencil of conics tangential to four lines $\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d$ expressed in line coordinates can be written:

$$\lambda_1[\mathbf{u}_a\ \mathbf{u}_c\ \mathbf{u}]\,[\mathbf{u}_b\ \mathbf{u}_d\ \mathbf{u}] \;-\; \lambda_2[\mathbf{u}_a\ \mathbf{u}_b\ \mathbf{u}]\,[\mathbf{u}_c\ \mathbf{u}_d\ \mathbf{u}] \;=\; 0\ . \tag{6}$$

Just as in the point case this can be expressed using homogeneous symmetric matrices $Q(\sigma), Q_1, Q_2$ as:

$$\mathbf{u}^T Q(\sigma)\mathbf{u} \;=\; \mathbf{u}^T(\lambda_1 Q_1(\mathbf{u}_a \ldots \mathbf{u}_d) \;+\; \lambda_2 Q_2(\mathbf{u}_a \ldots \mathbf{u}_d))\mathbf{u} \;=\; 0\ . \tag{7}$$

A projective transformation T that maps points $\mathbf{x}_i$ into $\mathbf{x}_i' = T\,\mathbf{x}_i$ will map the matrices $P(\sigma)$ and $Q(\sigma)$ to matrices $P'(\sigma)$ and $Q'(\sigma)$. If we apply this

transformation to the coordinates in the pencils (4) and (6) the conic matrices can be shown to be related as:

$$P(\sigma) = T^T P'(\sigma) T \qquad Q'(\sigma) = T Q(\sigma) T^T . \tag{8}$$

That is, the conic with cross-ratio $\sigma = \lambda_1/\lambda_2$ maps into a conic with the same cross-ratio.

For a given shape we will consider the class of ellipses having four contact points with the contour making up the shape outline. A contact point is defined as a point of second-order contact of the ellipse and the contour; that is, at that point both point and line coordinates of the contour and ellipse will coincide. Point and line coordinates of points on a conic are related as:

$$\mathbf{u}_a = P\mathbf{x}_a \quad \mathbf{u}_b = P\mathbf{x}_b \quad \mathbf{u}_c = P\mathbf{x}_c \quad \mathbf{u}_d = P\mathbf{x}_d . \tag{9}$$

For four points on a conic the cross-ratio will equal that computed from the tangents of the points. This follows easily from the fact that

$$[\mathbf{u}_a\ \mathbf{u}_b\ \mathbf{u}] = [P\mathbf{x}_a\ P\mathbf{x}_b\ P\mathbf{x}] = [P]\,[\mathbf{x}_a\ \mathbf{x}_b\ \mathbf{x}] . \tag{10}$$

which is applied to all the brackets in (6). An ellipse with four contact points can therefore be expressed in point and line coordinates as:

$$\mathbf{x}^T P(\sigma)\mathbf{x} = 0 , \qquad \mathbf{u}^T Q(\sigma)\mathbf{u} = 0 , \tag{11}$$

where P and Q are related to the point and line coordinates of the contact points according to (5) and (7).

Since line and point coordinates of a conic are related by $\mathbf{u} = P\mathbf{x}$ the relation $\mathbf{u}^T Q\mathbf{u} = 0$ can be written as $\mathbf{x}^T P^T Q P\mathbf{x} = 0$ from which we see that $P = P^T Q P$; that is, we have the important relation:

$$P(\sigma) = Q^{-1}(\sigma) . \tag{12}$$

This equation relates point and line coordinates of contact points and will play an important part in the design of an algorithm for actually locating contact points.

2.2 Four and Five Contact Point Ellipses

The cross-ratio computed from the contact points is invariant and can be used to identify corresponding ellipses in projective transform pairs. In the general case, for a given shape the class of ellipses with four contact points will be infinite. It is, however, a one-parameter infinite family as can be seen from the following crude equation-counting argument. Suppose that the curve can be parameterized with a parameter s. The contact points are then represented by parameters s_a, s_b, s_c, and s_d. Together with the five parameters for representing the ellipse, this gives us a total of nine parameters to be determined. These parameters are subject to the constraint that the ellipse should be in contact with the shape at the four points. This gives four constraints for coincidence of points and four

constraints for coincidence of tangents, making up a total of eight constraints. Nine parameters and eight constraints implies that the class of four contact point ellipses will be a one-parameter family. The important consequence of this is that in the generic case there will only be a finite number of ellipses with a *specific* cross-ratio; that is, we will have at most a finite ambiguity when identifying ellipses in projective transform pairs.

An analogous argument can be applied to the case of five contact points. In this case we will have ten unknown parameters and ten constraints, which in general will give a finite number of ellipses.

In general there will be a multiplicity of cross-ratios for four points or lines depending on the order in which the points are chosen. Permutations of the order of the points will give new values of the cross-ratio. However, if we adopt the convention of choosing the points on the ellipse in a certain order, e.g. clockwise, and restrict the projective transformations to those leaving the order of the points invariant, it can be shown that only two different cross-ratios can be computed for a set of four points on an ellipse. These two values of the cross-ratio will be each other's inverses. By choosing ellipses with cross-ratio = 1, this multiplicity is also taken care of. It can be shown that the projective transformations leaving the order of the points invariant are those corresponding to the case of relative camera motions with the object in front of the camera, which is a most natural constraint in real imaging situations.

The parameters and cross-ratio of the ellipses with four contact points to a smooth curve will in general vary smoothly with the contact points. When such an ellipse makes a fifth contact with the curve, five different cross-ratios can be computed, each corresponding to a selection of four contact points. A five contact point ellipse is therefore a common member of five different families of four contact point ellipses with smoothly varying parameters and cross-ratios. Five contact point ellipses can be seen as the analogues of the circles of the branching points in the skeleton of the medial axis transform.

Figure 1 shows five families of four contact point ellipses with varying cross-ratio. The shape is a polygon and these ellipses were computed by an exhaustive search of all combinations of four lines in the polygon. Figure 2 shows the five contact point ellipse common to these families and ellipses with four contact points and cross-ratio equal to one. As can be seen they correspond quite well to the intuitive decomposition of the shape in convex subparts.

3 Iterative Algorithm for Fitting Four or Five Contact Point Ellipses

3.1 Structure of the Algorithm

The ellipses of the polygon in Figs. 1 and 2 were computed by an exhaustive search among all combinations of lines in the polygon. It is however desirable to work with more general shapes extracted from grey-level images. The initial step of shape extraction from grey-level images is edge detection. The contours are

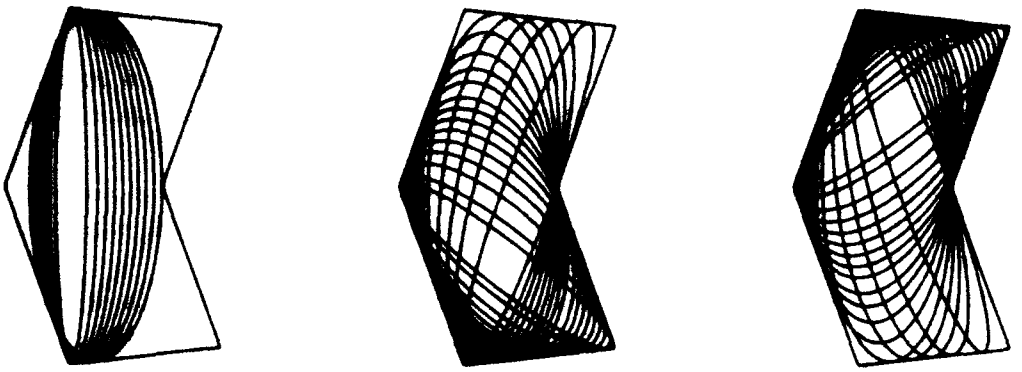


Fig. 1. Five families of four contact point ellipses. Left: one family. Center: two families. Right: two families. The families in the center and right figure are symmetrically related by a reflexion in the horizontal symmetry axis of the shape. All these families have the five contact point of Fig. 2 in common.

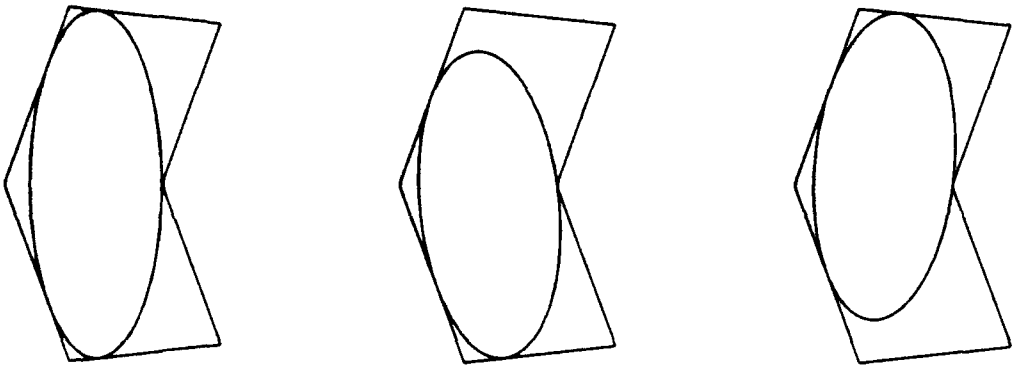


Fig. 2. Left: Five contact point ellipse. Center and right: four contact point ellipses with cross-ratio equal to one.

then represented by the sampled image coordinates (x_i, y_i) of the edge points. These coordinate lists are of finite precision, determined by the sampling grid, and long contours are often fragmented into smaller parts. An algorithm for computing contact point ellipses has to take all these facts into account.

The algorithm works iteratively, starting with an initial ellipse and updating it in a manner similar to that of active contour finding [10]. It is designed to have a fixed point for ellipses with four contact points and unit cross-ratio, and to break whenever five contact points are encountered.

3.2 Main Iteration Loop

The relation between the point and line coordinate matrices of an ellipse with four contact points was derived in the previous section (12). This relation is the

basis for deciding whether four points on a curve can be the contact points of an ellipse.

Given four points on a curve $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d$ with tangents $\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d$, the quantity:

$$\min_{\sigma} ||P(\sigma)\, Q(\sigma) - I||^2 \ . \tag{13}$$

will be zero iff the four points are contact points of the ellipse $P(\hat{\sigma})$ where $\hat{\sigma}$ is the value of the cross-ratio that minimizes the quantity. P and Q here are properly normalized. For arbitrarily chosen points the quantity will not be zero, but the ellipses $P(\hat{\sigma})$ and $Q(\hat{\sigma})$ will represent a certain best compromise.

Minimization of the criterion in (13) is one step in the iterative algorithm, mapping points on the curve to an updated ellipse. We actually have two choices for the updated ellipse, $P(\hat{\sigma})$ and $Q^{-1}(\hat{\sigma})$. By choosing $P_{update} = Q^{-1}(\hat{\sigma})$ we will get a four contact point ellipse in one iteration in the case the shape is a four sided polygon, since all ellipses $Q^{-1}(\sigma)$ are four contact point ellipses in this case.

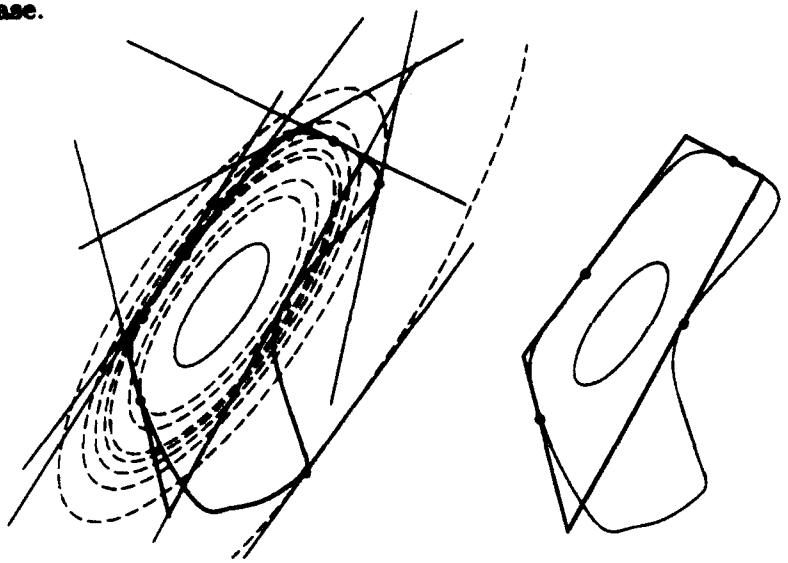


Fig. 3. Left: Ellipse in iterations, level curves of algebraic distance (broken) tangent lines of points of local extrema of algebraic distance (thin) and minimal polygon (wide). Right: four-line polygon and contact points for updating of ellipse.

In order to have a complete iteration a method of updating the four points on the curve is required. For a given ellipse P we would like to find the four new candidates for contact points. Referring to Fig. 3 start with computing the algebraic distance δ to each point x_i, y_i on the curve. Using a normalized matrix P and homogeneous coordinates $\mathbf{x} = (x_i, y_i, 1)^T$ we have

$$\delta(x_i, y_i) \ = \ \mathbf{x}^T P \mathbf{x} \ . \tag{14}$$

A contact point on the curve will have its tangent coinciding with the tangent of the ellipse $P\mathbf{x}$. It will also be a local extremum of the algebraic distance

$\delta(x_i, y_i)$ along the curve. Since the level curves of constant algebraic distance corresponds to scaled versions of the ellipse, any local extremum of algebraic distance along the curve will correspond to a contact point of a scaled version of the ellipse, as shown in Fig. 3. The points of local extrema of algebraic distance therefore form good candidates to be used as tentative contact points in the iterative algorithm. There are however in general more than four local extrema of algebraic distance along the curve. In order to select four specific points proceed as follows.

1. For each local extremum $\mathbf{x}_i$, the tangent $\mathbf{u}_i$ is computed using the matrix P: $\mathbf{u}_i = P\mathbf{x}_i$.
2. From all the tangents $\mathbf{u}_i$, the polygon enclosing the center point of the ellipse and not containing any other tangents is computed. This is referred as the minimal polygon.
3. From all the tangents in the minimal polygon select the four that enclose the center point of the ellipse and have minimal area.

This procedure gives four tangent lines to the curve $(\mathbf{u}_a \ldots \mathbf{u}_d)$ and their corresponding tangent points $(\mathbf{x}_a \ldots \mathbf{x}_d)$, for a given ellipse P. This closes the iteration loop. Note that four contact points will be fixed points of the algorithm.

The iterative algorithm is started by selecting an initial ellipse. This initial selection is a critical step for the convergence properties of the algorithm. This problem is treated in the experimental section

The iterative algorithm for finding four contact point ellipses to a curve represented by its edge coordinates can be summarized in the following steps:

1. Select an initial ellipse.
2. Compute points on the curve that are local extrema of the algebraic distance function generated by the ellipse. Compute tangents at these points.
3. Compute the minimal polygon from tangents enclosing the center of the ellipse.
4. Select the four lines in the minimal polygon that encloses the minimum area. Together with step 2 this gives four points and four lines.
5. Use these four points and lines to compute matrices $P(\sigma)$ and $Q(\sigma)$ parameterized by the cross-ratio σ.
6. Find cross-ratio $\hat{\sigma}$ that minimizes $||P(\sigma)\, Q(\sigma) - I||^2$.
7. Update ellipse $P_{update} = Q^{-1}(\hat{\sigma})$.
8. Check for convergence to four or five contact point ellipse.
9. If no convergence, go to step 2.

Convergence is in two steps: (1) Convergence to an ellipse with four contact points but arbitrary cross-ratio. (2) Convergence to an ellipse with four contact points and unit cross-ratio or an ellipse with five contact points.

The criterion for judging whether a point is a contact point is simply its distance from the ellipse. If this is below a threshold the point is declared to be a contact point. Since all points have the property of being local extrema of algebraic distance along the curve, this will automatically mean that a point

coinciding with the ellipse will have its tangent parallel to that of the ellipse that is, it will be a contact point.

When convergence to four contact points has been declared, the cross-ratio $\hat{\sigma}$ computed by the minimization in (13) is checked to lie within a threshold of unity. If not, points of local extrema of algebraic distance along the curve whose tangents are in the minimal polygon, are checked to lie within threshold distance of the ellipse. If such a point is found, convergence to five contact points is declared.

If no convergence is declared, the ellipse is updated in order to find unit cross-ratio ellipses. This updating is simply

$$P_{update} = Q^{-1}(1) \ . \tag{15}$$

That is, P is updated as the ellipse with unit cross-ratio in the pencil of ellipses inscribed in the four-sided polygon made up of the tangents of the four points. This can be seen as a linear extrapolation since if the curve to which we try to fit ellipses is actually a four-sided polygon, we will get convergence in one step.

The iterations are stopped if no convergence is found within 40 iterations. Iterations are also stopped if less than four lines are found in the minimal polygon.

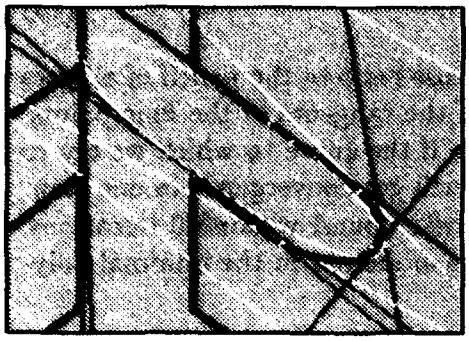
3.3 Experimental Results on Synthetic Shapes

Figure 4 shows an example of the evolution of the initial steps of the iterative algorithm on a synthetic shape. The tangent lines corresponding to local extrema of the algebraic distance of the ellipse are shown in white and the resulting minimal polygon in black on the left side of the figure. The right side shows the polygon of minimal area selected from the minimal polygon and the updated ellipse. In this case convergence to a four contact point ellipse is in 3 iterations. Typically the convergence to a four contact point ellipse is in 2 to 5 steps, while the convergence to a unit cross-ratio ellipse is 5 - 15 iterations.

A critical step in the algorithm is of course the selection of the initial ellipse. The algorithm that has been described will converge to a single four or five contact point ellipse. For a given shape there will in general be several four contact point ellipses with unit cross-ratio and five contact point ellipses. Selection of the initial ellipse will of course determine to which of these ellipses the algorithm will converge. This means that the parameter space of ellipses can be partitioned into regions, with each region corresponding to the convergence to a certain unit cross-ratio four contact point ellipse or a five contact point ellipse. At least one initial ellipse must be selected from every such region in order that all four contact point ellipses are found. Since four contact point ellipses with unit cross-ratio seem to correspond rather well with a decomposition of the object into convex subparts, a natural way to find initial ellipses is to use a very simple initial decomposition into convex subparts, and use a simple ellipse approximation of these subparts. This initial decomposition need not be perfect in any way. The important thing is that it should yield an ellipse within


Tangents: iteration 1


Min. area 4 side polygon: iteration 1


Tangents: iteration 2

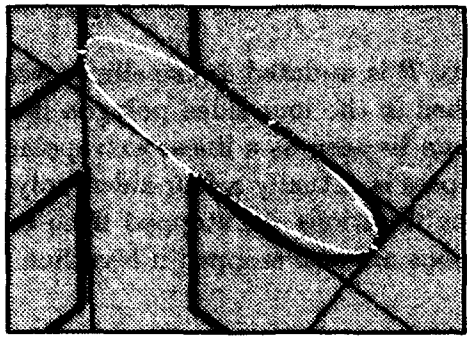

Min. area 4 side polygon: iteration 2

Fig. 4. Steps in the iterative ellipse fitting algorithm. The shape contour is in full black. Left: tangents of points of local extrema of algebraic distance (white,) lines of minimal polygon (black). Right: updated ellipse and minimum area four-side polygon.

the convergence region for the four contact point ellipse corresponding to that subpart.

In the case of synthetic images the algorithm of [9], where convex subparts are found by computing points of local minima of negative curvature along the shape contour, will be used. For the contour points $p_n = (x_n, y_n)^T$ between two such local minima we find the center of gravity p_{cg} and positive definite 2×2 scatter matrix S:

$$p_{cg} = \sum_{n=1}^{N} p_n/N \ , \qquad S = \sum_{n=1}^{N} (p_n - p_{cg})(p_n - p_{cg})^T/N \ .$$

With $p = (x, y)^T$, the initial ellipse is chosen as

$$(p - p_{cg})^T S^{-1} (p - p_{cg}) = 1 \ .$$

This ellipse will give a reasonably good approximation to the shape of the points $p_1 \ldots p_N$. In this process the end points of the contour segment between the local curvature minima are given extra weight, which will bias the initial ellipse towards the junction of the subpart with the main shape.

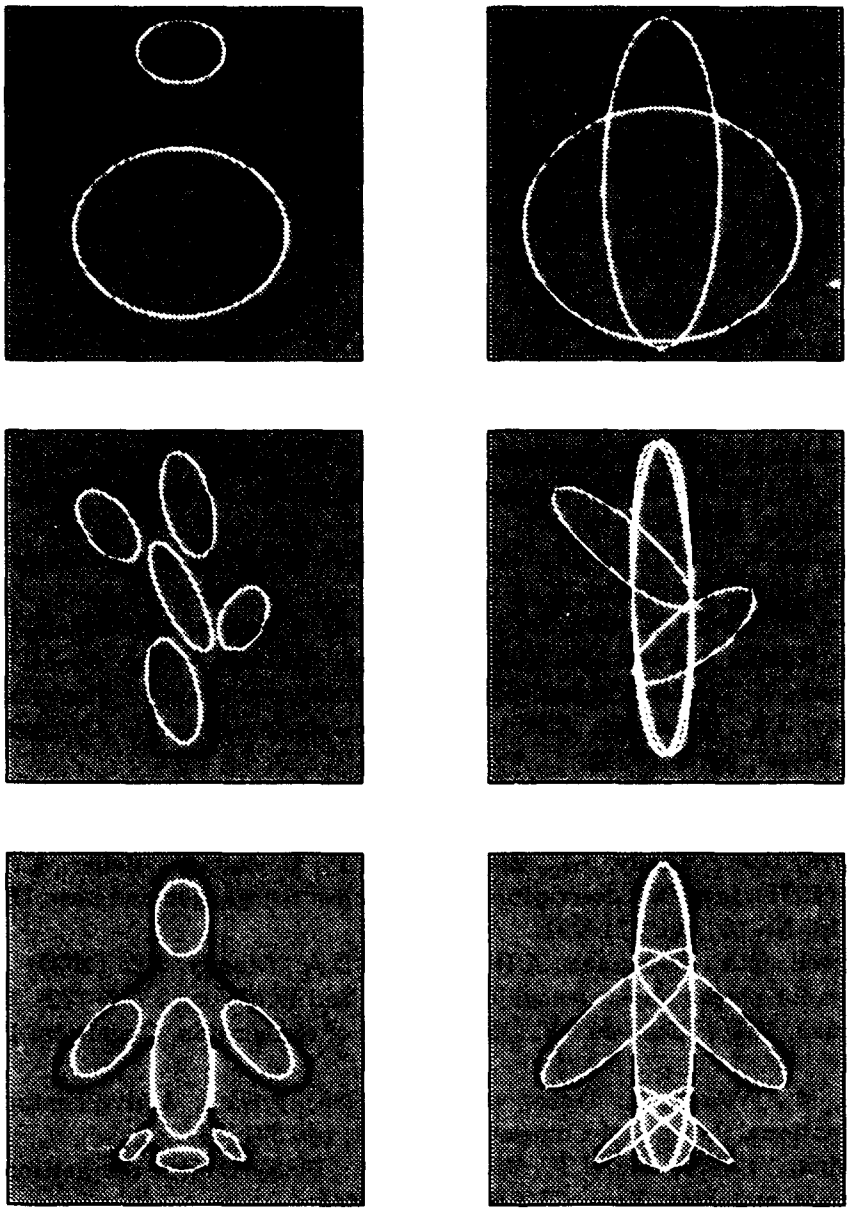


Fig. 5. Initial ellipses from subpart approximation using negative curvature maxima (left) and resulting ellipses with unit cross-ratio after applying the iterative algorithm (right)

The results of choosing initial ellipses in this way are shown in Fig. 5 on the left side. The right side shows the resulting unit cross-ratio four contact point ellipses and the five contact point ellipses. The result confirms the initial assumption that unit cross-ratio four contact point ellipses correspond to the decomposition into convex subparts. Several alternative initializations with randomly selected starting ellipses gave the same results and no alternative four contact point unit cross-ratio ellipses.
Convergence in those cases needed more iterations however.

4 Summary

An algorithm has been presented for computing ellipses from edge coordinates of an object. For planar objects these ellipses will be related to the shape in a projectively invariant way and can therefore be used for viewpoint independent recognition. For synthetic shapes with perfect edge data the algorithm will converge very fast if the initial ellipses are chosen as approximations to the convex subparts of the shape.

References

1. Barrett, E.B., Payton, P.M., Haag, N.N., Brill, M.H. (1991). General methods for determining projective invariants in imagery. CVGIP-IU 53, pp. 46–65.
2. Blum, H. (1973). Biological shape and visual science, Journal of Theoretical Biology, 38, pp. 205–287.
3. Bookstein, F. (1979). Fitting Conic Sections to Scattered Data, Computer Graphics and Image Processing 8, pp. 56–71.
4. Brady, M. (1983). Criteria for representation of shapes. In: Rosenfeld, A., Beck (eds), Human and Machine Vision, Erlbaum, Hillsdale NJ, pp. 39–84.
5. Carlsson, S. (1992). Projectively invariant decomposition of planar shapes. In: Mundy, J.L., Zisserman, A.P. (eds), Geometric Invariance in Computer Vision, MIT-Press , pp. 267–273.
6. Forsyth, D.A., Mundy, J.L., Zisserman, A.P., Brown, C. (1990). Invariance-a new framework for vision, Proc. of 3rd ICCV, pp. 598–605.
7. Forsyth, D.A., Mundy, J.L., Zisserman, A.P., Coelho, C., Heller, A., Rothwell, C.A. (1991). Invariant descriptors for 3-D object recognition and pose, IEEE PAMI Vol. 13, No 10 , pp. 971–991.
8. Rothwell, C.A., Zisserman, A.P., Forsyth, D.A., Mundy, J.L. (1992). Canonical frames for planar object recognition, Proc. 2nd ECCV, pp. 757–772.
9. Hoffman, D.D., Richards, W. (1984). Parts of recognition, Cognition, 18, pp. 65–96.
10. Kass, M., Witkin, A., Terzopoulos, D. (1988). Snakes: Active contour models, International Journal of Computer Vision 1, pp. 321–331.
11. Lamdan, Y., Schwartz, J.T., Wolfson, H.J. (1988). Object recognition by affine invariant matching, Proc. CVPR, pp. 335–344.
12. Lee, D.T. (1982). Medial axis transform of a planar shape, IEEE Trans. on Pattern Analysis and Machine Intelligence, 4, pp. 363–369.
13. Richards, W., Hoffman, D.D. (1985). Codon constraints on closed 2D shapes, CVGIP 31, pp. 265–281.
14. Van Gool, L.J., Moons, T., Pauwels, E.J., Oosterlinck, A. (1992). Semi-differential invariants. In: Mundy, Zisserman (eds), Geometric Invariance in Computer Vision, MIT-Press, pp. 157–192.
15. Weiss, I. (1988). Projective invariants of shapes, Proc. CVPR, pp. 291–297.

Blind Approximation of Planar Convex Shapes

Michael Lindenbaum and Alfred M. Bruckstein

Computer Science Department, Technion - Israel Institute of Technology, Haifa 32000, Israel

Abstract. This paper considers the process of learning the shape of an unknown convex planar object through an adaptive process of simple measurements called line probings, which reveal tangent lines to the object. A systematic probing strategy is suggested and an upper bound on the number of probings it requires to yield an approximation of the unknown object with a pre-specified precision is derived. A lower bound on the number of probings required by any strategy for such an approximation is also derived, showing that the gap between the number of probings required by the suggested strategy and the number of probings required by the optimal strategy is a logarithmic factor in the worst case. The proposed approach overcomes a basic deficiency of the classic geometric probing approach, which is based on the assumption that objects are polygonal and therefore is not applicable for a variety of real robotic tasks.

Keywords: shape, convex shape approximation, geometric probing.

1 Introduction

In this paper the process of learning the shape of an unknown convex planar object through an adaptive process of probing is considered. A probing is done by choosing a direction on the plane, and moving a line perpendicular to this direction, from infinity until it touches the object. Each such line probing reveals a tangent line (see Fig. 1). It is necessary to find a systematic procedure which guarantees that, after a given number of probings, the best possible approximation to the unknown object may be generated.

The problem, suggested as a simplified theoretic model for the robotic task of learning about an object from tactile sensors, is related to the class of geometric probing problems, but is very different from them. The geometric probing paradigm formulates the shape learning process in an algorithmic setting and usually addresses the following type of problem, initiated by Cole and Yap [3]: Determine the shape of an object from an adaptive sequence of simple measurements, called probings. The unknown object is known to belong to a restricted class, such as convex polygons or polyhedra, a certain type of geometric probe

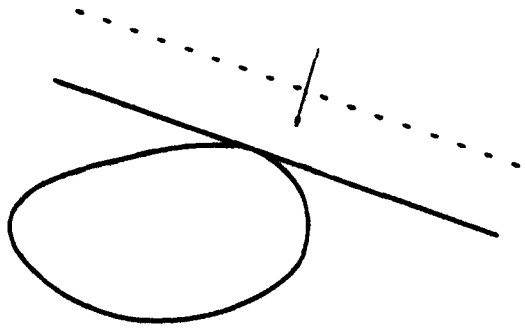


Fig. 1. Line probing

is defined, and a reconstruction strategy is suggested and rigorously analyzed. The performance of the probing strategy is measured by the number of probings which guarantee exact reconstruction. This number is a function of the object complexity (V), and is usually compared with a proven lower bound on the number of probings required by any strategy.

One of the many examples of geometric probing [1, 4, 5, 9, 10, 11, 14, 17, 18], which is related to the problem we consider here, is the work of Li [8], who considered the reconstruction of a convex polygon from line probings identical to the ones we investigate here. Li improved earlier results [4, 6] and suggested a probing strategy which guarantees complete reconstruction after no more than $3V + 1$ probings. He also derived a lower bound of $3V + 1$ on the number of measurements required by any strategy for a guaranteed reconstruction, thereby proving the optimality of his strategy.

The polygonal assumption, together with the desire for perfect reconstruction, implies that geometric probing results are not applicable for real robotic tasks. Here these deficiencies are overcome by not restricting the objects considered to be polygonal and by modifying the exact reconstruction problem into an approximation problem while maintaining the classical structure of geometric probing. An adaptive strategy for approximate reconstruction is sought, and its figure of merit is defined as the number of probings it requires, in the worst case, in order to achieve a certain certified approximate reconstruction. An upper bound on this figure of merit is proved in a rigorous way, and compared to a lower bound on the number of probings required by any strategy for achieving an approximation with the same precision.

The specific goal considered is to obtain enough information to approximate the unknown object. We intend to specify the shape of another object, referred to as a *certified approximation*, whose Hausdorff distance from the unknown probed object does not exceed a certain prespecified value ϵ. This probing procedure is referred to as *blind approximation* since the approximated object is not known fully and the approximation is inferred and evaluated without seeing (knowing) it.

2 Approximating Convex Shapes with Polygons

First, the problem of approximating convex shapes with polygons under the Hausdorff metric is considered. The Hausdorff distance between two planar sets S and P, is

$$\delta^H(P,S) = \max\{\sup_{x\in P}\inf_{y\in S}\|x-y\|, \sup_{y\in S}\inf_{x\in P}\|x-y\|\} \ , \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm in R_2.

The number of vertices in a polygon providing an ϵ-approximation of convex planar sets, has been investigated in many variations (see, e.g. [7, 13]; more references are cited in [10]). In a less quoted result [2], Bolour and Cover provide the following simple constructive method for building an approximating external polygon to a convex set S. Choose the first vertex v_0 as any point, external to S, whose distance from its boundary is ϵ. Through this point, pass a tangent to S and extend it to the point v_1 whose distance from the boundary of S is also ϵ. Let v_1 be the successive vertex. Repeat this process, each step obtaining a new vertex v_i and a new side $v_{i-1}v_i$, as long as the curve $v_0, \cdots, v_i$ does not cut itself. To complete the approximating polygon, connect the last vertex v_m and v_0 by a line segment and let it be the last side (see Fig. 2). It is clear that the polygon $v_0, \cdots, v_m$ is external and convex, and that its Hausdorff distance from S is exactly ϵ. Bolour and Cover [2] proved that every convex set S with perimeter L can be approximated by an external convex polygon, built by the method described above, whose number of vertices, n is bounded by

$$n \le \pi\left(\frac{L}{2\pi\epsilon}\right)^{1/2} + 1 \ . \tag{2}$$

(See [2] or [10] for a proof.) Asymptotically, this result coincides with the result derived by Popov, (see, e.g., Gruber [7]), and by McClure and Vitale [13].

3 The Task

Consider the following discrete measurement process applied to an unknown convex object S. Before each measurement, a direction is arbitrarily specified and a line probe perpendicular to this direction is moved from infinity until it touches the object, then a half-plane known to include the object S is revealed (see Fig. 1).

After some measurements, inference about the object shape may be made. The polygon R_j, defined as the intersection between all half-planes revealed by the probings, must include the object and must have a point of the object on each of its sides. The intention is to use the data gathered by the measurements to specify a set S' whose Hausdorff distance from the unknown set is guaranteed to be ϵ or less. Such a set S' is called a *certified approximation* and is denoted CA_ϵ. If the certified approximation CA_ϵ is a polygon, then it is denoted CAP_ϵ.

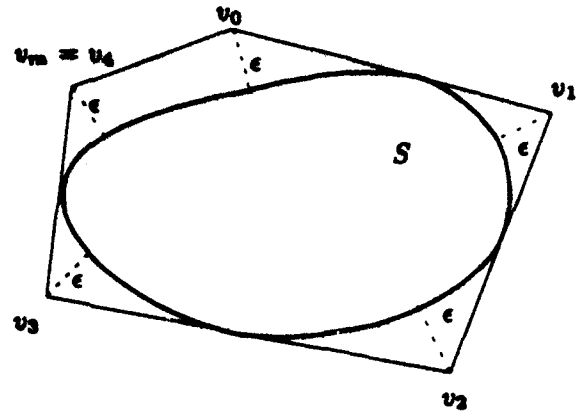


Fig. 2. Building external polygon to S, in distance ϵ

Define a probing strategy to be a rule for choosing the probing direction, which may depend adaptively on all previous probing results. A strategy is considered to be a better one if it requires, in the worst case, a smaller number of probings to achieve a certified approximation with a certain precision.

4 A Condition for Specifying a Certified Approximation

First, a necessary and sufficient condition on the probing results, for specifying a certified approximation, is derived. Let the height of the vertex v_i in a polygon P be the distance between v_i and the line $v_{i-1}v_{i+1}$. The height of the polygon is h if $h_i \leq h$ for all i. Define a central polygon P^c of a polygon P as the (convex) polygon whose vertices $v_1^c, \cdots, v_m^c$ are the centres of the sides of P. See Fig. 3, in which a polygon P and its central polygon P^c are described. With this notation, the following result can be proved.

Theorem 1. *Let R_j be a polygon created as the result of the probing process after j probings. Then R_j being of height 2ϵ is a necessary and sufficient condition for inferring a certified approximation CA_ϵ to the unknown object S. If the condition holds, then the Hausdorff distance of R_j^c, the central polygon of R_j, from the unknown object S does not exceed ϵ, i.e. R_j^c is a CAP_ϵ.*

The above theorem (proved in [10]) provides a stopping rule, and a direct method for finding the required certified approximation from the given probing results. One can look now for a strategy that achieves this approximation using a small number of probings.

5 A Lower Bound

Before specifying a strategy the performance achievable from any strategy is investigated by deriving a lower bound on the number of probings required to

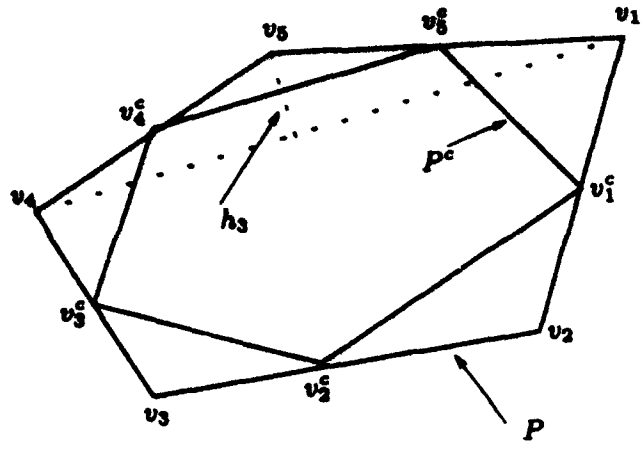


Fig. 3. A convex polygon P and its central polygon P^c

achieve a certified approximation. The bound is derived by considering an approximation problem. A convex set S is given and a polygon P with the following properties is sought. The polygon P should include S with every side being tangent to the boundary of S, its height (as defined before) must not exceed the number 2ϵ, and its number of vertices should be minimal. Such a polygon, from which the certified approximation to S (P^c) may be clearly inferred, could be thought of as a result of an optimal probing strategy which gets correct information about the shape of S and needs only to produce a certified approximation. The number of sides of this polygon is a lower bound on the number of probings required to find a certified approximation using the best possible strategy, as the additional information about S cannot degrade the performance of the probing strategy.

Since an optimal strategy with respect to the worst case is sought, then, in order to show the limitations of any strategy, it suffices to treat a single case. Thus, the lower bound LB on the number of sides of P is developed for S being a circle with radius R, and this bound turns out to be

$$m \geq 2\left(\frac{L}{2\pi\epsilon}\right)^{1/2} = LB \ . \tag{3}$$

(See [10] for a proof.)

6 The Strategy

The suggested strategy is very simple and is based on the following principles: The jth probing is associated with a certain vertex of R_{j-1}, which is the first to meet it. The probing directions are chosen from fixed sets, such that in the kth stage the probing directions are taken only from the set

$S_k = \left\{ i \cdot \frac{2\pi}{2^k} \mid i = 1, 2, 3, \cdots, 2^k \right\}$, which includes an exponentially increasing number of directions. The probing, however, is done only if the height of the associated vertex exceeds 2ϵ. This probing deletes the corresponding vertex of R_j and creates two adjacent vertices in place of it.

First, all the directions are sampled uniformly and coarsely, but after the height of some vertices becomes smaller than the threshold 2ϵ, no further probing is done on them and all probing effort is concentrated in the places where the uncertainty is still higher than allowed. The strategy depends on the parameter 2ϵ and the probing process terminates only after the height of the polygon R_j is 2ϵ or less. This hierarchical and adaptive nature allows the strategy to use fewer probings than a strategy based on a uniform sampling method.

The strategy proposed is similar, in principle, to the hierarchical representations of images [16]. The region quad-tree, for example, is based on examining a square part of the image, and if it is not uniform it is recursively split to four equal square cells. This process yields a representation of the image made of square uniform cells whose sizes are adapted to the local uniformity of the image. In the strategy proposed here, the amount of uncertainty is examined for each of the angular intervals whose end points are directions of probings, and if the uncertainty is above a certain level, another probing is done, the interval is halved, and the uncertainty is recursively examined for each of its halves.

7 An Upper Bound on the Number of Probings Done Until Reconstruction

To develop an upper bound on the number of probings done until reconstruction, an indirect approach is used. Consider the circumscribing imaginary polygon $P(S)$ built around S using the method described in [2]. The result of probing S with (exact) line probes is tangent lines which may be considered also as the results of probing the imaginary polygon $P(S)$ with a special probe that is allowed to penetrate into the polygon up to a maximal depth ϵ (see Fig. 4). Such

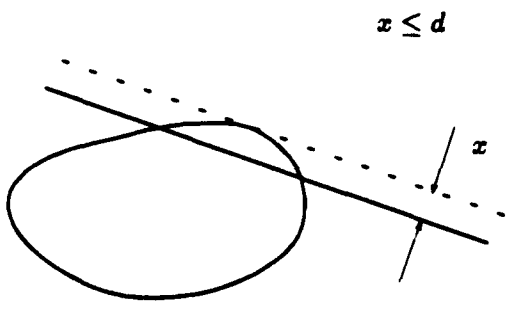


Fig. 4. Probing with positive error

probes are denoted *positive error line probe*, and considered *consistent* if there is a convex object that is tangent to all of them [14]. Then, the following claim is made. (See [10] for the proof.)

Lemma 2. *Let P be a convex polygon with V vertices and perimeter L. Assume $V \leq \frac{8L}{e\epsilon}$ where e is the natural logarithm base. Then, probing according to the proposed strategy with consistent positive error line probes associated with error ϵ guarantees that after no more than $j = V \log_2\left(\frac{16L}{\epsilon V}\right)$ probings, the height of R_j^d is 2ϵ or less.*

Recall now that $P(S)$ does not have more than $V \leq \pi\left(\frac{L}{2\pi\epsilon}\right)^{1/2} + 1$ vertices. Inserting this value for the number of vertices V of the imaginary polygon now yields the following main result. (See [10] for the proof.)

Theorem 3. *Let S be a general convex object with perimeter L. Then probing according to the proposed probing strategy (with parameter 2ϵ) guarantees that the height of R_j cannot exceed 2ϵ and that R_j^c is a certified approximation to S after no more than $j = UB$ probings, where asymptotically $UB \approx \sqrt{\frac{\pi}{8}}\sqrt{\frac{L}{\epsilon}}\log_2\left(\frac{512}{\pi}\frac{L}{\epsilon}\right)$.*

8 Discussion

This paper shows how geometric probing can be made into a useful technique for shape estimation from a sparse set of partial measurements. The demand for exact reconstruction is replaced by an easier and practical demand of finding a certified approximation from an adaptive sequence of line probings. A lower bound LB on the number of probings required to achieve an approximation with a certain precision, is derived. A probing strategy, relying on the basic notion of starting by uniform probing and focusing adaptively on higher uncertainty directions, is proposed. The performance of the proposed strategy is investigated by deriving an upper bound UB on the number of probings it requires to guarantee an approximate reconstruction with a certain precision. Both bounds depend on the normalized required precision ϵ_n defined as $\epsilon_n = \frac{\epsilon}{L}$, where ϵ is the required precision of the approximation and L is the perimeter of the unknown object. It follows that the number of probings n required by the optimal strategy satisfies

$$\Omega\left(\sqrt{\frac{1}{\epsilon_n}}\right) = LB < n_{\text{optimal}} < UB = O\left(\sqrt{\frac{1}{\epsilon_n}}\log\frac{1}{\epsilon_n}\right) . \qquad (4)$$

Thus, the number of probing steps is different only by a logarithmic factor from the absolute minimal number of sides required to represent the object approximately when it is fully known [7, 13]. The task considered here is similar to the more general problem of approximating a function given only a discrete set of samples [15]. It could be formulated in this context as the probing results may be considered to be samples of the support function. However, the question discussed here is more complicated as the partial data are collected dynamically and actively, and the strategy for collecting the data is also be considered.

This new approach leads to many interesting open problems. The first obvious one is to close the gap between the lower bound and the proved performance of the proposed strategy. Extending the results to hyperplane probing and to higher dimensions remain open problems too. Extending the results to non-convex objects is impossible in the context of line probings but may be possible if other types of probes, for example, finger probes, are considered. Finding certified approximations using different metrics will also provide completely different problems. It is interesting to note, in this context, that if the difference of area is considered as the metric, and if a line probe which reveals the tangency point is used, then an optimal blind approximation method could be obtained [12].

References

1. Alevizos P.D. , Boissonnat J.D., Yvinec M. (1989). Probing non convex polygons, Proc. IEEE Int. Conf. on Robotics and Automation, pp. 202–207.
2. Bolour, A., Cover, T.M. (1972). On the number of convex sets on the unit square, Technical Report No. 2, Department of Statistics, Stanford University.
3. Cole, R., Yap, C.K. (1987). Shape from probing, J. of Algorithms 8, pp. 19–38.
4. Dobkin, D., Edelsbrunner, H., Yap, C.K. (1986). Probing convex polytopes, Proc. 8th ACM Symposium on Theory of Computing, pp. 424–432.
5. Edelsbrunner, H. Skiena, S.S.,(1988). Probing convex polygons with X-rays, SIAM J. on Computing 17, pp. 870–882.
6. Greschak, J.P. (1985). Reconstructing Convex Sets, Ph.d. Diss. Elec. Eng., M.I.T.
7. Gruber,P.M. (1983). Approximation of convex bodies, In: Gruber, P.M. Wills, J.M., Convexity and its Applications, Birkhauser.
8. Li, S.Y.R. (1988). Reconstruction of polygons from projections, Information Processing Letters 28, pp. 235–240.
9. Lindenbaum, M., Bruckstein, A. (1990). Reconstruction of convex polygon from binary perspective projections, Pattern Recognition 23, pp. 1243–1350.
10. Lindenbaum, M., Bruckstein, A. (1990). Blind approximation of planar convex sets, CIS Report No. 9008, Center for Intelligent Systems, Technion.
11. Lindenbaum, M., Bruckstein, A. (1992). Parallel strategies for geometric probings, J. of Algorithms 13, pp. 320–349.
12. Lew, J.S., Quarles, D.N. Jr. (1989). Optimal inscribed polygons in convex curves, Am. Math. Month., pp. 886–902.
13. McClure, D.E., Vitale, R.A. (1975). Polygonal approximation of plane convex bodies, J. of Mathematical Analysis and Applications 51, pp. 326–358.
14. Prince, J.L., Wilsky, A.S. (1990). Reconstructing a convex set from noisy support line measurements, IEEE Trans. Patt. Anal. Mach. Intell. PAMI-12, pp. 377–389.
15. Rivlin, T.J. (1969). An Introduction to the Approximation of Functions, Braisdel Publ.
16. Samet, H. (1989). The Design and Analysis of Spatial Data Structures, Addison-Wesley, Reading, MA.
17. Skiena, S.S. (1988). Geometric Probing, Doctoral Dissertation, Department of Computer Science, University of Illinois, Urbana, IL.
18. Skiena, S.S. (1989). Problems in geometric probing, Algorithmica, pp. 599–605.

Recognition of Affine Planar Curves Using Geometric Properties

Craig Gotsman[1] *and Michael Werman*[2]

[1] Department of Computer Science, Technion, Haifa 32000, Israel
Email: gotsman@cs.technion.ac.il.
[2] Department of Computer Science, The Hebrew University, Jerusalem 91904, Israel
Email: werman@cs.huji.ac.il.

Abstract. An algorithm for the recognition of a digital image of a planar curve which has undergone an affine transformation is presented. The algorithm is based on affine-invariant extremal geometric properties of curves, uses existing computational-geometric methods, and is relatively insensitive to noise. Its time complexity is linear in the number of image pixels on the curve.

Keywords: shape, affine matching.

1 Introduction

A common problem in computer vision is the recognition of objects in images. Usually, the objects are represented as three-dimensional models, which are matched to their two-dimensional projections in the image. The typical viewing transformation is a perspective projection, characterized by 8 parameters. Matching under this transformation has proved difficult, forcing various approximation schemes to be employed. A common approach is to assume that the objects are relatively flat and rigid, hence the relation between the model and its image is approximated well by a two-dimensional affine transformation (characterized by 6 parameters); see [16] on the use of affine approximation to the perspective viewing transformation. Furthermore, it is assumed that the object models are their contours, and that such contours are extracted from the image. The object recognition problem then reduces to the matching of a digitized planar curve to a model planar curve.

Matching under affine transformations and its special cases (congruence and similarity transformations) has been the focus of much attention. The pure computational - geometric problems deal with point sets, ranging from exact matching of equal size point sets [12, 20], through exact matching of unequal size point sets [13], to approximate matching [4, 3]. The more applied computer vision problems are concerned with contour matching, where the contours are groups of pixels extracted from images. This paper addresses these problems. The current approaches to contour matching in the image processing literature can be categorized into two major classes: "global" and "local" methods. For global methods,

the entire contours are required to perform the matching. These include algorithms based on global shape properties, such as perimeter, area, and moments [15, 12]. Any such property invariant under a viewing transformation could serve as a basis for a recognition algorithm. The main advantage of such methods is their relative insensitivity to noise. However, by definition, these methods cannot deal with cases where any part of the curve is occluded, a major limitation in practical implementations. The local methods, considered the more modern approach, are based either on the detection of special-feature points on the contour, such as breakpoints or inflection points [16], or on differential invariants of the curves [7, 8, 21]. Relying on tensor theory, Cyganski *et al.* [9] first proposed the use of affine invariant curvature functions for recognition purposes. In their original paper [9], mainly due to complex tensor derivations and global information based normalization, it was not clear whether their method could be used to solve the recognition problem under partial occlusion. However, Vaz and Cyganski [21] did derive a local curvature invariant and employ a Hough-type method to solve the affine transformation recognition problem. Independently, Bruckstein and Netravali [8] developed a similar differential invariant, and generalized it to deal with arbitrary projective distortions.

Our approach to solving the contour matching problem can be considered "semi-local". It is reminiscent of that of Hopcroft and Huttenlocher [13] to the solution of the equal size point set matching problem. Both rely on the mathematical fact that the areas of an arbitrary finite region in the plane before and after the affine transformation are related by a constant factor. Hopcroft *et al.* took advantage of this by observing that the *ratio* of areas of transformed triangles is invariant. Their method is applicable only to the exact matching of equal cardinality point sets. It breaks down completely when confronted with two noisy point sets of unequal cardinality. However, we use a different implication of the same mathematical fact — geometric shapes with *extremal areas* among a number of possible shapes are invariant under affine transformations. We identify such shapes and use existing computational-geometric algorithms for their computation. The algorithms are adapted to the contour matching problem without degrading their performance significantly. This yields an algorithm for affine matching, presented in Sect. 3, whose time complexity is linear in the number of pixels on the image contour. Section 4 shows that this algorithm is, in a sense, optimal in the presence of noise.

2 Problem Definition

In its pure form, the problem to be solved is the exact matching of two continuous planar curves. In real-world problems, however, the curves are quantized, so that only a limited number of sample points on each curve are available. Moreover, the sampled points are corrupted by sensor noise and quantization errors, making exact matching unrealistic, so a tolerance parameter ϵ has to be imposed on the solution. Our approach to the solution of the inexact problem is essentially the same as that of the pure continuous problem, adapted to point matching and

accommodating a tolerance parameter. The solution carries over if the samples of the curves are dense enough, and ϵ is not too large.

A more exact mathematical formulation of the recognition problem under affine transformations is as follows: The model is assumed to be accurate, represented by an ordered list of planar points on its contour: $P = \{\mathbf{p}_i = (\mathbf{p}_i^x, \mathbf{p}_i^y) : i = 1, .., n\}$ (throughout this paper vectors are assumed to be row vectors). The contour extracted from the image is digitized and possibly noisy, represented by an ordered list of m points (pixels) $Q = \{\mathbf{q}_i = (\mathbf{q}_i^x, \mathbf{q}_i^y) : i = 1, .., m\}$. Given P, Q, and $\epsilon \geq 0$, we say that Q ϵ-matches P under an affine transformation iff there exists a 2×2 non-singular matrix $\mathbf{A}$, a 2-D vector $\mathbf{b}$ such that for every j there is an i and for every i there is a j such that $||\mathbf{p}_i\mathbf{A} + \mathbf{b} - \mathbf{q}_j|| \leq \epsilon$.

It is more convenient to formulate the problem in the planar homogeneous coordinate system. Now P and Q are sets of 3-D vectors with 1 as their last coordinate. Given P, Q and ϵ, we say that Q ϵ-matches P under an affine transformation iff there exists a 3×3 non-singular matrix $\mathbf{A}$ of the form

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} , \tag{1}$$

such that $||\mathbf{p}_i\mathbf{A} - \mathbf{q}_j|| \leq \epsilon$. In this case, we say that Q ϵ-matches P under $\mathbf{A}$.

3 Matching Under Affine Transformations

Since a planar affine transformation has 6 parameters, it is determined uniquely by aligning a triplet of non collinear model points with a triplet of non collinear image points. Consequently, the recognition problem reduces to that of finding such triplets efficiently. For complexity-analysis purposes, we assume that m and n are of the same order of magnitude (otherwise the correct order parameter would be $\max(m, n)$). The "brute-force" method (introduced by Huttenlocher and Ullman [15]) is to choose 3 arbitrary well-spaced points in P, and for each triplet of points of Q, to solve for $\mathbf{A}$, and check if it aligns the remaining points. Note that checking the remaining points requires only $O(n)$ time because of the monotonicity of the points around the perimeter. Since there are $O(n^3)$ possible triplets of points in Q, the total time-complexity of this algorithm is $O(n^4)$. If it were possible to extract from P (and Q), in time $O(T(n))$, unique triplets of points such that a correspondence between P and Q exists *only if* these triplets correspond, checking the remaining points would require another $O(n)$, bringing the total time to $O(T(n) + n)$. A crucial matter is characterizing these unique triplets, then devising efficient algorithms for their detection. Fortunately, such triplets exist, and can be extracted in linear time, yielding a linear time recognition algorithm.

3.1 Affine-Invariant Extremal Geometric Properties

Using homogeneous coordinates, consider a triangle with vertices $T = \{\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$, and its image $S = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ under the affine transformation $\mathbf{A}$. Denote by

Area(T) the area of the triangle whose vertices are T, and by T the 3×3 matrix where the ith row is the homogeneous vector $\mathbf{t}_i$. Then

$$\text{Area}(S) = \frac{1}{2}\det(\mathbf{S}) = \frac{1}{2}\det(\mathbf{TA}) = \frac{1}{2}\det(\mathbf{T})\det(\mathbf{A}) = \det(\mathbf{A})\text{Area}(T) \ .$$

Since A is of the form (1), expansion of the determinant by the third column yields:

$$\det(\mathbf{A}) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \ .$$

This implies that for a fixed affine transformation, the area of a transformed triangle relates to the area of the original triangle by the same constant factor. For shape classes closed under affine transformations, such as the class of d-gons or the class of ellipses, this "area-preservation" property can lead to simple matching of such shapes across affine transformations. The "area-preservation" property implies that given a set P of planar points, the d-gon whose vertices are a subset of P and has extremal (minimal or maximal) area among all possible d-gons, is an affine invariant, that is, the transformed vertices are the vertices of the extremal d-gon in the image of P. This indicates that the vertices of an extremal area triangle are candidates for the "characterizing triplets" mentioned in the previous section. In our application, where there are many triplets of close points, the *maximal* area triangle is to be preferred over the minimal area triangle. Moreover, there exist efficient algorithms for its detection.

There is a wealth of literature describing efficient algorithms for computing extremal-area geometric shapes from point sets [6, 1, 2]. It is easily verified that both maximal area inscribed d-gons and minimal area circumscribing d-gons of a point set are functions of its convex hull only, so that a preprocessing stage computing the convex hull of the point set is likely to reduce the complexity of subsequent stages by reducing the number of points to be considered. Determining the convex hull of an n-point set usually requires $O(n \log n)$ time (see [19] for a variety of algorithms), but in the special case that the point set is given as an ordered list of the vertices of a polygon, the convex hull may be computed in *linear* time [17]. Effectively, this reduces our treatment of the point sets representing the model and the image to convex polygons. The most practical extremal geometric properties of a convex polygon P with n vertices are maximal area inscribed d-gons, whose vertices coincide with a subset of those of P. For the special cases $d = 3, 4$, this extremal d-gon may be computed in linear time, given the convex hull [10], otherwise (for $d \geq 5$) $O(dn + n \log n)$ is required [1]. Similarly, minimal area circumscribing d-gons require only $O(dn + n \log n)$ time [2]. On a similar note, computing minimal area circumscribing ellipses requires $O(n^2)$ time [18].

3.2 The Matching Algorithm

Based on the above, the basic model-image matching algorithm (without occlusion) is described.

The Algorithm.

Input: Polygons $P = \{\mathbf{p}_1, .., \mathbf{p}_n\}$ of model points,
$Q = \{\mathbf{q}_1, .., \mathbf{q}_m\}$ of image points and $\epsilon \geq 0$.
Output: The (possibly many) affine transformation matrices $\mathbf{A}$ such that P ϵ-matches Q under $\mathbf{A}$.

Method:

1. Compute the vertices $T = \{\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$ of the maximal area inscribed triangle of P.
2. Compute the vertices $S = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ of the maximal area inscribed triangle of Q.
3. For each (one of the 3) cyclic permutations of T:
   (a) Solve the matrix equation $\mathbf{TA} = \mathbf{S}$ for A.
   (b) If P ϵ-matches Q under $\mathbf{A}$ then output $\mathbf{A}$.

All the steps of the algorithm can be performed in linear time (Step 3(a) in constant time), so the total time complexity of the algorithm is $O(n)$. There is one major problem with the algorithm: the triangles T and S may not be unique, and even if they are, the presence of (even minor) noise in the image points might result in completely different triangles. The latter will typically occur when the maximal area inscribed triangle has area just slightly larger than that of the "runner up", and the noise corrupts the points in such a way that the "runner up" has an advantage and wins. One approach to solving the uniqueness problem is computing both the maximal-area inscribed triangle and quadrilateral of each point set, with the hope that at least one of them will lead to a correct match. In the case of the quadrilateral, step 3 of the algorithm is performed for each one of the 4 cyclic permutations of the vertices, and $\mathbf{A}$ is obtained as the *least-squares* solution of the resulting overdetermining set of equations (8 linear equations in 6 unknowns). The results of this approach will be particularly good if the triangle and quadrilateral are "uncorrelated", in the sense that their vertices are (relatively) disjoint. Figure 1 shows the maximal-area inscribed triangle and quadrilateral in a model and image curve pair. Another way of dealing with the uniqueness problem is to compute a constant number of the largest area triangles, instead of just the largest. Step 3 of the algorithm is then performed for all possible pairs of the triangles extracted from P and Q. From experiments, we have found that computing the *three* largest area triangles is sufficient, in the sense that if Q is an affine image of P, there is always a triangle among the three extracted from P that correctly matches one of the three extracted from Q. The algorithm of [10] may be simply adjusted to compute the three largest area triangles, increasing the run-time of the algorithm by only a constant factor. In practice, if the points on the curves are dense, the three largest triangles will be almost identical. In this case, they should be treated as one "nominal" triangle. Three "different" triangles in this sense should be extracted from each curve.

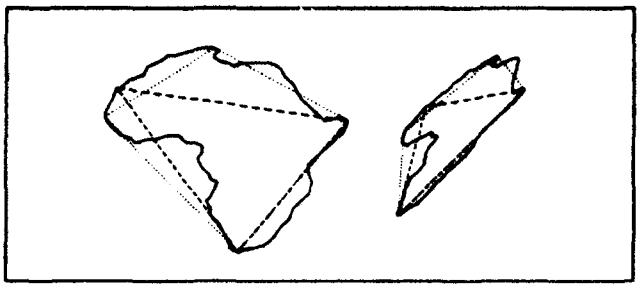


Fig. 1. The largest triangle and quadrilateral in a model curve and its affine transformed noisy image. In this case, solving for **A** from either the triangle or quadrilateral yields a correct match.

4 Sensitivity Analysis

First we explain why we prefer to use maximal area *inscribed* d-gons, and not minimal area *circumscribing* d-gons. The matrix **A** is obtained from the vertices of the computed d-gon. Since most of the edges of the circumscribing d-gon are tangent to the polygon representing the contour, even small displacements of the polygon vertices (due to noise) could lead to large displacements of the vertices of the circumscribing d-gon. This oversensitivity to noise, which does not exist in the case of inscribed d-gons, renders these types of extremal geometric shapes useless in our context.

Returning to inscribed triangles, the basic equation to be solved (for **A**) is:

$$\mathbf{TA} = \mathbf{S} \ .$$

This is obviously invariant to translations and similarity transformations, that is, such a transformation on **S** will result in the transformation being composed with **A**. Denote:

$$\Delta X(S) = \max\{\mathbf{s}_1^x, \mathbf{s}_2^x, \mathbf{s}_3^x\} - \min\{\mathbf{s}_1^x, \mathbf{s}_2^x, \mathbf{s}_3^x\} \ ,$$
$$\Delta Y(S) = \max\{\mathbf{s}_1^y, \mathbf{s}_2^y, \mathbf{s}_3^y\} - \min\{\mathbf{s}_1^y, \mathbf{s}_2^y, \mathbf{s}_3^y\} \ .$$

This invariance implies that, without loss of generality, we may transform to a coordinate system where the origin is contained in S, and

$$\Delta X(S)\Delta Y(S) > \text{Area}(S) \ . \tag{2}$$

This will enable us to obtain more precise estimates on the sensitivity of **S** to noise. Indeed, the solution **A** is sensitive to noise in the vertex matrices **T** and **S**. Assuming that **T** is exact and the true image triplet **S** is corrupted by noise

to $\mathbf{S}+\mathbf{E}$, where $\mathbf{E}$ is an error matrix, the relative error in the solution $\mathbf{A}+\Delta$ of the equation

$$\mathbf{T}(\mathbf{A}+\Delta)=(\mathbf{S}+\mathbf{E})$$

may be bounded by:

$$\frac{||\Delta||}{||\mathbf{A}||} \leq \kappa(\mathbf{S})\frac{||\mathbf{E}||}{||\mathbf{S}||} \ , \tag{3}$$

where $\kappa(\mathbf{S})$ is the *condition number* of the matrix $\mathbf{S}$ defined as

$$\kappa(\mathbf{S}) = ||\mathbf{S}|| \cdot ||\mathbf{S}^{-1}|| \tag{4}$$

(see [14], Chap. 5.8); hence the sensitivity of the solution $\mathbf{A}$ to noise in $\mathbf{S}$ is given by $||\mathbf{S}^{-1}||$. Note that $||\mathbf{S}^{-1}||$ is not dimensionless. This is because we assume the magnitude of the error $||\mathbf{E}||$ is independent of the magnitude of $\mathbf{S}$, as frequently happens in typical sampling scenarios. It would be convenient if the matrix corresponding to the specific inscribed triangle we work with, namely that with maximal area, had minimal $||\mathbf{S}^{-1}||$ among all triangles inscribed in $\mathbf{Q}$, but this is not the case. However, we are able to prove that it is not much larger than the smallest possible value among all inscribed triangles. Equations (3) and (4) are valid for any matrix norm. In what follows, we use the l_∞ norm, to simplify the calculations. This is not restrictive, as for any $0 < p < \infty$, there exist positive constants α_p, β_p such that

$$\alpha_p||\mathbf{A}||_\infty \leq ||\mathbf{A}||_p \leq \beta_p||\mathbf{A}||_\infty$$

(see [5] p. 32). Consequently, any relationship based on the l_∞ norm holds for any other l_p norm, with appropriate multiplicative factors.

The inverse of $\mathbf{S}$ is

$$\mathbf{S}^{-1} = \frac{1}{\det(\mathbf{S})}\begin{pmatrix} s_1^x - s_2^x\, s_1^y - s_2^y & s_1^x s_2^y - s_1^y s_2^x \\ s_2^x - s_3^x\, s_2^y - s_3^y & s_2^x s_3^y - s_2^y s_3^x \\ s_3^x - s_1^x\, s_3^y - s_1^y & s_1^x s_3^y - s_1^y s_3^x \end{pmatrix} \ . \tag{5}$$

Since S contains the origin, the third column of $\mathbf{S}^{-1}$ is just twice the area of S, and (5) reduces to

$$||\mathbf{S}^{-1}||_\infty = \frac{1}{2\text{Area}(S)}\max\{2\text{Area}(S), 2\Delta X(S), 2\Delta Y(S)\} \tag{6}$$

$$= \max\{1, \Delta X(S)/\text{Area}(S), \Delta Y(S)/\text{Area}(S)\} \ . \tag{7}$$

By (2), we obtain for the sensitivity of S to noise:

$$\text{Sens}(S) = ||\mathbf{S}^{-1}||_\infty = \max\{\Delta X(S)/\text{Area}(S), \Delta Y(S)/\text{Area}(S)\} \ . \tag{8}$$

A straightforward argument shows that for any triangle S

$$\frac{1}{4}\text{Perim}(S) \leq \max\{\Delta X(S), \Delta Y(S)\} \leq \frac{1}{2}\text{Perim}(S) \ .$$

Consequently, $\text{Sens}(S)$ of the triangle S which minimizes $\text{Perim}(S)/\text{Area}(S)$ is within a factor of 2 of the minimum value of (8) among all triangles inscribed

in Q, so the triangle minimizing $\text{Perim}(S)/\text{Area}(S)$ is in a sense the least sensitive to noise. This is the "fattest" triangle in Q. Unfortunately, this triangle is invariant only under simple similarity transformations, not under general affine transformations (see Fig. 2). However, we are able to prove that the maximal-area inscribed triangle has a perimeter/area ratio which is not much larger than this optimal one. This is not obvious as any triangle's perimeter/area ratio may be increased by an arbitrary factor λ by scaling it down by λ, so small triangles tend to have a large perimeter/area ratio. To prove our claim, we need the following two geometric inequalities:

Lemma 1. *If K is a planar convex region and T a triangle inscribed in K, then*

$$\text{Perim}(T)/\text{Area}(T) \geq \text{Perim}(K)/\text{Area}(K) \ .$$

Proof. It is easily verified (see Fig. 3) that for any planar convex region K with in-radius R (the radius of the largest inscribed circle), the inequality $2/R \geq \text{Perim}(K)/\text{Area}(K)$ holds, with equality (among other cases) when K is a triangle. Denote by r the in-radius of T. Since obviously $r \leq R$,

$$\text{Perim}(T)/\text{Area}(T) = 2/r \geq 2/R \geq \text{Perim}(K)/\text{Area}(K) \ .$$

□

Lemma 2 Appendix of [11]. *If S is an inscribed triangle of maximal area in a convex hexagon K, then*

$$\text{Area}(S) \geq 4/9 \ \text{Area}(K) \ .$$

We now relate the perimeter/area ratio of the maximal-area inscribed triangle to the optimal value. We show that it is not larger than 9/4 times this optimal value.

Theorem 3. *Let K be a convex planar region. Denote by S the maximal area triangle inscribed in K and by T the triangle inscribed in K with minimal perimeter/area ratio. Then*

$$\text{Perim}(S)/\text{Area}(S) \leq 9/4 \ \text{Perim}(T)/\text{Area}(T) \ .$$

Proof. Let K' be the (at most) hexagon whose vertices are those of S and T. Since S is maximal in K, it is maximal in K' too. Obviously $\text{Perim}(S) \leq \text{Perim}(K')$. This and Lemma 2 applied to K' yields:

$$\text{Perim}(S)/\text{Area}(S) \leq 9/4 \ \text{Perim}(K')/\text{Area}(K') \ . \tag{9}$$

But by Lemma 1

$$\text{Perim}(K')/\text{Area}(K') \leq \text{Perim}(T)/\text{Area}(T) \ . \tag{10}$$

Combining (9) and (10) completes the proof. □

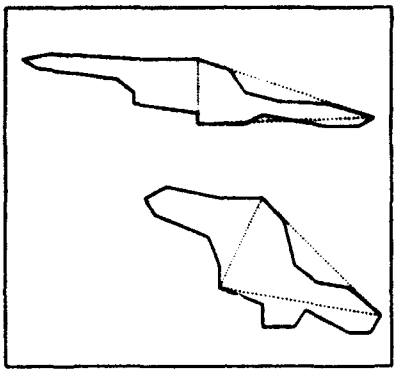


Fig. 2. The triangle with minimal perimeter/area ratio is not an affine invariant.

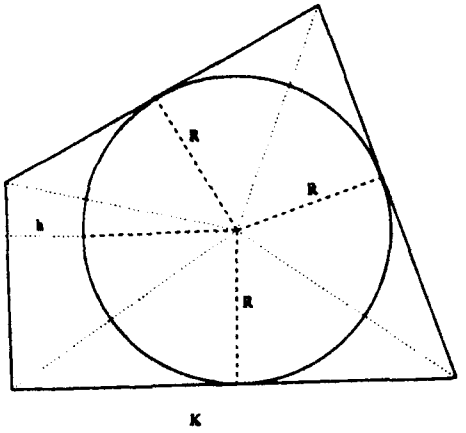


Fig. 3. The proof that Area(K) $\geq$ Perim(K)Inradius(K)/2 for a convex region K. Observe that the heights of the triangles composing K are no shorter than Inradius(K).

References

1. Aggarwal, A., Klawe, M.M., Moran, S., Shor, P., Wilber, R. (1987). Geometric applications of a matrix-searching algorithm, Algorithmica 2, pp. 195–208.
2. Aggarwal, A., Park, J. (1988). Notes on searching in multidimensional monotone arrays. In: Proc. 29th Symposium on Foundations of Computing, IEEE, pp. 497–512.
3. Arkin, E.M., Kedem, K., Mitchell, J.S.B., Sprinzak, J., Werman, M. (1991). Matching points into noise regions: Combinatorial bounds and algorithms, Proc. Second Symposium on Discrete Algorithms, SIAM, pp. 42–51.
4. Baird, H.S. (1985). Model Based Image Matching Using Location, MIT Press, Cambridge, MA.
5. Ben-Israel, A., Greville, T.N.E. (1974). Generalized Inverses, Wiley, New York.

6. Boyce, J.E., Dobkin, D.P., Drysdale, R.L., Guibas, L.J. (1985). Finding extremal polygons, SIAM Journal of Computing 14, pp. 134–147.
7. Bruckstein, A.M., Katzir, N., Lindenbaum, M., Porat, M. (1990). Similarity-invariant recognition of partially occluded planar curves and shapes, Proc. 7th Israeli Symposium on Artificial Intelligence and Computer Vision, pp. 45–59.
8. Bruckstein, A.M., Netravali, A.N. (1991). On differential invariants of planar curves and recognising partially occluded planar shapes, Proc. Int. Visual Form Workshop, pp. 89–98.
9. Cyganski, P., Orr, J., Cott, T., Dodson, P. (1987). Development, implementation, testing and application of an affine invariant curvature function, Proc. First Int. Conf. on Computer Vision, pp. 496–500.
10. Dobkin, D.P., Snyder, L. (1979). On a general method for maximising and minimising among certain geometric problems. In: Proc. 20th Symposium on Foundations of Computing, IEEE, pp. 9–17.
11. Fleischer, R., Mehlhorn, K., Rote, G., Welzl, E., Yap, C.K. (1990). On simultaneous inner and outer approximations of shapes, Proc. 6th Symposium on Computational Geometry, ACM, pp. 216–224.
12. Hong, J., Tan, X. (1988). A new approach to point pattern matching, Proc. 10th Int. Conf. on Pattern Recognition, IEEE, pp. 82–84.
13. Hopcroft, J.E., Huttenlocher, D.P. (1989). On planar point matching under affine transformation, Technical Report TR 89-986, Dept. of Computer Science, Cornell University.
14. Horn, R.A., Johnson, C.R. (1988). Matrix Analysis, Cambridge University Press, Cambridge, UK.
15. Huttenlocher, D.P., Ullman, S. (1987). Object recognition using alignment. In: Proc. First Int. Conf. on Computer Vision, pp. 102–111.
16. Lamdan, Y., Schwartz, J.T., Wolfson, H.J. (1988). Object recognition by affine invariant matching. In: Proc. Conf. on Computer Vision and Pattern Recognition, IEEE, pp. 335–344.
17. Melkman, A. (1987). On-line construction of the convex hull of a simple polyline, Information Processing Letters 25, pp. 11–12.
18. Post, M.J. (1981). A minimal spanning ellipse algorithm. In: Proc. 22nd Symposium on Foundations of Computing, IEEE, pp. 115–122.
19. Preparata, F.P., Shamos, I.M. (1985). Computational Geometry, Springer-Verlag, New York.
20. Sprinzak, J., Werman, M. (1990). Exact point matching. In: Proc. 7th Israeli Symposium on Artificial Intelligence and Computer Vision, pp. 31–43.
21. Vaz, R.F., Cyganski, D. (1990). Generation of affine invariant local contour feature data, Pattern Recognition Letters 11, pp. 479–483.

Recognizing 3-D Curves from a Stereo Pair of Images: a Semi-differential Approach *

Theo Moons[1], *Eric J. Pauwels*[1], *Luc J. Van Gool*[1], *Michael H. Brill*[2], *Eamon B. Barrett*[3]

[1] Katholieke Univ. Leuven, ESAT-MI2, K. Mercierlaan 94, 3001 Leuven, Belgium
[2] Science Appl. Int. Corp., 1710 Goodridge Drive, 1-11-1, McLean, VA 22102, USA
[3] Lockheed Missiles and Space Co., Space Systems Div. 1111 Lockheed Way, Sunnyvale, CA 94089-3504, USA

Abstract. This study investigates how the stereo view of a 3-D curve can be characterized in a way that is invariant under 3-D Euclidean motions of the curve. Depending on the knowledge about and the variability of the parameters of the stereo-setup, several transformation groups are identified as the framework relevant to the extraction of invariants. The point-derivative exchange principle can be used to construct a new set of semi-differential invariants for stereo pairs. The resulting invariants range from the quite simple to the rather complicated, but the information needed for their computation remains limited at all times. Finally, invariants from motion sequences as well as surface shape descriptors are proposed as special cases.

Keywords: shape description, semi-differential invariants, stereo view, space curves and surfaces, differential geometry, shape-from-motion.

1 Introduction

This paper is concerned with the recognition of 3-D curves from stereo-image stereo pairs. One way to proceed would be to recover a 3-D description of the curve and to calculate its Euclidean invariants, for example, curvature and torsion as a function of arc length [2]. Another strategy, however, is to try and to find invariants on the level of the image projections directly, that is, features that are invariant under the transformation group governing the curve projections when the curve undergoes 3-D Euclidean motions. This is the strategy followed in the paper. In particular, the possibility of eliminating the need for camera calibration via such an approach will be considered. To this end, one first has to investigate what kind of transformations a 3-D Euclidean motion of the 3-D scene induces in the image planes. Emphasis will be on identifying the *best suited* transformation group for the problem at hand. Here 'best suited' is to be understood as the *smallest* transformation group on the stereo image that

* The support of ESPRIT Basic Research Action EP 6448 'VIVA' and the Post-Doctoral Research Grants (T. Moons and E. Pauwels) of the Katholieke Universiteit Leuven are gratefully acknowledged.

encompasses all the transformations of the stereo pair induced by 3-D Euclidean motions of the scene curve. In a second step, invariants are computed for these transformations. They are obtained by the methods expounded in the literature [1, 4]. In particular, emphasis is put on the extraction of *semi-differential invariants* (i.e. invariants combining low order coordinate derivatives at several points simultaneously).

2 Towards the Optimal Transformation Group

Suppose a pair of images is taken with the stereo configuration of Fig. 1. Note

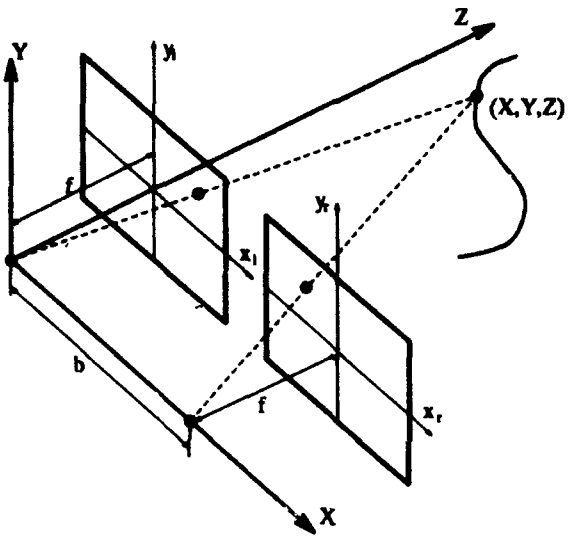


Fig. 1. Configuration with a parallel, aligned stereo-image pair.

that the image planes are coplanar. Furthermore, assume that the 3-D scene undergoes a 3-D Euclidean motion. Recall that a Euclidean motion is defined thus: a point with 3-D coordinates $\mathbf{x} = (X, Y, Z)^T$ is transformed into the point $\mathbf{x}' = R\,\mathbf{x} + V$, with rotation matrix $R = (r_{ij})$ and translation vector $V = (v_1, v_2, v_3)^T$. The more convenient linear representation $\mathbf{X}' = \mathbf{E}\,\mathbf{X}$ of this transformation is obtained by putting

$$\mathbf{X} = (X, Y, Z, 1)^T \quad \text{and} \quad \mathbf{E} = \begin{pmatrix} R & V \\ 0 & 1 \end{pmatrix} , \tag{1}$$

The two projections of the point will change accordingly. In fact, if the coordinates for the left and the right images before the Euclidean motion are written as (x_l, y) and (x_r, y) respectively, then the relation between the 3-D- and image-coordinates is

$$x_l = f\frac{X}{Z} , \quad x_r = f\frac{X-b}{Z} , \quad \text{and} \quad y = f\frac{Y}{Z} , \tag{2}$$

where f denotes the focal length of both cameras and b is the intercamera distance. Moreover, if the image coordinates after the Euclidean motion are (x_l', y') and (x_r', y'), then the new coordinate triple (x_l', x_r', y') is related to the old one (x_l, x_r, y) by a three-dimensional projective transformation, namely

$$x_l' = f\frac{(r_{11} + v_1/b)x_l + (-v_1/b)x_r + r_{12}y + r_{13}f}{(r_{31} + v_3/b)x_l + (-v_3/b)x_r + r_{32}y + r_{33}f},$$

$$x_r' = f\frac{(r_{11} - 1 + v_1/b)x_l + (1 - v_1/b)x_r + r_{12}y + r_{13}f}{(r_{31} + v_3/b)x_l + (-v_3/b)x_r + r_{32}y + r_{33}f}, \tag{3}$$

$$y' = f\frac{(r_{21} + v_2/b)x_l + (-v_2/b)x_r + r_{22}y + r_{23}f}{(r_{31} + v_3/b)x_l + (-v_3/b)x_r + r_{32}y + r_{33}f}.$$

The main disadvantage of projective transformations is their non-linearity. In mathematics, this disadvantage is circumvented by introducing homogeneous coordinates. For practical implementations, however, homogeneous coordinates are not suitable. It was observed in [1] that the above transformation formulae can be linearized by a change of coordinates:

Definition 1. The stereo pair of images of a scene point $(X, Y, Z)^T$ is (according to the situation depicted in Fig. 1) characterized by the 4-tuple

$$\mathbf{r} = \frac{1}{x_l - x_r}\begin{pmatrix} x_l \\ x_r \\ y \\ 1 \end{pmatrix}.$$

Using the *stereo coordinates* $\mathbf{r}$ has two advantages:

1. It follows from (2) that $\mathbf{r}$ depends linearly on the scene coordinates:

$$\mathbf{r} = \mathbf{U}\,\mathbf{X} \quad \text{with} \quad \mathbf{U} = \frac{1}{b}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{4}$$

2. A Euclidean motion $\mathbf{X}' = \mathbf{E}\,\mathbf{X}$ induces a linear transformation of the stereo coordinates $\mathbf{r}$:

$$\mathbf{r}' = \mathbf{U}\,\mathbf{E}\,\mathbf{U}^{-1}\mathbf{r}. \tag{5}$$

If an accurate and reliable estimate for the focal length f and the intercamera distance b can be obtained, then by (4) the scene coordinates $\mathbf{x}$ of a scene point can be computed directly from the stereo coordinates $\mathbf{r}$. Explicitly, if $\mathbf{r} = (p, p-1, q, w)^T$, then $\mathbf{x} = (bp, bq, bfw)^T$. A Euclidean motion (1) of the 3-D scene induces a Euclidean transformation $\mathbf{p}' = R\,\mathbf{p} + V/b$ of the image parameters $\mathbf{p} = (p, q, fw)^T$, for which we have the obvious two-point invariant $(p_1 - p_2)^2 + (q_1 - q_2)^2 + f^2(w_1 - w_2)^2$.

In general, determining the parameters b and f for the vision system accurately (part of the *calibration problem*) may well involve tedious procedures. A first reaction is to persevere with the projective transformations (3) and to look

for descriptions of the image curve which are invariant under all projective transformations. But then the question arises whether or not the transformations (3) really constitute the entire 3-D projective group. Put differently, does there exist a proper subgroup of the 3-D projective group that contains all the transformations (3)? The advantage of having a proper subgroup is clear: a smaller group allows one to distinguish more space curves on the basis of their stereo images; and moreover, smaller groups admit less complex invariants.

Depending on the knowledge about and the variability of the parameters of the stereo set-up, one can distinguish four different possibilities:

1. *b and f are constant and (accurately) known;*
2. *b and f are constant but unknown (or insufficiently well known);*
3. *b and f are variable (and unknown for that matter);*
4. *cameras with different focal lengths (but parallel image planes).*

The first case has already been solved above and leads to the 3-D Euclidean motion group, which is a 6-dimensional (Lie) group. The last one — which is, strictly speaking, not covered by the mathematical formalism introduced above — was discussed in [5] and yields the 3-D projective group as the relevant one. Note that this is a 15-dimensional (Lie) group. The large difference in dimensions strengthens our hope of finding proper subgroups for the remaining cases. Intermediate cases with one of the configuration parameters b or f fixed and the other variable are reducible to one of the cases considered earlier.

3 The General Stereo Group $GS(4)$

First consider case 3 in which b and f are variable (and unknown). This situation corresponds, for instance, to camera pairs having identical, but variable focal lengths. The relation between the stereo coordinates of a scene point before ($\mathbf{r}$) and after ($\mathbf{r}'$) a 3-D Euclidean motion is derived as follows: before the motion the image is obtained by a stereo configuration with parameters b_1 and f_1. According to (4), the stereo coordinates of a scene point $\mathbf{X}$ are given by $\mathbf{r} = \mathbf{U_1}\,\mathbf{X}$ where $\mathbf{U_1}$ is characterized by b_1 and f_1. After the motion an image is obtained by a stereo configuration with (possibly different) parameters b_2 and f_2. Consequently, the stereo coordinates of the scene point $\mathbf{X}'$ are given by $\mathbf{r}' = \mathbf{U_2}\,\mathbf{X}'$ with $\mathbf{U_2}$ defined by b_2 and f_2. Using $\mathbf{X}' = \mathbf{E}\,\mathbf{X}$, these expressions yield $\mathbf{r}' = \mathbf{U_2}\,\mathbf{E}\,{\mathbf{U_1}}^{-1}\,\mathbf{r}$. Hence, the relevant transformations are $\{\mathbf{U_2}\,\mathbf{E}\,{\mathbf{U_1}}^{-1} \mid \mathbf{E}$ as in (1) and $\mathbf{U_1}, \mathbf{U_2}$ as in (4)$\}$. Unfortunately, this set of transformations does *not* constitute a group, since it is not closed under matrix multiplication. In order to get invariants, we first have to determine the enclosing group. An arbitrary element of this set of transformations is of the form

$$\mathbf{G} = \frac{1}{b_2}\begin{pmatrix} b_1 r_{11} + v_1 & -v_1 & b_1 r_{12} & b_1 f_1 r_{13} \\ b_1 r_{11} + v_1 - b_2 & -v_1 + b_2 & b_1 r_{12} & b_1 f_1 r_{13} \\ b_1 r_{21} + v_2 & -v_2 & b_1 r_{22} & b_1 f_1 r_{23} \\ (b_1 r_{31} + v_3)/f_2 & -v_3/f_2 & b_1 r_{32}/f_2 & b_1 f_1 r_{33}/f_2 \end{pmatrix}, \qquad (6)$$

and thus belongs to the following subgroup of the general linear group $GL(4)$.

Definition 2. The *general stereo group* is defined as

$$GS(4) = \{(g_{ij}) \in GL(4) \mid g_{21} = g_{11} - 1, \quad g_{22} = g_{12} + 1, \quad g_{23} = g_{13}, \\ g_{24} = g_{14}, \text{ and } \det(g_{ij}) > 0\} \ .$$

Note that $GS(4)$ is a 12-dimensional Lie group.

On the other hand, replacing the 3×3 rotation matrix R, i.e. $R \in SO(3)$, in $\mathbf{E}$ by a general invertible 3×3 matrix A, i.e. $A \in GL(3)$, *does* yield a group! This means that by extending the possible transformations on the original 3D curve from Euclidean to affine motions, affine transformation causes the correspondingly induced tranformations on $\mathbf{r}$ to constitute a subgroup of the general stereo group $GS(4)$. Unfortunately, the following proposition shows that all these groups actually coincide with $GS(4)$. The notation $\mathrm{grp}\, S$ stands for the smallest subgroup of $GS(4)$ containing the subset S.

Proposition 3.

$$GS(4) = grp\ \{\mathbf{U_1} \begin{pmatrix} R & V \\ 0 & 1 \end{pmatrix} \mathbf{U_2}^{-1} \mid R \in SO(3),\ V \in \mathbb{R}^3 \textit{ and } \mathbf{U_1}, \mathbf{U_2} \textit{ as in (4)}\}$$

$$= \{\mathbf{U_1} \begin{pmatrix} A & V \\ 0 & 1 \end{pmatrix} \mathbf{U_2}^{-1} \mid A \in GL(3),\ \det A > 0 \textit{ and } \mathbf{U_1}, \mathbf{U_2} \textit{ as in (4)}\} \ .$$

Proof. Denote the group at the right-hand side of the first equality by H, and that of the second equality by K. It is clear that $H \subset K \subset G(4)$ are Lie subgroups of $GS(4)$. It suffices to show that $H = GS(4)$. The Lie algebra $\mathcal{H}$ of H is a Lie subalgebra of the Lie algebra $\mathcal{GS}(4)$ of $GS(4)$. If these two Lie algebras coincide, then so do H and $GS(4)$, since $GS(4)$ is a connected Lie group. (All properties of Lie groups and their algebras used in this proof can be found in [3].) Hence, it suffices to show that $\mathcal{H} = \mathcal{GS}(4)$. But Lie algebras being vector spaces, one needs only to prove that they have equal dimensions: $\dim \mathcal{H} = \dim \mathcal{GS}(4) = \dim GS(4) = 12$. Since $\mathcal{H}$ is a linear subspace of $\mathcal{GS}(4)$, the claim will be settled if $\mathcal{H}$ has *at least* dimension 12. To prove this, we compute the tangent vectors at the identity matrix to some particular curves in H. With the notation of (6), the following curves are used:

- $(r_{ij}) = \tilde{R}_i(\theta)$ is the 1-parameter subgroup of $SO(3)$ formed by the rotations about the X, Y, Z-axis respectively, $v_j = 0$, and $b_j = f_j = 1$;
- $(r_{ij}) = \mathbf{I_3}$ the (3×3)-identity matrix, $V = \tilde{T}_i(\theta)$ the 1-parameter subgroup of $\mathbb{R}^3$ formed by the translations in the X, Y, Z-directions respectively, and $b_j = f_j = 1$;
- $(r_{ij}) = \mathbf{I_3}$, $v_j = 0$, $b_1/b_2 = e^{\theta}$ and $f_j = 1$;
- $(r_{ij}) = \mathbf{I_3}$, $v_j = 0$, $b_j = 1$, $f_1 = 1$ and $f_2 = e^{-\theta}$.

The corresponding tangent vectors at $\theta = 0$ are:

$$\mathbf{R_1} = \begin{pmatrix} 0&0&-1&0\\ 0&0&-1&0\\ 1&0&0&0\\ 0&0&0&0 \end{pmatrix}, \mathbf{R_2} = \begin{pmatrix} 0&0&0&1\\ 0&0&0&1\\ 0&0&0&0\\ -1&0&0&0 \end{pmatrix}, \mathbf{R_3} = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&-1\\ 0&0&1&0 \end{pmatrix},$$

$$\mathbf{T_1} = \begin{pmatrix} 1&-1&0&0\\ 1&-1&0&0\\ 0&0&0&0\\ 0&0&0&0 \end{pmatrix}, \mathbf{T_2} = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 1&-1&0&0\\ 0&0&0&0 \end{pmatrix}, \mathbf{T_3} = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 1&-1&1&0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1&0&0&0\\ 1&0&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&1 \end{pmatrix}.$$

These matrices belong to the Lie algebra $\mathcal{H}$. Therefore, their Lie bracket $[\mathbf{P}, \mathbf{Q}] = \mathbf{PQ} - \mathbf{QP}$ also belongs to $\mathcal{H}$. In particular, this holds for

$$\mathbf{S_2} = [\mathbf{R_2}, \mathbf{F}] = \begin{pmatrix} 0&0&0&1\\ 0&0&0&1\\ 0&0&0&0\\ 1&0&0&0 \end{pmatrix}, \quad \mathbf{S_3} = [\mathbf{R_3}, \mathbf{F}] = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&0&-1\\ 0&0&-1&0 \end{pmatrix},$$

$$\mathbf{S_1} = [\mathbf{R_2}, \mathbf{S_3}] = \begin{pmatrix} 0&0&-1&0\\ 0&0&-1&0\\ -1&0&0&0\\ 0&0&0&0 \end{pmatrix}, \quad \mathbf{C} = [\mathbf{R_3}, \mathbf{S_3}] = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&2&0\\ 0&0&0&-2 \end{pmatrix}.$$

A routine calculation shows that $\mathbf{R_1}, \mathbf{R_2}, \mathbf{R_3}, \mathbf{T_1}, \mathbf{T_2}, \mathbf{T_3}, \mathbf{B}, \mathbf{F}, \mathbf{S_1}, \mathbf{S_2}, \mathbf{S_3}, \mathbf{C}$ are 12 linearly independent elements of $\mathcal{H}$. □

An important consequence of this proposition is that stereo invariants for the curve undergoing a 3-D Euclidean motion are also invariant for 3-D (positive) affine transformations of that curve. Consequently, *it is impossible to distinguish between affine and Euclidean transformations of a space curve on the basis of a stereo pair of images alone when the configuration parameters (b and f) are variable.*

4 Semi-Differential Invariants for $GS(4)$

Once the appropriate transformation group is identified, one can compute sets of independent invariants for this group. Since Euclidean motions in the scene induce linear transformations $\mathbf{r}' = \mathbf{G}\,\mathbf{r}$, with $\mathbf{G} \in GS(4)$, on the stereo coordinates $\mathbf{r}$, the value of a determinant $|\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3\ \mathbf{r}_4|$ whose columns are the stereo coordinates of 4 scene points will be multiplied with $|\mathbf{G}| = \det G$ if each of the $\mathbf{r}_i$ is replaced by $\mathbf{r}'_i$. If the scene point belongs to a 3-D curve, its image traces out a curve in the space of stereo coordinates, and the derivatives of $\mathbf{r}$ w.r.t. an (arbitrary) parameter transform according to the same linear transformation.

Therefore replacing a column $\mathbf{r}_i$ in the determinant above by a jth-order derivative $\mathbf{r}_i^{(j)}$ does not disadvantage its transformation behaviour. In particular, one can distinguish between the following essentially different determinants:

$$|\mathbf{r}'\ \mathbf{r}'_1\ \mathbf{r}'_2\ \mathbf{r}'_3| = |\mathbf{G}|\ |\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|, \qquad |\mathbf{r}'\ \mathbf{r}'^{(1)}\ \mathbf{r}'_1\ \mathbf{r}'^{(1)}_1| = |\mathbf{G}|\dot{u}_1\dot{u}|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}^{(1)}_1|\ ,$$
$$|\mathbf{r}'\ \mathbf{r}'^{(1)}\ \mathbf{r}'_1\ \mathbf{r}'_2| = |\mathbf{G}|\dot{u}\,|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_2|, \qquad |\mathbf{r}'\ \mathbf{r}'^{(1)}\ \mathbf{r}'^{(2)}\ \mathbf{r}'_1| = |\mathbf{G}|\dot{u}^3|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}_1|\ ,$$
$$|\mathbf{r}'\ \mathbf{r}'^{(1)}\ \mathbf{r}'^{(2)}\ \mathbf{r}'^{(3)}| = |\mathbf{G}|\dot{u}^6|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}^{(3)}|\ ,$$

where subscripts denote measurements at fixed reference points, superscripts indicate the order of derivative with respect to the curve parameter, and $\dot{u}$ is the derivative of the curve parameter used before the transformation with respect to the curve parameter which was used after the transformation. Invariants are now obtained by taking appropriate ratios of products of these determinants. Using the techniques explained in [4], one arrives at the following list of semi-differential invariants for $GS(4)$ (abs denoting the absolute value).

Absolute semi-differential invariants:

Case 1: $\frac{|\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|}{|\mathbf{r}_4\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|}$.

Case 2: $\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_2|}{|\mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|}$ and $\int \mathrm{abs}\left(\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_2|}{|\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|}\right) dt$.

Case 3: $\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}|}{|\mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}\ \mathbf{r}_2|}$ and $\int \mathrm{abs}\left(\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}|}{|\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}\ \mathbf{r}_2|}\right) dt$.

Case 4: $\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}_1|}{|\mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}|}$ and $\int \mathrm{abs}\left(\frac{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}_1|}{|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}|}\right) dt$.

Admitting derivatives in the building determinants introduces the reparameterization factor $\dot{u}$ in the transformation formulae, as indicated above. Quotients of two determinants are invariant for the transformation group $GS(4)$, but some combinations are not invariant under reparameterization of the given curve. However, these combinations F can easily be made invariant under reparameterization by integration: $\int \mathrm{abs}(F)\,dt$. The absolute value is taken in order to obtain a nondecreasing function which may then be used as an *invariant* parameter.

Invariants invoking derivatives of higher order may also be derived, but we stop the list here, because they are very hard to obtain in practical situations. Removing the denominator from the expressions above also increases accuracy. By doing so, one gets the following, essentially different, relative invariants.

Relative semi-differential invariants:

Case 1: $|\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3|$ is a relative invariant.

Case 2: $\int \mathrm{abs}\left(|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_2|\right) dt$ is a relative invariant parameter.

Case 3: $\int \mathrm{abs}\left(|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1)}|\right) dt$ is a relative invariant parameter.

Case 4: $\int \mathrm{abs}\left(|\mathbf{r}\ \mathbf{r}^{(1)}\ \mathbf{r}^{(2)}\ \mathbf{r}_1|^{\frac{1}{3}}\right) dt$ is a relative invariant parameter.

5 The Special Stereo Group $SS(4)$

Now consider the case in which b and f are constant but unknown. This situation corresponds, for example, to the case where images are taken from a moving 3-D curve with a fixed camera set-up. In this case, the relation between the scene and the stereo coordinates of a scene point is, before as well as after a 3-D Euclidean motion, given by the same configuration matrix $\mathbf{U}$ which is characterized by b and f. More precisely, $\mathbf{r} = \mathbf{U}\,\mathbf{X}$ and $\mathbf{r}' = \mathbf{U}\,\mathbf{X}'$ (cf. (4)). Since $\mathbf{X}' = \mathbf{E}\,\mathbf{X}$, these expressions yield $\mathbf{r}' = \mathbf{U}\,\mathbf{E}\,\mathbf{U}^{-1}\,\mathbf{r}$, and the relevant transformations are $\{\mathbf{U}\,\mathbf{E}\,\mathbf{U}^{-1} \mid \mathbf{E}$ as in (1) and $\mathbf{U}$ as in (4)$\}$. Again, this set of transformations does *not* form a group, and, in order to get invariants for these particular transformations, we first have to determine the group generated by them:

$$\text{grp}\,\{\mathbf{U}\begin{pmatrix} R & V \\ 0 & 1 \end{pmatrix}\mathbf{U}^{-1} \mid R \in SO(3),\ V \in \mathbb{R}^3 \text{ and } \mathbf{U} \text{ as in (4)}\}\ . \qquad (7)$$

As each generator $\mathbf{U}\mathbf{E}\mathbf{U}^{-1}$ of this group has determinant 1, the same property holds for all group elements. This imposes an additional constraint on the $GS(4)$ parameters, yielding to a lower-dimensional subgroup. Hence this group is contained in the following subgroup of $GS(4)$.

Definition 4. The *special stereo group* is defined as

$$SS(4) = \{\mathbf{G} \in GS(4) \mid \det \mathbf{G} = 1\}\ .$$

$SS(4)$ is an 11-dimensional Lie group. Following the same lines of reasoning as in the proof of Proposition 3, one shows that the transformation group defined in (7) actually coincides with $SS(4)$.

Proposition 5.

$$SS(4) = grp\,\{\mathbf{U}\begin{pmatrix} R & V \\ 0 & 1 \end{pmatrix}\mathbf{U}^{-1} \mid R \in SO(3),\ V \in \mathbb{R}^3 \textit{ and } \mathbf{U} \textit{ as in (4)}\}\ .$$

All relative invariants of $GS(4)$ are absolute invariants of $SS(4)$. Moreover, all invariants of $SS(4)$ are absolute invariants, since $SS(4)$ is the semi-direct product of a simple Lie group with a commutative one. Allowing affine transformations instead of the rotations R in the defining equation (7) of this group would destroy the property of having a unit determinant. In particular, allowing affine transformations brings us back to the (larger) group $GS(4)$. Taking only Euclidean motions into account therefore effectively simplifies the absolute invariants: 4 points and/or derivatives suffice. Moreover, the singularities complicating the use of $GS(4)$'s absolute invariants are eliminated, since the invariants of $SS(4)$ no longer have denominator determinants.

6 Shape from Image Sequences

The previous ideas on shape extraction through stereo find their natural translation to shape-from-motion by considering two subsequent frames as a stereo pair. In order for the coplanar setup to be relevant, the camera motion is restrained to translation along the image x-axis (i.e. horizontal translations leaving the image plane invariant). Similarly, the camera is assumed to have fixed intrinsic parameters during the motion fragment from which subsequent frames are taken. The motion of the object(s) under observation is assumed negligible between the two frames. In that case, one arrives at the $GS(4)$ set-up. This adaptation corresponds to the first step in the sequence below, where v denotes the image velocity of the point and dt the time elapsed between the frames. Since all the invariants are known to be based on determinants of such vectors, the x in $x + vdt$ can be eliminated. Carrying out this simplification and rearranging the vector components yields the second step:

$$\frac{1}{x_l - x_r}\begin{pmatrix} x_l \\ x_r \\ y \\ 1 \end{pmatrix} \longrightarrow -\frac{1}{vdt}\begin{pmatrix} x \\ x + vdt \\ y \\ 1 \end{pmatrix} \longrightarrow -\frac{1}{dt}\begin{pmatrix} x/v \\ y/v \\ 1/v \\ dt \end{pmatrix} .$$

As before, one can consider coordinate derivatives, where care has to be taken since the original vectors already contain velocity, i.e. temporal derivatives. This leads to the semi-differential invariants found for $GS(4)$, but now simply re-expressed in terms of the motion-sequence oriented vectors. Elapsed time dt should be known to replace the vectors in cases where $SS(4)$ is relevant.

7 Surface Shapes

The techniques put forward in this paper might also prove useful for the description of textured surfaces. When the surface contains point markings or a curve drawn on it, the previous results apply. However, surfaces might contain some specific texture types that add to the possibilities of the semi-differential technique. For instance, several drawn (or otherwise identifiable) curves might run through a single point on the surface. Although in principle there may be many, for the sake of argument let there be two. The parameter of one will be called u, that of the other v. The availability of several curves leads to simpler invariants as is illustrated by the following results for two points with two curves through each:

$$\frac{|\mathbf{r}\ \mathbf{r}^{(1:u)}\ \mathbf{r}^{(1:v)}\ \mathbf{r}_1||\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:u)}\ \mathbf{r}_1^{(1:v)}|}{|\mathbf{r}\ \mathbf{r}^{(1:u)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:v)}||\mathbf{r}\ \mathbf{r}^{(1:v)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:u)}|} \quad \text{and} \quad \frac{|\mathbf{r}\ \mathbf{r}^{(1:u)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:u)}||\mathbf{r}\ \mathbf{r}^{(1:v)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:v)}|}{|\mathbf{r}\ \mathbf{r}^{(1:u)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:v)}||\mathbf{r}\ \mathbf{r}^{(1:v)}\ \mathbf{r}_1\ \mathbf{r}_1^{(1:u)}|} .$$

If the surface pattern consists of dense point textures with some identifiable points embedded, then spatial derivatives at the identifiable points can be estimated. Analysing the problem with different sets of curvilinear coordinates

implies that we also have to guarantee invariance under transitions from one set of curvilinear coordinates to another:

$$\begin{pmatrix} \partial f/\partial u' \\ \partial f/\partial v' \end{pmatrix} = \mathbf{T} \begin{pmatrix} \partial f/\partial u \\ \partial f/\partial v \end{pmatrix} \qquad \text{where} \qquad \mathbf{T} = \begin{pmatrix} \partial u/\partial u' & \partial v/\partial u' \\ \partial u/\partial v' & \partial v/\partial v' \end{pmatrix} .$$

Using the image coordinates as the parameterization of a surface patch gives

$$|\mathbf{r}^{(1:x')}\ \mathbf{r}^{(1:y')}\ \mathbf{r}_1\ \mathbf{r}_2| = |\mathbf{T}|\ |\mathbf{r}^{(1:x)}\ \mathbf{r}^{(1:y)}\ \mathbf{r}_1\ \mathbf{r}_2| .$$

It is supposed that $\mathbf{r}, \mathbf{r}_1, \mathbf{r}_2$ are all 'identifiable' points and that the derivatives are calculated on the basis of neighbouring texture points. One cannot expect the same texture points to be found in different sequences. It is assumed only that there are sufficient texture points to always allow the estimation of the spatial derivatives along the image coordinate axis directions by tracking some texture points over a limited number of subsequent frames. Reparameterization is more restrictive than in the two-curve case, in the sense that it requires the combined use of both partial derivatives. The following expression is invariant under changes in viewpoint:

$$\frac{|\mathbf{r}\ \mathbf{r}_1\ \mathbf{r}^{(1:x)}\ \mathbf{r}^{(1:y)}|}{|\mathbf{r}\ \mathbf{r}_2\ \mathbf{r}^{(1:x)}\ \mathbf{r}^{(1:y)}|} .$$

8 Conclusions

This study considers the problem of constructing the simplest possible semi-differential invariants for the recognition of nonplanar curves from stereo images and motion sequences. The importance of selecting the appropriate transformation group for each set of conditions was highlighted. The rationale behind the different conditions was the elimination of camera calibration.

References

1. Brill, M.H., Barrett, E.B., Payton, P.M. (1992). Projective invariants for curves in two and three dimensions. In: Mundy, J., Zisserman, A. (eds.), Geometric Invariance in Computer Vision, MIT Press, Cambridge, Massachusetts, pp. 193–214.
2. Kishon, E., Hastie, T., Wolfson, H. (1991). 3-D curve matching using splines, Journal of Robotic Systems 8(6), pp. 723–743.
3. Sagle, A.A., Walde, R.E. (1973). Introduction to Lie Groups and Lie Algebras, Pure and Applied Mathematics, Vol. 51, Academic Press, New York.
4. Van Gool, L., Moons, T., Pauwels, E.J., Oosterlinck, A., (1992). Semi-differential invariants. In: Mundy, J., Zisserman, A. (eds.), Geometric Invariance in Computer Vision, MIT Press, Cambridge, Massachusetts, pp. 193–214.
5. Van Gool, L.J., Brill, M.H., Barrett, E.B., Moons, T., Pauwels, E.J. (1992). Semi-differential invariants for nonplanar curves. In: Mundy, J., Zisserman, A. (eds.), Geometric Invariance in Computer Vision, MIT Press, Cambridge, Massachusetts, pp. 293–309.

Statistical Shape Methodology in Image Analysis *

John T. Kent and Kanti V. Mardia

Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

Abstract. One of the main goals of high-level image analysis is object recognition. It is assumed that from training data or otherwise, a template for the object is available and the image contains a deformation of this underlying template. This paper describes and unifies some of the statistical methodology in this field. Some operational definitions of shape are given and then methods of registration including local versus global considerations are decribed. Using the Bayesian paradigm, methods for reconstructing shapes from deformable templates are described.

Keywords: shape, registration, Bayesian paradigm, landmarks, grey-level object, deformation, Procrustes analysis, shape coordinates, principal component analysis.

1 Introduction

One of the goals of high-level image analysis is the identification and description of objects in images. Further, when objects can vary from an underlying template, suitable allowance for deformations must be incorporated. The purpose of this paper is to describe and unify some of the statistical methodology which can be used in this field. The objective is not to be comprehensive but merely to illustrate statistical ideas which have been found illuminating in some interesting applications.

In Sect. 2 several different sorts of objects of interest are described, including sets of landmarks, outlines, solid objects and grey-level images. The concept of "shape" arises when certain transformations applied to an object (e.g. changes in size and rotation) are deemed to leave the essential properties of the object unchanged. The study of shape will be approached from a statistical point of view.

* The authors are grateful to Fred Bookstein for generous comments, and to Ian Dryden, Colin Goodall, John Haddon, Charles Taylor and Alistair Walder for helpful discussions. This work was supported by an SERC-MOD grant under the Complex Stochastic Systems Initiative.

The topic of registration is discussed in Sect. 3. Registration is important so that objects can be superimposed for the purposes of averaging and comparison. The actual fitting of objects to images is reviewed briefly in Sect. 4 using the Bayesian paradigm.

Bold-face type is used to represent vectors $\mathbf{z}$ and matrices $\mathbf{A}$. Also $\mathbf{z}^T$ denotes the transpose of a vector, $\bar{\mathbf{z}}$ the complex conjugate, and $\mathbf{z}^* = (\bar{\mathbf{z}}^T)$ the adjoint. Let $\mathbf{1}$ and $\mathbf{0}$ denote column vectors of ones and zeros, respectively. The dimension will be clear from the context.

2 Definitions of Shape

Intuitively the *shape* of an object is the underlying structure or pattern in the object after ignoring irrelevant transformations. Possible irrelevant transformations include translations, changes in scale, rotations, and shear or some subset thereof. The set of transformations is assumed to form a group which we shall call the *equivalence group*. The exact definitions of "object", "shape", "pattern", etc., depend on the context of the problem, and several approaches to the study of shape are listed below.

There are also other ingredients in shape analysis. First, for data analysis it is helpful to have a finite-dimensional space of shapes, or parametric model. In practice the dimensions of this model will be much lower than the number of pixels in an image. The next step is to describe a distance between shapes. This distance often depends upon an underlying idealized shape or "template". In many applications the objective is to identify a deformed version of the template in an image. For the rest of this section some of the statistical approaches that have been used to describe shape are outlined.

2.1 Landmark-based Objects

Let an "object" be a set of n points or *landmarks* in $\mathbb{R}^2$ or $\mathbb{R}^3$, or more generally $\mathbb{R}^p$, arranged as an $n \times p$ matrix $\mathbf{X} = \{\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T\}$. In practice the landmarks are locations which can be consistently identified on different geometric bodies. Examples include tips of fingers on a two-dimensional view of a hand, the tip of the nose on a three-dimensional human face, etc. The equivalence group is usually taken to be the special similarity group, formed from translations, scale changes and rotations. That is, two objects $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are regarded as having the same shape if $\mathbf{x}_i^{(1)} = \mathbf{a} + \gamma \mathbf{B} \boldsymbol{x}_i^{(2)}$, $i = 1, \ldots, n$, where $\mathbf{a}$ is a translation vector of length p, $\gamma > 0$ is a scale change and $\mathbf{B}(p \times p)$ is a rotation matrix.

Various distances have been proposed in this context. For simplicity we focus attention on the 2-dimensional case with a point $\mathbf{x}$ in $\mathbb{R}^2$ regarded as a complex number z. An object is then a complex vector $\mathbf{z} = (z_1, \ldots, z_n)^T$. Without loss of generality suppose $\mathbf{z}$ has been centered and standardized so that $\sum z_i = 0$, $\mathbf{z}^*\mathbf{z} = \sum |z_i|^2 = 1$. Rotation has not been standardized here. Possible distances include the following:

1. *Procrustes distance*: $d^2(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = 1 - |\mathbf{z}^{(1)*}\mathbf{z}^{(2)}|^2$. This is closely related to Kendall's [19] geodesic distance in shape space. Goodall [15] and Kent [20] give some other minor variations.
2. *Euclidean distance in Bookstein [6] coordinates*: $(\mathbf{w}^{(1)} - \mathbf{w}^{(2)})^*(\mathbf{w}^{(1)} - \mathbf{w}^{(2)})$ where $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ are two shape vectors in Bookstein coordinates. Bookstein coordinates ware defined in terms of the above spherical coordinates $\mathbf{z}$ by

$$w_i = (z_i - z_1)/(z_2 - z_1) \in \mathbb{C} \ , \ i = 1, \ldots, n.$$

Thus in Bookstein coordinates the object has been transformed so that the first landmark lies at $w_1 = 0$ and the second landmark at $w_2 = 1$.

2.2 Outline-based Objects

In this case the "object" is a (nonintersecting) closed curve in $\mathbb{R}^2 = \mathbb{C}$, $\{f(t) : t \in [0, 1]\}$ with $f(0) = f(1)$. Typically f represents the outline of a geometric body. It is assumed that the parameter t means the same thing on different curves and this correspondence is often obtained by fitting smooth curves (possibly straight lines) between a few landmarks identified along the outline.

Given a set of n landmarks along a template curve it is straightforward to construct a finite-dimensional space of objects. First the n landmarks are allowed to move freely in $\mathbb{C}$, forming a vector space of real dimension $2n$. Next, to interpolate between a pair of consecutive landmarks, a similarity transformation can be uniquely defined to map the two consecutive landmarks in the template to the two corresponding landmarks on the new object. This linear transformation can then be used to map the template curve between the two landmarks to the new object. In practice only small deformations are used so that the new curve will remain nonintersecting.

Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ denote the landmarks of the template and $\mathbf{z} = (z_1, \ldots, z_n)^T$ the landmarks of the deformed template. Set $\eta_i = \mu_i - \mu_{i-1}$ and $e_i = z_i - z_{i-1}$ to be the edges in the template and its deformation, respectively, with the index i interpreted cyclically on $\{1, 2, \ldots, n\}$. The similarity transformation between η_i and e_i can be written as a complex number, $1 + t_i$, where

$$e_i = (1 + t_i)\eta_i \ , \qquad i = 1, \ldots, n.$$

Grenander *et al.* [18] proposed a conditional autoregressive Gaussian model for $\mathbf{t} = (t_1, \ldots, t_n)^T$. This model has the property that the conditional distribution of t_i given the values at the other sites depends only on the nearest neighbours,

$$E[t_i \mid t_j, j \neq i] = \lambda(t_{i-1} + t_{i+1}) \ , \qquad \mathrm{var}[t_i \mid t_j, j \neq i] = \sigma^2,$$

where $0 \leq \lambda < \frac{1}{2}$ and $\sigma^2 > 0$. The model penalizes two aspects of the deformation: the size of the deformation (large values of $|t_i|$) and the lack of smoothness (when t_i is unlike its neighbours).

It can be shown that the quadratic form in the normal density for this model can be written in the form of a squared distance $d^2(\mathbf{z}, \boldsymbol{\mu}) = (\mathbf{z} - \boldsymbol{\mu})^T \mathbf{A}(\mathbf{z} - \boldsymbol{\mu})$,

where $\mathbf{A}$ is an $n \times n$ positive semidefinite matrix satisfying $\mathbf{A1} = \mathbf{0}$. See [25] for a detailed calculation of the entries of $\mathbf{A}$.

In this case the equivalence group contains only translations (since $\mathbf{A1} = \mathbf{0}$). Matching of scale and rotation must be done by pre-registration (see Sect. 3).

2.3 Solid Planar Objects

Let an "object" now denote a parameterized region $D \subset [0,1]^2 \subset \mathbb{R}^2$, expressed as a $0-1$ function $f(t) \in \{0,1\}$ for $t \in [0,1]^2$. Thus $f(t) = 1$ if and only if $t \in D$. As for outline-based objects, attention is usually focused on a finite set of landmarks $z_1, \ldots, z_n$ so that when comparing two objects, points in one object can be identified with the corresponding points on the other object. Such objects typically arise as a silhouette of a three-dimensional body.

Bookstein ([7], [8, Sect. 2.2]) proposed a parametric family of objects based on deforming an underlying template using a pair of thin-plate splines. A natural distance between a test object with landmarks $\mathbf{z}$ and a template with landmarks $\boldsymbol{\mu}$ can be defined by a quadratic form proportional to *bending energy*,

$$d^2(\mathbf{z}, \boldsymbol{\mu}) = (\mathbf{z} - \boldsymbol{\mu})^* \mathbf{B} (\mathbf{z} - \boldsymbol{\mu}) \equiv \mathbf{z}^* \mathbf{B} \mathbf{z} \ , \tag{1}$$

where the real-valued $(n \times n)$ matrix $\mathbf{B}$ is determined by the underlying template $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ and the thin-plate spline construction. Write $\boldsymbol{\mu} = \boldsymbol{\nu} + i\boldsymbol{\eta}$ and set $\mathbf{X} = (\mathbf{1}, \boldsymbol{\nu}, \boldsymbol{\eta})$ to be an $(n \times 3)$ matrix. Let $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ denote the projection matrix onto the orthogonal complement of the column space of $\mathbf{X}$, and define a matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ by $\sigma_{ij} = |\mu_i - \mu_j|^2 \log |\mu_i - \mu_j|$. Then it can be shown that $\mathbf{B} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}$. Details are given in [25].

Strictly speaking, the equivalence group is just the translation and rotation group here. However, the matrix $\mathbf{B}$ has the property that $\mathbf{B1} = \mathbf{0}$, $\mathbf{B}\boldsymbol{\nu} = \mathbf{0}$ and $\mathbf{B}\boldsymbol{\eta} = \mathbf{0}$. Hence if $\mathbf{z}$ is an arbitrary affine transformation of $\boldsymbol{\mu}$ in $\mathbb{R}^2$, (i.e., $(\mathrm{Re}\ \mathbf{z}, \mathrm{Im}\ \mathbf{z}) = \mathbf{1}\mathbf{c}^T + (\boldsymbol{\nu}, \boldsymbol{\eta})\mathbf{E}$ for some 2×1 vector $\mathbf{c}$ and 2×2 nonsingular real-valued matrix $\mathbf{E}$, regarded as an $(n \times 2)$ matrix equation) then $d(\mathbf{z}, \boldsymbol{\mu}) = 0$. Hence the distance focuses attention on those aspects of $\mathbf{z}$ which are not an affine transformation of $\boldsymbol{\mu}$. In some ways it is best to think of the template not as a point but as a three-dimensional subspace of $\mathbb{R}^n$ spanned by $\mathbf{1}, \boldsymbol{\nu}$ and $\boldsymbol{\eta}$. In practice it is also often useful to introduce a second distance within this subspace (incommensurate with (1)) to measure the affine difference between $\mathbf{z}$ and $\boldsymbol{\mu}$ ([8, Sect. 6.3]).

2.4 Grey-level Objects

Next consider situations in which "texture" is also important. Thus we might define an "object" to be a grey-level image on the unit square $[0,1]^2$.

Amit *et al.* [1] developed a framework to deal with such objects. Starting with an underlying template image on $[0,1]^2$, they consider a parametric family of images obtained by deforming the unit square $[0,1]^2$ in such a way as to hold the edges fixed. They give their deformations a Fourier series representation and define a distance between two images in terms of the Fourier coefficients.

This framework is similar to Sect. 2.3 since deformations of regions in $\mathbb{R}^2$ are involved. However, no landmarks are needed or used in the approach of [1]. Further, the equivalence group here is trivial so it is necessary to carefully preregister a test image on top of a template image before attempting any deformation.

3 Registration

Registration involves transforming objects so that they can be superimposed on one another in Euclidean space as closely as possible. When comparing shapes, registration involves the choice of an appropriate element from the equivalence class of objects. Most methods of registration depend on first identifying landmarks in an object.

It is useful to distinguish two basic types of registration. *Global* registration involves a single transformation applied to the whole object. Usually a linear transformation is used, chosen from the equivalence class for a shape. On the other hand, in *local* registration a nonlinear transformation is used, chosen to get a closer local match between the landmarks of the two objects.

3.1 Global Methods of Registration

Global registration methods can be classified as *absolute* if the objects are registered with respect to a fixed frame of reference, or *relative* if objects are only registered with respect to one another. One method of absolutely registering landmark-based objects in two dimensions is to apply a similarity transformation to send one landmark L_1, say, to $(0,0)$ and another landmark L_2 to $(1,0)$. This registration procedure is due to Bookstein [6]. However, some caution is required in selecting which landmarks to label L_1 and L_2. For example, if L_1 and L_2 are too near or are highly variable, then distortion in the shape representation is likely. In three dimensions one way to carry out the analogous registration is to apply a similarity transformation to do the following: (i) send L_1 to $(0,0,0)$, (ii) send L_2 to $(1,0,0)$, and (iii) send L_3 to the $x-y$ plane with positive y component [16]. Bookstein's method of registration is popular because it is easy to calculate and straightforward to display.

Another method of (relative) registration is Procrustes analysis [15]. Here a landmark object is transformed relative to a template to minimize the sum of the squared Euclidean distances between landmarks. A closely related approach is due to Kendall [19] and involves partial registration. Landmark vectors $\mathbf{z}$ are centered and scaled ($\sum z_i = 0$, $\sum |z_i|^2 = 1$) so that they lie on the unit complex sphere in $\mathbb{C}^n$. However, the registration is not complete because the rotated vector $e^{i\theta}\mathbf{z}$ and $\mathbf{z}$ represent the same shape. For most purposes statistical analyses based on the Procrustes, Kendall, and Bookstein methods yield similar interpretations.

An exception to this statement is Principal Component Analysis (PCA). As is well-known in multivariate analysis [23], the underlying metric can have a significant effect on the interpretation of PCA; for example, PCA based on a

covariance matrix is different from PCA on a correlation matrix. However, this property has not always been widely appreciated in the image analysis literature. Let $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$ be a set of m centered and scaled landmark vectors, concentrated about another landmark vector $\boldsymbol{\mu}$, for example, the dominant eigenvector of the complex covariance matrix of $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$. Without loss of generality suppose the $\mathbf{z}^{(j)}$ have been rotated so that $\mathbf{z}^{(j)*}\boldsymbol{\mu} > 0$, and define *Kendall tangent space coordinates* by $\mathbf{v}^{(j)} = (\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^*)\mathbf{z}^{(j)}$, the projection of $\mathbf{z}^{(j)}$ onto the tangent space of the unit sphere at $\boldsymbol{\mu}$. The vectors $\mathbf{v}^{(j)}$ in this tangent space satisfy complex constraints $\mathbf{v}^{(j)*}\mathbf{1} = 0$ and $\mathbf{v}^{(j)*}\boldsymbol{\mu} = 0$; hence the tangent space can be regarded as a real subspace of $\mathbb{R}^{2n}$ of dimension $2n-4$. A version of PCA based on Kendall tangent space coordinates can be defined by carrying out a conventional PCA on the vectors $\mathbf{v}^{(j)}$, regarded as $2n$-dimensional real vectors in a $(2n-4)$-dimensional subspace of $\mathbb{R}^{2n}$.

Cootes *et al.* [11] have essentially used Kendall tangent space coordinates to compare images of resistors on which landmarks could be identified. They used "derived" landmarks (consisting of averages of actual landmarks at each end of the resistor) to carry out the registration.

On the other hand, a conventional PCA can also be carried out in Bookstein coordinates using the $n-2$ nonfixed landmarks, each with 2 components. The mapping between Kendall coordinates and Bookstein coordinates is approximately linear, but is not orthogonal; hence the two approaches to PCA can yield different results. Kent [21] gives further details. Limitations of PCA in Bookstein coordinates are pointed out in Bookstein ([8, p.340]). It should be noted that a PCA in Bookstein coordinates can also be sensitive to the pair of landmarks which are chosen for the registration.

Note that Bookstein registration ([8, Sect. 5.3.2]) can give rise to "spurious" correlations. Let $\boldsymbol{\mu} = (-\frac{1}{2}, \frac{1}{2}, \mu_3, \mu_4)$ be a fixed vector and construct a random vector $\mathbf{z}$ by adding independent complex random variables with small variance τ^2 to each of the four landmarks. Let $\mathbf{w}$ denote the representation of $\mathbf{z}$ in Bookstein coordinates. Then it can be shown that

$$\operatorname{cov}(w_3, w_4) = (1 + 4\mu_3\bar{\mu}_4)\tau^2$$

([8, 24]). When one is examining shape variability, there is always an effect on perspective depending on the pair of landmarks chosen for registration.

3.2 Preregistration

When landmarks cannot be identified in an image then it is useful to roughly match the template to the image by other means — Grenander *et al.* [18] and Mardia *et al.* [25] identify hands in an image by modified thresholding and use principal components to superimpose the template on the image before fitting a deformed template by more detailed procedures.

3.3 Local Registration

In face analysis the concept of shape arises at two levels ([2, 9, 10, 12, 28]). First landmarks (sometimes called "control points") are identified in a two-

dimensional image of a human face. The relative positions of these landmarks form one definition of shape. Next the grey-level texture of the face itself forms a second notion of shape. Further, to compare two grey-level shapes it is important to match up the control points first, using a nonlinear map (deformation) from $\mathbb{R}^2$ to $\mathbb{R}^2$. The resulting grey-level images can then be directly compared.

More specifically, let $\mu_i, i = 1, \ldots, n$ and $z_i, i = 1, \ldots, n$ denote the landmarks on a template and test face respectively, and let $f(t), g(t)$, $t \in \mathbb{R}^2$ be the two grey-level images on the unit square; typically f and g take integer values 0 to 255. First we construct a deformation $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ with the following properties:

1. Φ is continuous and bijective,
2. $\Phi(\mu_i) = z_i$, $i = 1, \ldots, n$,
3. Φ is equivariant under linear changes to the data.

Then $g(\Phi(t))$ represents the second image, deformed to the landmark spacing of the first image, and as such, can be directly compared to $f(t)$.

Two simple ways to construct the deformation from a template image to a test image are as follows:

1. Form the Delaunay triangulation of landmarks for the template face, and construct the piecewise linear map which takes the landmarks of the template to the landmarks of the test image and which is linear on each Delaunay triangle. If $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are the vertices of a Delaunay triangle, then the equation
$$\mathbf{y}_i = \mathbf{A}x_i + \mathbf{b} \ , \qquad i = 1, 2, 3,$$
gives 6 equations for the 6 quantities in $\mathbf{A}(2\times 2)$ and $\mathbf{b}(2\times 1)$ determining the linear mapping. It is also possible to triangulate $\mathbb{R}^2$ outside the convex hull of the landmarks with each triangle having one or two vertices at infinity, but a bit more care is needed to guarantee a well-specified piecewise linear mapping.
2. The thin-plate spline mapping from the template to the test image.

Some details of the thin-plate spline mapping can be found in [7, 25, 29]. The advantage of the thin-plate spline mapping is that it minimizes bending energy.

If the $\{z_i\}$ differ only slightly from a linear transformation of $\{\mu_i\}$ then both these deformations will be bijective, the usual situation in practice. However, if a gross nonlinearity is involved, then the deformation may fold over on itself, leading to a lack of bijectivity.

3.4 Shape Analysis without Registration

It is also possible to compare objects without first registering them, by focusing attention on the interlandmark distances. (See [17, 22].) However Bookstein ([8, pp. 225–227]) points out some limitations in this point of view.

4 Image Analysis and the Bayesian Paradigm

In this section we describe how Bayes theorem together with appropriate statistical models can be used to locate objects in images. The popularity of this approach is largely due to the efforts of Grenander and his co-workers over the years.

Consider a grey-level image represented as a real-valued function $f(t), t \in \mathbb{R}^2$. In practice this is observed on a rectangular grid of pixels indexed by $t = \ell$, $\ell \in D$, say, where

$$D = \{\ell = (\ell_1, \ell_2) : 1 \leq \ell_1 \leq L_1, 1 \leq \ell_2 \leq L_2\} \ .$$

Suppose an image contains an object S which is assumed to be a deformation of some underlying template object S_0, and suppose the image is also subject to observational noise. One possible model is

$$y_\ell = \nu_1 + \epsilon_\ell \quad \text{if} \quad \ell \in S \ , \qquad y_\ell = \nu_2 + \epsilon_\ell \quad \text{if} \quad \ell \notin S \ ,$$

where $y_\ell \in \mathbb{R}$ denotes the observed "grey-level" in the ℓ^{th} pixel and ℓ labels pixels in an $L_1 \times L_2$ grid. In the simplest version of the model we suppose the noise variables ϵ_ℓ are independent $N(0, \sigma_\epsilon^2)$ random variables. The mean levels ν_1 and ν_2 indicate the difference between the shape and the background. Thus given S the model for $\mathbf{y} = \{y_\ell\}$ has probability density function

$$P(\mathbf{y} \mid S) \propto \exp\{-\frac{1}{2\sigma_\epsilon^2}[\sum_{\ell \in S}(y_\ell - \nu_1)^2 + \sum_{\ell \notin S}(y_\ell - \nu_2)^2]\} \ .$$

More realistic models might include an allowance for blurring [4, 14].

Next let $P(S)$ denote the prior probability density of S under a deformable template model, for example, the conditional autoregressive model in Sect. 2.2. Then by Bayes theorem $P(\mathbf{y} \mid S)P(S)$ is proportional to the posterior density of S given the data. We seek an estimate of S to maximize the posterior density. This estimate is known as the MAP or *maximum a posteriori* estimate. If we let $d(S, S_0)$ be a distance between S and S_0, where S is an arbitrary deformation of S_0, a natural prior for S is given by

$$P(S) \propto \exp\{-\frac{1}{2\sigma^2} d^2(S_0, S)\} \ .$$

Preregistration (see Sect. 3) can be carried out to obtain an initial fit of the object to the image if needed. In particular, the search for objects in images based on the models described in Sects 2.2 - 2.4 can be cast in this framework. Popular methods for the computation include Iterated Conditional Modes and simulated annealing, [3, 5, 13, 18].

Other statistical contributions to shape analysis in a wider context include the work of Ripley and Sutherland [27] who attempted to identify the arms of a spiral galaxy, and Mardia *et al.* [26] who provide a gesture recognition method.

References

1. Amit, Y., Grenander, U. and Piccioni, M. (1991). Structural image restorations through deformable templates, J. Am. Statist. Assoc., 86, pp. 376–387.
2. Benson, P.J. and Perrett, D.I. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural image, European J. Cognitive Psychology, 3, pp. 105–135.
3. Besag, J.E. (1986). On the statistical analysis of dirty pictures (with discussion), J. R. Statist. Soc. B, 36, pp. 192–236.
4. Besag, J.E. (1989). Towards Bayesian image analysis, J. Appl. Statist., 16, pp. 395–407.
5. Besag, J.E. and Green, P.J. (1993). Spatial statistics and Bayesian computation, J. R. Statist. Soc. B, 55, pp. 25–37.
6. Bookstein, F.L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion), Statist. Science, 1, pp. 181–242.
7. Bookstein, F.L. (1989). Principal warps; thin-plate splines and the decomposition of deformations, IEEE Trans. Pattern Anal. Machine Intell., PAMI-11, pp. 567–585.
8. Bookstein, F.L. (1991). Morphometric Tools for Landmark Data. Cambridge Univ. Press.
9. Bruce, V. (1988). Recognising Faces. London, Lawrence Erlbaum Associates.
10. Coombes, A.M., Moss, J.P., Linney, A.D., Richards, R. and James, D.R. (1991). A mathematical method for the comparison of three dimensional changes in the facial surface, European J. of Orthodontics, 13, pp. 95–110.
11. Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J. (1992). Training models of shapes from sets of examples, Proc. British Machine Vision Conference, Leeds 1992, pp. 9–18.
12. Craw, I. and Cameron, P. (1992). Face recognition by computer, Proc. British Machine Vision Conference, Leeds 1992, pp. 498–507.
13. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs' distributions and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intell., PAMI-6, pp. 721–741.
14. Geman, D. and Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities, IEEE Trans. Pattern Anal. Machine Intell., PAMI-14, pp. 367–382.
15. Goodall, C.R. (1991). Procrustes methods in the statistical analysis of shape (with discussion), J.R. Statist. Soc. B, 53, pp. 285–339.
16. Goodall, C.R. and Mardia, K.V. (1991). Multivariate aspects of shape theory with applications, Ann. Statist. to appear.
17. Gower, J. (1991). Discussion to a paper by C. Goodall, J. R. Statist. Soc. B, 53, pp. 326–327.
18. Grenander, U., Keenan, D.M. and Chow, Y. (1991). Hands: A pattern theoretic study of biological shapes. Berlin, Springer Verlag.
19. Kendall, D.G. (1984). Shape manifolds, procrustean metrics and complex projective spaces, Bull. London Math. Soc., 16, pp. 81–121.
20. Kent, J.T. (1992a). New directions in shape analysis. In: Mardia, K.V. (ed.), The Art of Statistical Science, Chichester, Wiley, pp. 115–127.
21. Kent, J.T. (1992b). The complex Bingham distribution and shape analysis, J. R. Statist. Soc. B, to appear.
22. Lele, S. (1991). Discussion to a paper by C. Goodall, J. R. Statist. Soc. B, 53, p 334.

23. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate Analysis. Academic Press, London.
24. Mardia, K.V. and Dryden, I.L. (1989). The statistical analysis of shape data, Biometrika, 76, pp. 271–281.
25. Mardia, K.V., Kent, J.T. and Walder, A.N. (1991). Statistical shape models in image analysis, Proc. Interface 1991, Seattle, pp. 550–557.
26. Mardia, K.V., Ghali, N.M., Howes, M., Hainsworth, T.J. and Sheehy, N. (1993). Techniques for on-line gesture recognition on workstations, Image and Vision Computing J., 11. (5), pp. 283–294.
27. Ripley, B.D. and Sutherland, A.I. (1990). Finding spiral structures in images of galaxies, Phil. Trans. Roy. Soc. A, 332, pp. 477–485.
28. Turk, M. and Pentland, A. (1991). Eigenfaces for recognition, J. Cognitive Neuroscience, 3, pp. 71–86.
29. Wahba, G. (1990). Spline models for observational data, SIAM, Philadelphia.

Recognition of Shapes from a Finite Series of Plane Figures

Nikolai Metodiev Sirakov

Institute of Mechanics and Biomechanics, Bulgarian Academy of Sciences
Akad. G.Bontchev.Bl 4, 1113 Sofia, Bulgaria
Email: imbm@bgearn.bitnet

Abstract. An object is defined as a rigid 3-D body that is bounded by surfaces. Surfaces of up to the second order will be considered. The models of the objects are constructed as a finite series of plane figures. Recognition is determined by the order and set of identification. Regularity and consistency are used for classification of these figures. An algorithm for the recognition of a single object or of several objects, situated one after the other, has been developed. The algorithm allows recognition of overlapping objects.

Keywords: shape description, plane figures, regularity, consistency.

1 Introduction

Recognition of 3-D objects is an important task in various applications. The problem can be solved by transformation from 3D to 2D.

In this article it will be assumed that the 3-D coordinates are obtained from the visible part of the objects by means of a stereo sensor [4]. Consider a fixed Cartesian coordinate system $O'x'y'z'$. The first octant is called the *operation environment*, and the plane $O'x'y'$ is be called the *ground plane*. With respect to this fixed coordinate system, a mobile Cartesian system $Oxyz$ is defined. The point O of this system is identified with the left camera of the stereo sensor.

In the initial position, the axes of the two coordinate systems are parallel. The mobile system $Oxyz$ may move by in-plane translations and rotations on $O'x'y'$, and rotate with respect to $O'z'$ and $O'x'$ (Fig. 1). The relation between the two coordinate systems is known, therefore measurements can be carried out in $Oxyz$.

The space encompassed by the stereo sensor is divided into a rectangular grid. In this grid, consider k finite planes which are parallel to the Oxz plane. The objects are motionless and situated in the operation environment on the ground plane of $O'x'y'z'$ and $Oxyz$ moves toward this octant. At the moment of recognition, $Oxyz$ is motionless, and only starts moving after a decision is taken.

An object is defined as a rigid 3-D body that is bounded by surfaces. Surfaces of up to the second order will be considered. The objects, which will be recognized are: parallelepipeds, cylinders, spheres, ellipsoids, hyperboloids, and paraboloids.

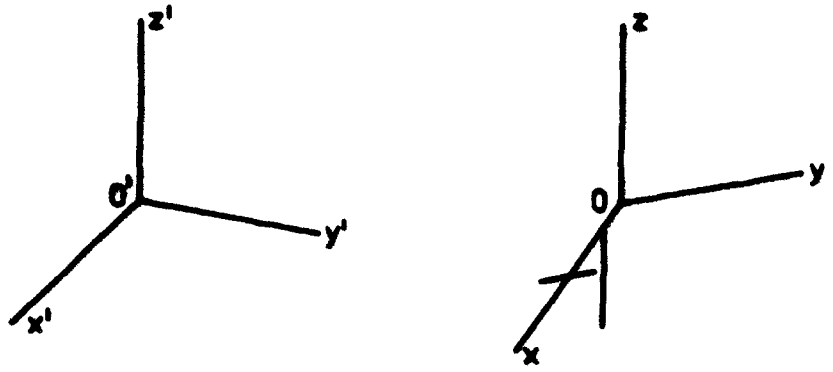


Fig. 1. The coordinate frames

2 Creation of the Object Models

Consider a set of 3-D objects $M = \{O_1, O_2, \ldots, O_n\}$ with surfaces of up to the second order. For every object $O_p \in M$ a transformation $O_p \rightarrow R_q$ can be defined, where R_q is a finite series of plane figures. Only the pairs, of which the object belongs to the space enfolded by the series of finite planes, will be taken into account, indicated by $T_k^s : O_k \rightarrow R_s$.

Definition 1. If for an object O_k there exists m different sets $T_k^{s_i} : O_k \rightarrow R_{s_i}$ for $s_i = 1, \ldots, m$, and $\bigcup_{i=1}^{m} T_k^{s_i} = T_k$, $R_{s_i} \neq R_{s_j}$ for $s_i \neq s_j$, the object has an *order of identification* $r_k = m$ and the sets $T_k^{s_i}$ are called the *sets of identification.*

Thus, every object O_k can be identified by the triplet $\{r_k, T_k, R_k\}$, where $R_k = \{R_{s_1}, \ldots, R_{s_m}\}$.

Definition 2. Two objects O_k and O_l are *equivalent* with respect to the series of plane figures R_s, $O_i \overset{R_s}{\approx} O_j$ if $T_k^s : O_k \rightarrow R_s$ and $T_l^s : O_l \rightarrow R_s$.

The equivalence relation satisfies the reflexivity, transitivity, and symmetry properties.

Definition 3. Two objects O_k and O_l are *identical* $O_k \cong O_l$ if and only if $r_k = r_l$ and $O_k \overset{R_{s_i}}{\approx} O_l$ for $s_i = 1, \ldots, r_k$.

Definition 4. Two objects are *different*, if and only if they are not identical according to $\cong$.

From Definitions 3 and 4 follow some corollaries.

Corollary 5. *Two objects O_k and O_l are different with respect to $\cong$, if $r_k \neq r_l$.*

Corollary 6. *Two objects O_k and O_l are different with respect to $\cong$, if $r_k = r_l$ but $|R_k \cap R_l| < r_k$.*

Corollary 7. *Two objects O_k and O_l are different with respect to $\cong$ and object O_k is part of object O_l, if $r_k < r_l$ but $|R_k \cap R_l| = r_k$.*

If an object has an order of identification $r = 1$, it is symmetrical. Such an example is the sphere.

The object O_k may be in different positions, such that different surfaces are on the ground plane, represented by $O_{k1}, \ldots, O_{kl}$, so that $r_{ki} \neq r_{kj}$, or if $r_{ki} = r_{kj}$, then $|R^{ki} \cap R^{kj}| < r_{ki}$ for $i \neq j$. In these expressions, r_{kj}, for $j = 1, \ldots, l$, is the order of identification, and R^{kj} are the sets of sequences corresponding to the sets of identification. It follows that every object O_k can be represented by the sequence $\{a_{kj}\} = a_{k1}, \ldots, a_{kl}$ denoting the positions of the object.

From Corollaries 5 and 6 it can be seen that $O_{ki} \neq O_{kj}$ for $i \neq j$, therefore the set M can be extended to $\{M\} = \{O_{11}, \ldots, O_{1k_1}, \ldots, O_{nk_n}\}$.

To a specific position of an object O_{kj} corresponds the finite sequence of plane figures $a_{kj} = R^{kj}_{s_1}, \ldots, R^{kj}_{s_m}$, where the index $m = r_{kj}$, R^{kj}_{si} are finite sequences of plane figures, and $R^{kj}_{s_i} \neq R^{kj}_{s_{i+1}}$, for $i = 1, \ldots, m$. The finite sequence of plane figures show the orientation of the object with respect to the sensor. Each pair of neighbouring elements corresponds to a pair of neighbouring sets of identification. The neighbourhood is defined as anti-clockwise.

From the above, it can be seen that every object $O_k \in M$ can be represented by a sequence $\{a_{kj}\} = a_{k1}, \ldots, a_{kl}$. If an index i can be associated to every a_{kj}, then the sequence $\{a_{kj}\}$ is a finite numerical sequence. Analogically, if $R^{kj}_{s_m} \rightarrow s_m$, every sequence a_{kl} is a finite numerical sequence.

Hence, an object may be in different positions on the ground plane, each associated with a series of plane figures. The models of the objects can be constructed by means of a series of plane figures $R^{kj}_{s_i}$, where the index k denotes a particular object, the index j denotes the different positions of the object on the ground plane, and the index s_i denotes the orientation with respect to the cameras. As a result, the models of 3-D objects can be represented by a tree, as shown in Fig. 2.

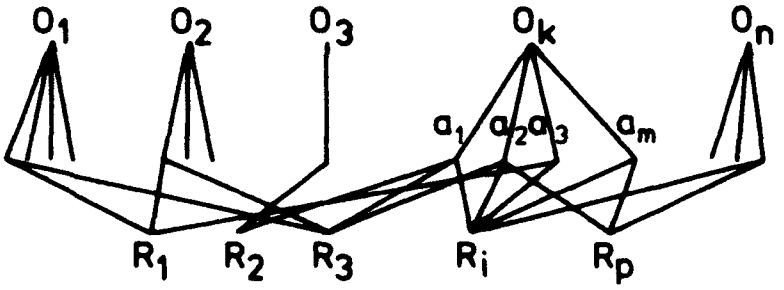


Fig. 2. Models of 3-D objects

3 Structure of the Series of Plane Figures

Since M is finite, the set of all series of plane figures is also finite. Since the R are finite, the set of plane figures $F = \{F_1, F_2, \ldots, F_n\}$ is finite.

If an index is juxtaposed to every plane figure of F, then a finite numerical series will correspond to every series R. The general element of the series is of the form

$$R = \{f_j, E_j, k_j\} \ , \tag{1}$$

where j is the index indicating the place of the element in the series, f_j is a number corresponding to a certain figure, k_j is a repetition factor showing how many times a figure f_j appears consecutively in the series, E_j is a subset of the set $E = \{A_w, B_w, \ldots\}$. Specifically, $A, B, \ldots$ are geometrical elements of the figure f, and $w \in \{\uparrow, \downarrow, \cong\}$. The signs denote the change in size of a geometrical element with respect to identical elements in the previous figure: $\uparrow$ denotes an increase, $\downarrow$ denotes a decrease, and $\cong$ denotes no change.

Hence, the task of recognition of 3-D objects is reduced to the recognition of 2-D plane figures by tracing a finite numerical series.

4 Regularities and Consistency of Regularities for Recognition of Plane Figures

Grenander [2] gives excellent and elegant theoretical concepts in the field of *regular structures*, but they are not efficient enough when applied in reality. In the following, methods that can be applied for real-time systems are proposed.

Let $D = \{D_1, \ldots, D_k\}$ be a certain set of actions, and $S = \{S_1, \ldots, S_l\}$ be a set of conditions.

Definition 8. S_i is *regular* with respect to D_j if when the action D_j is fulfilled, the condition S_i is satisfied. Regularity will be indicated by (S_i, D_j).

Let the conditions be satisfied and the actions be fulfilled in the same space. Let x be a point in this space.

Definition 9. The regularity (S_i, D_j) is said to have changed at a point x, if (S_i, D_j) is satisfied at point $x - dx$ and $(S_k, D_j) \ncong (S_i, D_j)$ at point x.

Assume that N is a set of points in the plane.

Definition 10. N is regular with respect to (S_i, D_j), if every time D_j is fulfilled, the points of N satisfy condition S_i .

The starting point x^b and end point x^e of regularity can be associated with the term *regularity*. Therefore, every regularity can be described in an interval $\Delta L = x^e - x^b$. Consider the finite part of the Oxz plane. Choose the movements on straight lines parallel to the Oz axis (i.e. $D = \{D_l\}$, where D_l is $x = l$ for $l = 0, 1, \ldots, n$) from all possible actions on the plane.

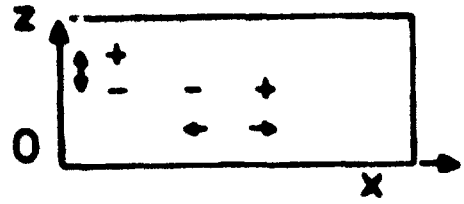


Fig. 3. Movements of the line in the plane

Let the line move in the plane parallel to Oz. The elements of S with respect to the action D_l are defined as follows:

- S_0 in the interval ΔL there are no points on the line;
- S_1 for $x = l$, where l is constant, there exists a set of points on the line;
- S_2 in the interval ΔL a set of points remains fixed on the line;
- S_3 in the interval ΔL a set of points moves in the positive direction on the line;
- S_4 in the interval ΔL a set of points moves in the negative direction on the line;
- S_5 in the interval ΔL a set of points moves in negative and positive directions on the line. The distance between the points increases;
- S_6 in the interval ΔL a set of points moves in positive and negative directions on the line. The distance between the points decreases;
- S_7 in the interval ΔL there exists a single point on the line.

Thus, the regularities with respect to $(S_0, D_l), \ldots, (S_7, D_l)$ to D_l are obtained. The curves which satisfy the regularities are shown in Fig. 4. Because there is only one action D_l, a shorter notation for regularities can be used, S_i.

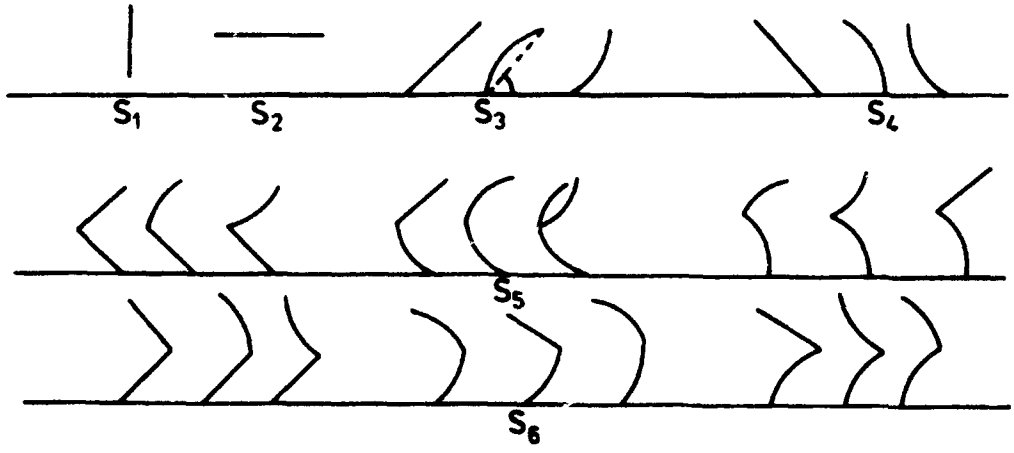


Fig. 4. Regularity curves

From the previous definitions, it may be concluded that the plane figures in F can satisfy several regularities.

Definition 11. The regularities S_i and S_j are *consistent* with respect to F, if there exists a figure $F_p \in F$, such that two subsets of points belonging to F_p can be found, satisfying the regularities S_i, S_j and the inequality: $|x_j^b - x_i^e| < d$. Consistency shall be denoted by the sign $\longrightarrow$.

If n regularities S_i are observed in the interval ΔL, their consistency shall be denoted by S_i^n. If several different regularities are observed, they can be connected in a positive direction with respect to the Oz axis. Therefore, the plane figures in F can be defined as consistencies of regularities.

From the last formulation, it also follows that a single figure (in its different orientations with respect to the coordinate system) can be represented by different consistencies. On the other hand, there may exist figures in F which are represented by a single consistency (see Fig. 5).

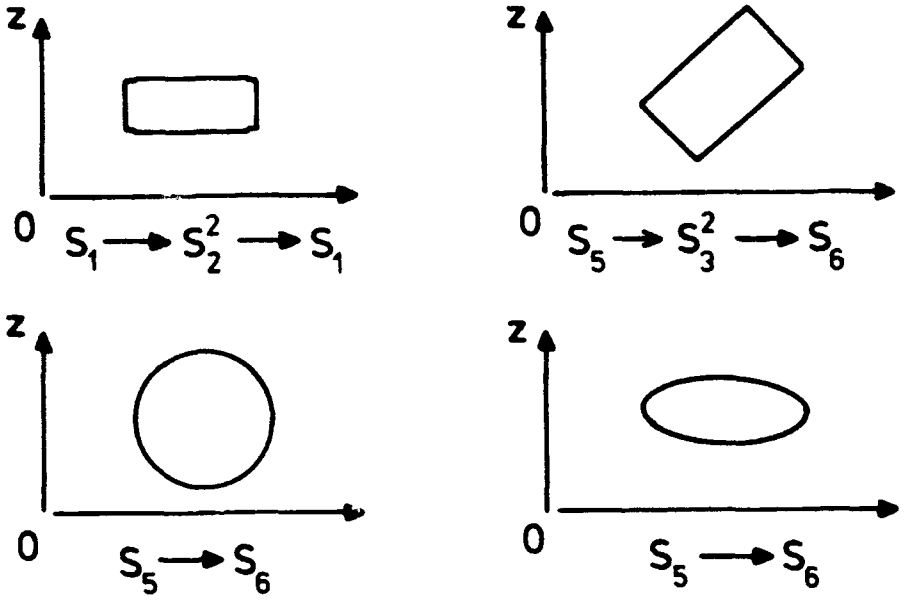


Fig. 5. Plane figures with consistencies

Definition 12. Two figures are *similar*, if the points of these figures satisfy exactly the same consistencies.

Let $F^p = \{F_1^p, \ldots, F_k^p\}$ indicate all the orientations of a figure F_p . Then, using Definitions 11 and 12 the set F can be extended to $F^0 = \{F^1, \ldots, F^n\}$. If $C = \{C_1, \ldots, C_l\}$ denotes the set of all consistencies, satisfied by the figures in F, then $|C| \leq |F^0|$. Hence, it follows that $F^0 = \bigcup_{i=1}^{l} m_i$, where m_i are the set of figures whose points satisfy the consistency C_i, and $m_i \cap m_j = 0$ for $i \neq j$.

5 Independent Observer

Assume that the plane is divided by a rectangular grid. The shape of the curves on the vertices of the grid are presented in [6]. Suppose that there is an independent observer (IO) who is tracking the movement of a curve point in the plane on the line $x = l$ when this line is moving in the plane. The IO may stop the moving line if one of the following situations occur:

1. One or more nonzero regularities are changed.
2. d columns of the grid are zero.
3. The end of the searching plane is found.

The IO can make the following observations:

1. Direction of the movement of the point on the line: positive, negative.
2. The shape of the segments induced by the movement of the line: vertical, horizontal.
3. The change of length of the segments: decrement, increment, no change.

As a result, the IO can determine the following criteria and conditions:

1. The type of regularities with respect to the direction of movement of the point on the line: $\uparrow$ for S_3, $\downarrow$ for S_4, $\downarrow\uparrow$ for S_5, $\uparrow\downarrow$ for S_6, $\cong$ for S_1, $\emptyset$ for S_2.
2. The angle between the direction of regularity and the axis Ox, using the information on the shapes of the segments and the type of regularities.
3. The shape of the arc: convex, concave, or straight line.

The following sequence can be defined:

$$\mathcal{S} = (S_{i1})^{k1} \longrightarrow (S_{i2})^{k2} \longrightarrow \ldots \longrightarrow (S_{in})^{kn} \ , \tag{2}$$

where the general element is the regularity $S_{in} = \{\mu_{in}, \nu_{in}, \ell_{in}\}$;

$\mu_{in} \in \{\uparrow, \downarrow, \downarrow\uparrow, \uparrow\downarrow, \cong, \emptyset\}$;

$\nu_{in} \in \{V, H, VV, HH, VH, HV\}$, V = vertical, H = horizontal ;

$\ell_{in} \in \{\ell_{\uparrow}, \ell_{\downarrow}, \ell_{\uparrow\uparrow}, \ell_{\uparrow\downarrow}, \ell_{\downarrow\uparrow}, \ell_{\downarrow\downarrow}\}$.

ℓ denotes the length of the vertical or horizontal segment of the regularity.

Each of the elements of the sequence, corresponding to all points, satisfies one and only one regularity. Hence, under the above assumptions, the boundaries of the figures can be represented using sequence (2), and its consistency of regularities. The IO can recognize several regularities simultaneously. The program which realizes this feature is presented in [5].

From this definition it follows that the IO determines sets of similar figures. If a set includes only one figure, then the IO concludes that the figure is recognized.

6 Method of the Sum of the Quadratic Error

The Method of the Sum of the Quadratic Error (MSQE) [8] is applied to determine a curve which gives the best approximation of a set of points if the set of similar figures includes more than one figure. Suppose that there is a set of points $N = \{(x_i, z_i); i = 1, \ldots, n\}$ on the Oxz plane, and the algorithm has recognized a set of similar plane figures. These points may satisfy several approximation criteria:

1. Circle with centre (x_0, z_0) and radius r, given by the inequality

$$E_1 = \sum_{i=1}^{n} \left|(x_i - x_0)^2 + (z_i - z_0)^2 - r^2\right| < n\, d^0\, (d^0 + 2r) = \delta_1 \ ,$$

where $x_0 = \dfrac{\sum_{i-1}^{n} x_i}{n}$, $z_0 = \dfrac{\sum_{i-1}^{n} z_i}{n}$, $r^2 = \dfrac{\mu_{20} - \mu_{02}}{n}$,

μ_{20}, μ_{02} are the second central moments of the points of N,

d_i denotes the distance between the point (x_i, z_i) and the circle,

and d_0 is the least distance greater than d_i, for $i = 1, \ldots, n$.

2. Ellipse with centre (x_0, z_0) and axes a and b, given by the inequality

$$E_2 = \sum_{i=1}^{n} \left|\frac{(x_i - x_0)^2}{\mu_{20}}\, n + \frac{(z_i - z_0)^2}{\mu_{02}}\, n - 1\right| < \frac{d_0^2}{2\mu_{02}}\, n^2 + 2\frac{d_0}{(2\mu_{02})^{\frac{1}{2}}}\, n^{\frac{3}{2}} = \delta_2,$$

where $a^2 = \dfrac{\mu_{20}}{n}$, $b^2 = \dfrac{\mu_{02}}{n}$.

3. Hyperbola with imaginary axis Oz or Ox, respectively given by inequalities

$$E_3 = \sum_{i=1}^{n} \left|\frac{(x_i - x_0)^2}{\mu_{20}}\, 2n + \frac{(z_i - z_0)^2}{\mu_{02}}\, n - 1\right| < n\, d_0^2 + 2n\, d_0 = \delta_3 \ ,$$

$$E_3 = \sum_{i=1}^{n} \left|\frac{(z_i - z_0)^2}{\mu_{02}}\, 2n + \frac{(x_i - x_0)^2}{\mu_{20}}\, n - 1\right| < n\, d_0^2 + 2n\, d_0 = \delta_3 \ .$$

Note that the MSQE gives good results if the points of N are normally distributed around the centre of the figure. For an arc of a plane figure of second order the MSQE is not recommended.

7 Recognition of Overlapping Objects

If objects which are in the scope of the sensor overlap, then some of the plane figures may overlap too, and parts of the boundaries of the figures are not visible. An approach for the determination of an arc of a curve of second order, which approximates the set of plane point, is presented in [7].

From the definitions of regularities and consistencies, it follows that the points of an arc of a curve of second order satisfy the regularities S_5 and S_6, or consistencies $C_3 = S_3 \longrightarrow S_4$ and $C_4 = S_4 \longrightarrow S_3$. Assume that the set of plane points

satisfies the noted regularities or consistencies. Consider the general equation of a curve of second order:

$$A\,x + 2B\,xz + C\,z + 2D\,x + 2E\,z + F = 0\ . \tag{3}$$

Theorem 13. *If N is a set of plane points, which satisfies one of the regularities S_5 or S_6, then $C \neq 0$.*

Theorem 14. *If N is a set of plane points, which satisfies one of the consistencies C_3 or C_4, then $A \neq 0$.*

From the theorems the following systems of equations can be constructed:

$$\frac{A}{C}\,x_i^2 + \frac{D}{C}\,x_i + \frac{E}{C}\,z_i + \frac{F}{C} = z_i^2\ , \qquad i = 1,2,3,4\ , \tag{4}$$

$$\frac{C}{A}\,z_i^2 + \frac{D}{A}\,x_i + \frac{E}{A}\,z_i + \frac{F}{A} = x_i^2\ , \qquad i = 1,2,3,4\ . \tag{5}$$

If the coordinates of four points are known from the sensor system, the coefficients of (4) or (5) can be calculated.

There are some important aspects of the behaviour of the different types of curves with regard to the regularities S_5 or S_6, and the consistencies C_3 or C_4.

The following are elliptic curves: an ellipse, an imaginary ellipse, a pair of complex conjugate lines. Because the plane is restricted to the real plane, imaginary curves cannot satisfy any of the regularities or consistencies. There are two hyperbolic curves: a hyperbola, a pair of real crossing lines. Obviously, the pair of real crossing lines does not satisfy S_5, S_6, C_3, or C_4. There are three parabolic curves: a parabola, a pair of real parallel curves, a pair of complex conjugate parallel curves. The latter has no real points.

From the definition of regularities and consistencies it follows that the points of a pair of real parallel curves satisfy one of the following regularities [5, 8]:

S_2^2 if the pair is parallel to the axis Ox;
S_1^2 if the pair is parallel to the axis Oz;
S_3^2 if the pair makes a sharp angle with the axis Ox;
S_4^2 if the pair makes a wide angle with the axis Ox.

The following criteria can be formulated:

1. If $A \cdot C > 0$, the set of plane points is approximated by an arc of an ellipse.
2. If $A = C$, the set of points is approximated by an arc of a circle.
3. If $A \cdot C < 0$, the set of plane points is approximated by an arc of a hyperbola.
4. If $C = 0$ or $A = 0$, the set of plane points is approximated by an arc of a parabola.

The system of equations (4) or (5) is solved using a specific choice of points. Three points are boundary points of regularities S_5 or S_6, and of consistencies C_3 or C_4. The fourth point is the middle point of one of the parts of the regularity or the consistency. If a middle point is not recorded by the sensor, the nearest point can be used.

8 Conclusion

Using the previously described concepts, an algorithm for the recognition of a single object or of several objects, situated one after the other, has been developed. The construction of the object models allows the use of parallel processes for checking finite numerical series [1].

The program is implemented in a language of the FORTH family. It can recognize an object in 200 ms, using an INTEL 80286 processor [5]. Three objects, one after the other, are recognized in 500 ms.

The second step was the construction of a program for the recognition of a 3-D object from multiple views [9]. The third step of this research is the recognition of several objects, which are in the scope of the sensor, using a single view [8]. Current research is on the recognition of complex scenes from multiple views, and the recognition of moving objects.

The system can be applied to robot control, navigation of vehicles [6], and detection of 3-D objects in the chemical and nuclear industries.

References

1. Cupic R., Sirakov N., Trebaticky I. (1991). Processing of the 3d information with aid of a systolic system, Proc. Int. Conf. Cybernetics and Informatics, Smolenize, Slovakia, pp. 43–48 (in Slovakian language).
2. Grenander U. (1981). Regular Structures-Lectures in Pattern Theory, Vol.3, Springer-Verlag, New York.
3. Sirakov N. (1988). Application of regularities in pattern classification, Proc. Int. Conf. Mech. Appl. in Field of Robotics and New Materials, Sunny Beach, Bulgaria, pp. 309–315.
4. Sirakov N., Nedev N. (1989). Approximation analysis of spatial Scenes with 3-D coordinates data, obtained by stereo TV projection in real time, Mathematical Research, Computer Analysis of Images and Patterns, Vol.55, Akademie Verlag, Berlin, 1989, pp. 123–128. (Proc. of the III-rd Int. Conf. CAIP-89, Leipzig, Germany, Sept. 8–10.1989)
5. Sirakov N., Dimitrov A. (1989). Software application for recognition of 3-D objects, represented as finite series of plane figures, Proc. VI Natl. Conf. Theoretical and Applied Mechanics, Varna, Bulgaria, Vol.1, pp. 428–432.
6. Sirakov N., Trebeticky I. (1991). Calculation of the curvature of the edge of a road with the help of regularities, Computer Analysis of Image and Patterns, Research in Informatics, Akademie Verlag, Germany, Vol. 5, pp. 195–201.
7. Sirakov N., Ivanov N. (1992). A fast method for recognition of a 3-D overlapping objects based on the regularities and principles of the independent observer, Proc. 4th Portuguese Conf. on Patt. Rec.- RecPad'92, Coimbra, Portugal, pp. 49–57.
8. Sirakov N. (1992). Regularity and consistency for recognition of several objects from a single view, Proc. IFAC ASQP'92, Istambul, Turkey, Vol.1, pp. 337–345.
9. Sirakov N. (1990). General algorithm for recognition of an object class in R^3 from multiple views, presented on the Int. Conf. Control'90, Lugano, Swiss.

Polygonal Harmonic Shape Characterization *

Anthony J. Maeder[1], *Andrew J. Davison*[2], *and Nigel N. Clark*[3]

[1] School of Electrical and Electronic Systems Engineering, Queensland University of Technology, Brisbane QLD 4001, Australia
[2] Victorian Centre for Image Processing and Graphics, Department of Computer Science, Monash University, Clayton VIC 3168, Australia
[3] Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown WV 26506-6101, USA

Abstract. This paper describes a shape characterization technique based on the polygonal harmonics formed when the boundary of a region in a digital image is traversed using various step lengths. The traversal is derived from a traditional method for estimating fractal dimension, where steps of constant length are taken along the boundary, as if with a pair of dividers. While the traversal for fractal dimension is closed in a single winding, usually with an unequal step length, the derived traversal continues until the step endpoint lands on a previously encountered endpoint to form a stable "harmonic polygon". Parameters based on the harmonic polygons formed using different step lengths provide information on specific aspects of shape and allow comparison of shapes.

Keywords: shape, shape descriptor, region, fractal, polygon.

1 Introduction

Descriptors for shapes of 2-D regions in digital pictures are important for scene understanding as they provide a means for characterizing and quantifying the regions. This characterization leads in turn to analysis and understanding of the picture content. Shape characterization lacks a consistent, unifying approach and consequently many different descriptors have been proposed, with no generally accepted rules for determining their applicability in a given problem. Furthermore, certain application areas have developed a favoured choice of shape descriptors for no convincing theoretical reasons. Improvements in the quality and efficacy of shape characterization would be obtained by curtailing this proliferation and arbitrary selection of techniques.

Often several aspects of region shape are of interest to the observer, so typically several different shape descriptors are used to aid analysis. These different descriptors may involve very different computational processes, adding to the computational expense of the analysis. It is thus desirable to develop techniques which allow several descriptors to be computed by a single process, or at least by closely related processes. The choice of which descriptors to use can be rather

* This work was sponsored in part by the Australian Research Council and CSIRO.

ad hoc but it is often necessary to select descriptors which will characterize shape properties at coarse, intermediate, and fine scales relative to the region size. If descriptors for different scales were available through a single process, computational expense would be further reduced.

The technique presented here offers both the advantages identified above. A set of harmonic numbers is constructed through repeated traversal of the region boundary and several parameters based on various combinations of these numbers are used as separate shape descriptors. A single process therefore gives rise to these different descriptors. The construction of the set of numbers requires traversals to be made using different step lengths, so shape information relating to different scales is extracted automatically.

In this paper, the overall technique for extracting the set of harmonic numbers from traversals will first be described. The shape descriptors based on this set will then be formulated. Finally, some examples of applying the technique to biological and engineering problems will be discussed.

2 Polygonal Harmonic Traversal

The technique described here uses a method of region boundary traversal, derived from fractal perimeter estimation, to construct a set of equilateral polygons of various side lengths with all vertices located on the region boundary [1]. The conventional traversal method used in estimating fractal dimension [7] consists of stepping along the boundary of the region with a fixed step length, constructing a succession of chords of equal length with coincident endpoints, and closing the polygon so formed with a single side usually of shorter length [9]. The variation caused in perimeter estimates for nearby step lengths due to the fluctuation in length of this shorter side have been shown to perturb the results obtained for the fractal dimension considerably [2]. This perturbation is particularly marked at large step lengths, often leading to severe miscalculation of the fractal dimension.

In the derived method, stepping is continued until a previous step endpoint from the same traversal is visited a second time, as a means of controlling the perturbation. It may take several windings of the boundary to satisfy this requirement but eventual termination is guaranteed in the discrete case by the finite number of pixels along the boundary. The resulting polygon, with all sides equal to the step length, is termed *harmonic* as it forms a locally stable structure. The number of steps in the harmonic polygon determines the *order*, which is regarded as the essential shape characteristic of the region for different step lengths.

Definition 1. Let L denote a step length used for traversal of a given region, $N(L)$ denote the number of steps of length L required to form a harmonic polygon for the given region, and, assuming that the final stable harmonic polygon requires only a single winding to traverse in its closed state, the *order* of a harmonic polygon $n = N(L)$.

Each harmonic polygon constitutes an equivalent shape for the region boundary, in the sense that any shape which has a polygon of that order for the same

step length will have the same harmonic perimeter estimate for that step length. The existence of polygonal harmonics of various orders as the step length changes therefore provides basic shape information about the region. Figure 1a shows a region boundary with a polygonal harmonic of order 3; note that a harmonic of order 2 would form a straight line, as illustrated by the dashed line.

Polygonal harmonic shape analysis first requires the construction of a set of harmonic numbers over a range of different step lengths. Since the step length is notionally a continuous variable, discrete samples of step length must be chosen and used to determine corresponding polygonal harmonics computationally for digital image data. The existence and order of harmonics must be assumed to remain constant between adjacent sample step lengths which produce the same order of harmonic. There are thus ranges of step lengths over which a single order of polygonal harmonic exists or *endures*. Within these ranges, especially where a change to a different order of harmonic occurs, some transitions between different orders may be encountered, providing subranges of interrupting or *impure* behaviour. In some cases, harmonics may exist which require more than one winding of the region boundary to attain closure in the stable state. These cases are termed *complex* harmonics, whereas those which close in a single winding are termed *simple* harmonics.

Definition 2. Let $W(L)$ denote the *winding number* of the polygon. The *order* n of a harmonic polygon is equal to $\frac{N(L)}{W(L)}$.

For simple harmonic polygons, $W(L) = 1$. Figure 1b shows a polygonal harmonic plot of results for the region in Fig. 1a, which exhibits some of the properties described here. A more lengthy discussion of these concepts can be found in [5,6].

It should be apparent that there are a number of variations to the traversal algorithm which must be considered. In particular it must decide how the traversal will be started, how to specify where the endpoint of a step lies, how the step is constructed, and how the traversal process is terminated. These issues are discussed at length in [3,4].

3 Shape Parameters from Polygonal Harmonics

Region shape properties that can be deduced from polygonal harmonic plots are based on the prominence of the numerical features of the plots identified above. Simple, coarse-scale, shape features (such as elongation) are indicated by the existence of low order harmonics (smallest values of n) over a wide range of step lengths. Single major features (such as angular protrusion) and repeated features (such as rotational symmetry) are indicated by the existence of particular harmonics. For example harmonics of orders 2 and 3 are strongly present in Fig. 1b due to the elongated and somewhat triangular nature of the region in Fig. 1a. Overall roughness and complexity of shape are indicated by the existence of a substantial proportion of impure and complex harmonics. Such properties have been investigated experimentally elsewhere and the conclusions are summarized in [5,3]. Formulations for several such parameters will be presented in this section and the reader is referred to [6] for rigorous discussion of these.

For a given region, polygonal harmonics of various orders endure over corresponding *ranges*. These ranges should be normalized by the largest possible step length to remove variations due to the absolute size of the region.

Let $Lmax_n$ and $Lmin_n$ respectively be the greatest and least step lengths at which a harmonic polygon of order n exists for the given region. $\mathcal{L}_n = Lmax_n - Lmin_n$ is the length of the range over which harmonic polygons exist for the given region. $D = Lmax_2$ is the greatest possible step length for the given region.

Within the range corresponding to order n, interruptions may occur as *subranges* within which different orders of harmonic polygons exist. These interruptions may be due to the formation of complex harmonics or due to the temporary existence of another simple harmonic owing to some shape feature. The effect of these interruptions in diluting the range must be taken in consideration by reducing the size of $\mathcal{L}_n$ accordingly.

Definition 3. Let $Smax_{n,i}$ and $Smin_{n,i}$ be the greatest and least step lengths at which the ith interrupting subrange occurs within the step length range for harmonic polygons of order n, $i = 1, 2, 3, \ldots, imax$. For $Smax_{n,i} \geq L \geq Smin_{n,i}$, $\frac{N(L)}{W(L)} \neq n$. $\mathcal{S}_{n,i} = Smax_{n,i} - Smin_{n,i}$ is the length of the ith interrupting subrange for the range of order n. The *endurance parameter* E_n of the nth harmonic is formulated as:

$$E_n = \frac{1}{D}\left(\mathcal{L}_n - \sum_{i=1}^{imax} \mathcal{S}_{n,i}\right)$$

$$= \frac{1}{D}\left((Lmax_n - Lmin_n) - \sum_{i=1}^{imax}(Smax_{n,i} - Smin_{n,i})\right) . \qquad (1)$$

Note that $Lmax_n > Smax_{n,i} \geq Smin_{n,i} > Lmin_n$.

A relatively high endurance value indicates a tendency for the region shape to be similar to that of an equilateral polygon of that order. As the order increases, the number of possible polygon configurations grows and the shape similarity is less definite, so the first few harmonics are the most important for gross shape characterization. The relative importance of endurance values for major harmonics can be used to compare several region shapes and to derive a description of how they differ. This importance can be assessed most easily by ranking the endurances for the first few harmonics of a region in monotonically decreasing order. For a region A, construct the ordered sequence $\{E^A_{n(k)}\}$, $k = 1, 2, 3, \ldots, kmax$ with $E^A_{n(1)} \geq E^A_{n(2)} \geq E^A_{n(3)} \geq \cdots \geq E^A_{n(kmax)}$. Furthermore, let $\{n(k)^A\}$ denote the set of polygonal harmonic orders which generate $\{E^A_{n(k)}\}$. A region A may be compared with a region B by considering the harmonics appearing in such rankings for the two regions as the only ones of *significance*. Note that complex harmonics may be included amongst these significant ones if their endurance is high enough. The absolute difference sum

for the significant endurances can then be computed:

$$\Delta_E^{AB} = \sum_{i \in \{n(k)^A\} \cup \{n(k)^B\}} |E_i^A - E_i^B| \ , \tag{2}$$

which would be small when a good shape match exists. Note that the terms in the absolute differences are commutative, so $\Delta_E^{AB} = \Delta_E^{BA}$.

The effects of the interrupting or impure harmonic behaviour in the polygonal harmonic plot can be assessed in a similar way to the endurance by summing the impure subranges associated with a particular harmonic range. Consider a pair of significant harmonic ranges $\mathcal{L}_n$ and $\mathcal{L}_m$ at neighbouring step lengths, where $m > n$ so $Lmin_m < Lmax_n$. Note that $Lmin_n$ may lie on either side of $Lmax_m$. If $Lmax_m > Lmin_n$ then no further impurities, other than those already accounted for, exist in either range. If $Lmax_m < Lmin_n$, there is another impurity subrange between these values. The existence of this extra impurity subrange increases the amount of impure shape information associated with both $\mathcal{L}_n$ and $\mathcal{L}_m$.

Definition 4. Let $S_{nm} = Smax_{nm} - Smin_{nm}$ denote the length of subrange of step lengths delimited by $Smin_{nm} = Lmax_m$ and $Smax_{nm} = Lmin_n$, where m and n, $m > n$, are immediately successive orders of harmonic for which step length ranges exist. The impurity for harmonic order n must thus be defined in conjunction with knowledge about the next significant harmonic of higher order, so that this *inter-harmonic* impurity subrange can be included. The *impurity parameter* I_n associated with harmonic order n can thus be formulated as:

$$\begin{aligned} I_n &= \frac{1}{D}\left(S_{nm} + \sum_{i=1}^{imax} S_{n,i}\right) \\ &= \frac{1}{D}\left((Smax_{nm} - Smin_{nm}) + \sum_{i=1}^{imax}(Smax_{n,i} - Smin_{n,i})\right) \ . \end{aligned} \tag{3}$$

High impurity values associated with very significant harmonics indicate a shape which is complicated, with many protrusions and cavities at a large or intermediate scale. By comparison, low impurities indicate a general smoothness or straightness.

As was done with endurances, an absolute difference sum can be computed between impurities:

$$\Delta_I^{AB} = \sum_{i \in \{n(k)^A\} \cup \{n(k)^B\}} |I_i^A - I_i^B| \ . \tag{4}$$

Note that the impurities corresponding to the ranked endurances are used in this calculation and the impurity values themselves are not ranked. This is because endurances have more direct descriptive power for shape than impurities. In practice Δ_I^{AB} values have been found to be weaker measures of shape match than Δ_E^{AB} values. A more useful function is discriminating shapes when endurance

values are fairly low and fairly similar, so that shape closeness can only be inferred weakly and should be accompanied by similar impurity values (i.e. small impurity absolute-difference sum).

A clear indication of the amount of shape information that has been extracted in a partial polygonal harmonic shape analysis is given by expressing the successive endurance and impurity values as cumulative quantities. Since $\sum_n (E_n + I_n) = 1$, the amount of structured (i.e. coarse-scale) shape information is given by $\sum_n E_n$ and the amount of unstructured (i.e. fine-scale) shape information is given by $\sum_n I_n$. Neither of these quantities is changed much by insignificant harmonics, so considering the way they increase for only the first few significant harmonics can reveal much about the shape characteristics.

The existence of many complex harmonics indicates that the shape has a tendency towards intermediate scale roundness, including the nature of any large-scale protrusions and cavities. Spikes and long protrusions decrease the likelihood of complex harmonic formation, as do straight sections of the region boundary as opposed to rounded or locally rough ones. The *complexity* of the harmonic for a given step length L is the number of windings $W(L)$ made by the harmonic in its stable position. The importance of complexity in analysing a particular shape can be estimated most directly by considering the proportion of complex harmonics. Some idea of this proportion can be gained visually by constructing a complexity plot of complexity values over the range of step lengths for the region as shown in Fig. 1c. Consider the set of all subranges of step lengths over which only complex harmonics occur.

Definition 5. Let $\mathcal{T}_j = Tmax_j - Tmin_j$ denote the length of a subrange of step lengths for which $W(L) \neq 1$, $Tmin_j \leq L \leq Tmax_j$, $j = 1, 2, 3, \ldots, jmax$. The *complexity ratio* parameter R^A for a region A can be formulated as:

$$R^A = \frac{1}{D} \sum_{j=1}^{jmax} \mathcal{T}_j = \frac{1}{D} \sum_{j=1}^{jmax} (Tmax_j - Tmin_j) \quad . \tag{5}$$

The value of R^A lies between 0 and 1, and is large when the region roundness is dominant and small when straight sides and protrusions are dominant.

4 Sample Results

Before discussing our overall results, let us examine how the parameters formulated in Sect. 3 apply to the example in Fig. 1.

We first obtain from the endurances an indication of the global shape. Endurance values for the first few orders of harmonics for the data of Fig. 1a are $E_2 = 0.28$, $E_3 = 0.18$, $E_4 = 0.01$, $E_5 = 0.08$ and $E_6 = 0.02$, suggesting a somewhat elongated region via the high E_2, a strong triangular tendency via the high E_3, a slight pentagonal tendency via the modest E_5, and very little tendency towards other polygons.

Further, by examining the impurity of the harmonics which arise, we can obtain some information about the local shape. Impurity values for the first few

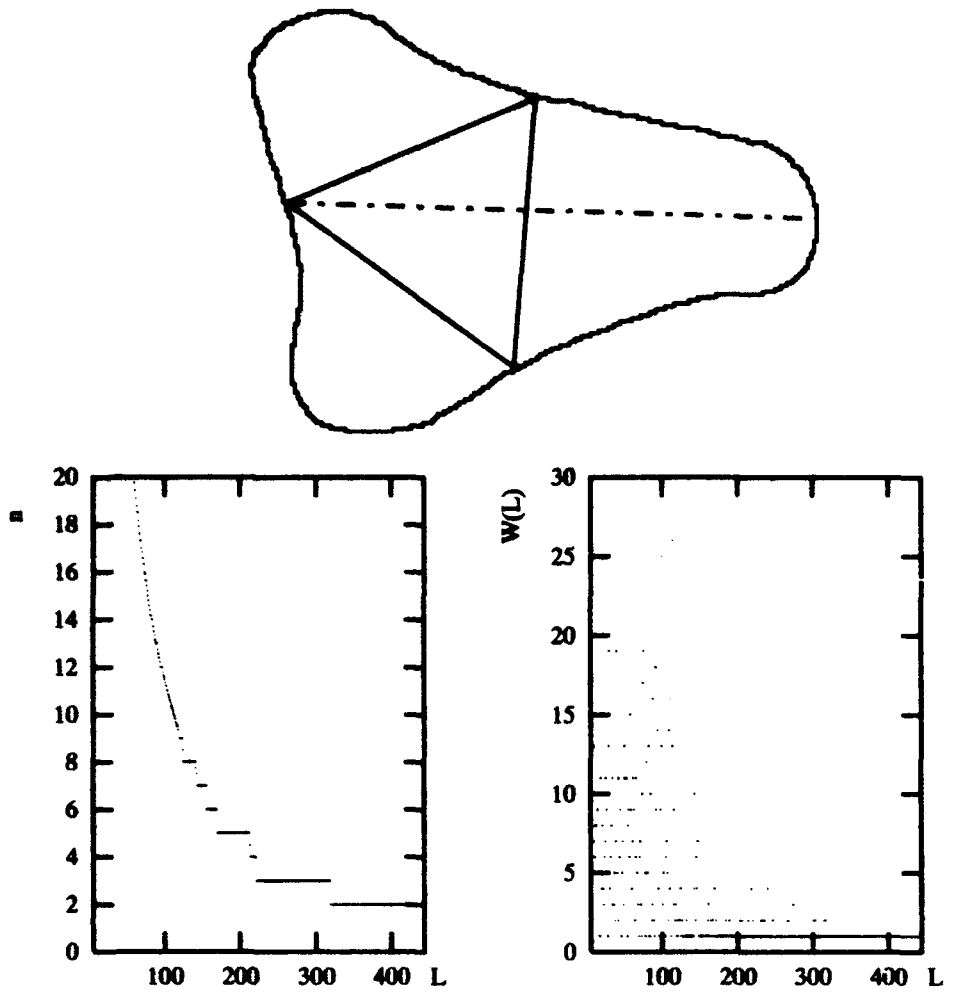


Fig. 1. (a) (top) A region boundary with superimposed polygonal harmonics of order 2 and 3; (b),(c) (left, right) Plots of polygonal harmonics and complexity for the region in (a).

orders of harmonics for the data of Fig. 1a are $I_2 = 0.00$, $I_3 = 0.04$, $I_4 = 0.01$, $I_5 = 0.01$ and $I_6 = 0.01$, indicating a region with a smooth boundary and very little local roughness. Comparing these impurity values with the corresponding endurance values, it can be seen that E_2 and E_3 give strong indications of shape, E_5 gives a weak indication of shape and the other endurances are not significant.

The moderately low complexity ratio of 0.31 for the region of Fig. 1a indicates a shape with significant protrusions, but is not low enough to suggest that there are many such protrusions.

Some examples from biological and engineering applications of shape characterization will now be considered to illustrate several of the concepts discussed above. Figure 2 shows a set of four leaf outlines consisting of two pairs of leaves from two different types of plant. Figure 3 shows a set of four particle outlines consisting of two pairs of particles resulting from two different production pro-

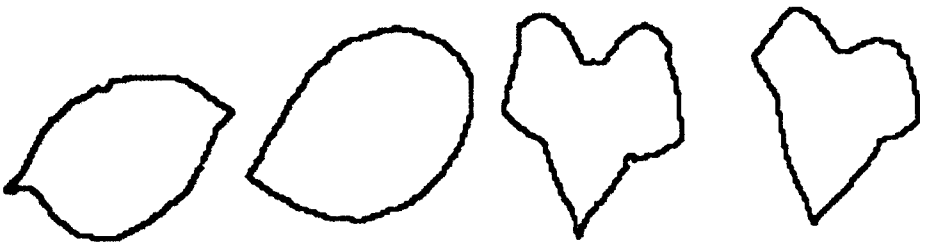


Fig. 2. (a-d) (left-right) Sample regions of four different leaves.

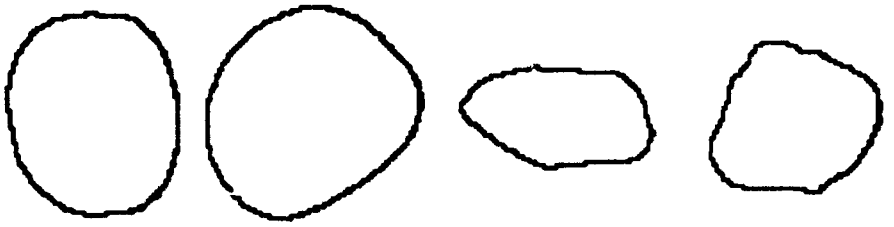


Fig. 3. (a-d) (left-right) Sample regions of four different particles.

cesses. In each case the results of polygonal harmonic shape analysis indicate basic shape facts about the outlines and allow similar shapes to be distinguished from dissimilar ones.

Table 1 provides the values of endurance and impurity for harmonic orders $n = 2, 3, \ldots, 8$ of these outlines, cumulative endurance and impurity values for three different values of $kmax$, and the complexity ratio R^A. The leaves of Figs. 2a and 2b have a somewhat elongated shape with some coarse local roughness and thus have a high E_2 characteristic, fairly low following harmonics and low impurity values. The moderate complexity ratios indicate several protrusions. Those of Figs. 2c and 2d have a much chunkier appearance, with order 3 and 5 harmonics being significant as there are dominant lobes in the shape. The complexity ratios are smaller than for 2a and 2b as the protrusions are more pronounced. Low impurity values imply that all the shapes are well characterized, but the best indicator of the extent to which this is true is seen in the cumulative endurance and impurity values, which show that between a half and two thirds of the total shape information is already contained in the harmonics considered. The four particle outlines in Fig. 3a-d are all similar in shape to a large extent, being roughly round with some elongation. Higher impurity values than those for the leaves indicate that the particle outlines are smoother (i.e. more curved in nature) and thus that the endurance shape information is less reliable. This is highlighted by the dominance of the cumulative impurity values by comparison with the cumulative endurance values for the harmonics considered. The high values for the complexity ratio in 3a and 3b distinguish these particles as being very smooth locally, while 3c and 3d are somewhat rougher.

Table 2 compares each of the shapes from Fig. 2 with the others from that same set, and likewise for Fig. 3, by displaying values for Δ_E^{AB} and Δ_I^{AB}. The

Table 1. Polygonal harmonic shape analysis of sample regions in Fig. 2 and 3.

E_n

Image	2a	2b	2c	2d	3a	3b	3c	3d
n=2	.26	.29	.05	.06	.17	.14	.15	.18
n=3	.02	.03	.17	.21	.02	.06	.01	.05
n=4	.05	.05	.00	.01	.02	.02	.01	.03
n=5	.03	.03	.11	.06	.00	.01	.00	.02
n=6	.02	.02	.01	.01	.01	.01	.00	.01
n=7	.01	.01	.00	.02	.00	.00	.00	.02
n=8	.01	.01	.02	.02	.00	.00	.00	.00

I_n

Image	2a	2b	2c	2d	3a	3b	3c	3d
n=2	.00	.00	.03	.00	.02	.01	.05	.03
n=3	.05	.07	.07	.03	.11	.11	.11	.08
n=4	.04	.05	.00	.00	.09	.09	.07	.06
n=5	.03	.04	.00	.02	.07	.07	.06	.05
n=6	.03	.03	.02	.02	.05	.06	.02	.04
n=7	.03	.03	.00	.03	.04	.04	.04	.03
n=8	.02	.02	.03	.01	.03	.04	.01	.02

$\sum E_{n(k)}$

Image	2a	2b	2c	2d	3a	3b	3c	3d
kmax=2	.31	.34	.28	.27	.19	.20	.16	.23
kmax=4	.36	.40	.35	.35	.22	.23	.17	.28
kmax=6	.39	.43	.36	.38	.22	.24	.17	.31

$\sum I_{n(k)}$

Image	2a	2b	2c	2d	3a	3b	3c	3d
kmax=2	.04	.05	.07	.03	.13	.12	.16	.11
kmax=4	.12	.16	.13	.08	.27	.28	.29	.22
kmax=6	.18	.21	.15	.09	.38	.38	.35	.29

$\sum E_{n(k)} + \sum I_{n(k)}$

Image	2a	2b	2c	2d	3a	3b	3c	3d
kmax=2	.35	.39	.35	.30	.32	.32	.32	.34
kmax=4	.48	.55	.48	.43	.49	.51	.46	.50
kmax=6	.56	.64	.51	.47	.60	.62	.52	.60

Image	2a	2b	2c	2d	3a	3b	3c	3d
R	.35	.41	.28	.22	.70	.69	.29	.42

Table 2. Comparison of sample region shapes using polygonal harmonic shape analysis. Note: Δ_E values are shown in the lower triangle and Δ_I values in the upper triangle of the sub-tables.

kmax=2

	2a	2b	2c	2d		3a	3b	3c	3d
2a		.01	.12	.06	3a		.01	.03	.04
2b	.03		.12	.09	3b	.07		.04	.05
2c	.49	.51		.09	3c	.03	.06		.05
2d	.43	.45	.05		3d	.04	.05	.07	

kmax=4

	2a	2b	2c	2d		3a	3b	3c	3d
2a		.04	.13	.07	3a		.02	.09	.10
2b	.04		.13	.11	3b	.08		.07	.10
2c	.52	.54		.14	3c	.05	.08		.07
2d	.47	.49	.12		3d	.09	.07	.11	

kmax=6

	2a	2b	2c	2d		3a	3b	3c	3d
2a		.04	.17	.09	3a		.02	.09	.11
2b	.04		.17	.13	3b	.08		.11	.13
2c	.52	.54		.14	3c	.05	.09		.10
2d	.49	.51	.13		3d	.09	.09	.15	

very low values for Δ_E for 2a, 2b, 2c, and 2d clearly confirm these shapes are very similar. In the case of 2a and 2b, the low value of Δ_I gives further weight to this conclusion, but for 2c and 2d this is not so. The values for both Δ_E and Δ_I for pairs from 3a-d are all much the same and fairly low, indicating that all of these shapes have much in common. The table also demonstrates that good shape matching can be performed using only the first few harmonics, since the strength of deductions made on the basis of $kmax = 2$ is as strong as that for $kmax = 6$.

5 Conclusion

The scope of polygonal harmonic shape analysis has been described, providing a variety of shape descriptors from a common set of harmonic numbers generated by traversal of the region boundary. The effectiveness of these descriptors has been shown by providing typical examples of shape characterization problems to which they have been applied. The descriptors have proved useful in characterizing individual shapes and in comparing shapes, as shown by the sample results.

References

1. Clark, N.N. (1987). A new scheme for particle shape characterization based on fractal dimension and fractal harmonics, Powder Technology 51, pp. 243–249.
2. Clark, N.N, Maeder, A.J, Reilly, S. (1992). Data scatter in Richardson plots, Particles and Particle System Characterization 9, pp. 9–18.
3. Davison, A.J, Maeder, A.J (1991). Properties of polygonal harmonics for coarse scale shape analysis, Proceedings of DICTA-91 Digital Image Computing: Techniques and Applications, Melbourne 4-6 December 1991, pp. 562–568.
4. Davison, A.J, Maeder, A.J (1992). PHCL: A code library implementation for shape analysis using polygonal harmonics and fractal dimension, Australian Computer Science Communications 14, pp. 243–252.
5. Maeder, A.J, Clark, N.N (1991). Harmonic endurance: a new shape descriptor derived from polygonal harmonics, Powder Technology 68, pp. 137–143.
6. Maeder, A.J, Clark, N.N (1991) Two-dimensional shape characterization in digital images using polygonal harmonic analysis. In: Cantoni, V., Ferretti, M., Levialdi, S., Negrini, R., Stefanelli, R. (eds.), Progress in Image Analysis and Processing II - Proceedings of the 6th International Conference on Image Analysis and Processing, pp. 123–130.
7. Mandelbrot, B.B (1979) Fractals: Form, Chance and Dimension. W.H. Freeman, San Fransisco.
8. Richardson, L.F. (1961). The problem of contiguity: an appendix to statistics of deadly quarrels. General Systems Yearbook, vol. 6, no. 1, pp. 139–187.

Shape Description and Classification Using the Interrelationship of Structures at Multiple Scales *

Gregory Dudek

McGill Research Centre for Intelligent Machines, McGill University, Montreal, Quebec, Canada H3A 2A7

Abstract. This paper deals with the classification of objects described by planar curves in an image. Invariance to deformation is an important aspect of shape representation and two representations are described with different degrees of such invariance. One of these is a new statistical method for shape description exhibiting a large degree of such invariance.

Using scale-space to describe shape statistically allows for a texture-like form of object classification. The scale-space used is one based on curvature-tuned smoothing (CTS). This allows a curve to be represented as a set of descriptors at various scales. The spatial correlation of these descriptors produces a statistical description of a contour that has similarities to a large-scale texture measure. The texture being measured is, in fact, the combination of substructures that define the object's shape.

Keywords: shape description, classification, recognition, tuned smoothing, active contours, statistical shape, scale-space, texture.

1 Introduction

For the purposes of object recognition, an object's shape is its most important characteristic. Computational approaches to shape-based recognition have largely focused on shape matching based on shape similarity as a template-like matching process combined with a limited amount of deformation (notwithstanding several exceptions noted below). Vision-based object recognition amounts to a process of finding the exemplar shape from a library of models whose contours best match the input shape according to some distance measure. This approach fails to describe the alternative types of shape-based recognition that is performed by humans. Consider the recognition of 2-dimensional objects such as the silhouettes of clouds or plants: such objects are eminently recognizable from their silhouettes but are often highly dissimilar in any template-like sense.

* The author gratefully acknowledges the financial support of the Natural Sciences and Engineering Research Council and the comments of M. Langer, J. K. Tsotsos and S. W. Zucker

1.1 Representational Constraint

Despite its intuitiveness, the concept of what it means for objects to have similar shapes is surprisingly hard to define. This may be, in part, because there are multiple mechanisms that contribute to the concept of shape [8, 24].

Computationally, there are several classes of shape description techniques which can be organized along a continuum or taxonomy according to their degree of *representational shape constraint*; that is, the degree of spatial freedom they permit in individual parts (or sub-parts) of an object without change to the representation (or the deformation invariance properties). Template-like representations are the most constraining allowing almost no deformation in an object's shape [3], metric representations with parameterized deformation are somewhat less constraining [20, 14], representations based on feature topology are less constraining still [1], and finally statistical shape description, a method described below, captures shape properties with extremely little positional constraint on the individual sub-shapes or features.

Two matching methods along this continuum are presented based on the same input primitives. One method is a minimum-deformation matching method, the other is a new method for shape description and representation based on statistical properties of an object's shape. The complex relationship between spatial scale and object structure has become apparent in attempts to describe object shape computationally [25, 4, 15, 12]. The statistical shape recognition method exploits the multi-scale aspect of object shape by describing objects in terms of the interrelationship between different shape features at a single location of an object contour. This leads to an object similarity measure that associates objects having similar structural properties even when they are dissimilar in a template matching or part-by-part sense. The notion of statistical shape properties and the relationship between different scales has some similarities to a texture measure [16, 18]. A key difference from conventional microtexture descriptors [23, 9] is that the primitive features here are large-scale shape primitives.

2 Curvature Scale-space Description

Shape primitives can be extracted using a variational method called curvature-tuned smoothing [5, 6]. This description has its basis in curvature measurements [2, 13], and tolerates sparse data or noise [19, 22]. The multi-scale nature of the representation allows multiple alternative descriptions for portions of a curve to be retained. It produces a description of a curve where a single region may be described in terms of one or more arcs of different curvatures (of one or more sizes), and hence makes the information at different spatial scales explicit. The term scale is used to refer to the size or spatial extent of a processing operation or feature.

The curve representation is produced by repeatedly minimizing the following energy functional with respect to a piecewise C^2 solution $\mathbf{u}(t) = (x(t), y(t))$:

$$E(\mathbf{u}(t), c) = \int_{t_a}^{t_b} ||\mathbf{u}(t) - \mathbf{d}(t)||^2 + \alpha p(\mathbf{u}(t)) + \lambda(c)(\kappa_{\mathbf{u}}(t) - c)^2 \, dt, \qquad (1)$$

where t is the arc length, $\mathbf{d}(t) = (x(t), y(t))$ is a list of initial data points estimating the input curve, $p(x, y)$ is a potential function derived from the input image (i.e. a measure of edge strength), $\kappa_{\mathbf{u}}(t)$ is the curvature of $\mathbf{u}(t)$, c is the *curvature tuning*, α is a constant, and λ is the stabilizing constant selected as a function of c. This solution is determined for various values of c, denoted by c_i. The first two terms constrain the solution to be consistent with an initial input description and with image support for the curve position. The third term expresses an *a priori* bias for a solution with a specific curvature given by c.

In practice, the discrete form of this equation is used:

$$\sum_{t \in \text{data}} ||\mathbf{u}_i(t) - \mathbf{d}(t)||^2 + \alpha p(\mathbf{u}_i(t)) + \lambda(c)(1 - l_i(t))(\kappa_{\mathbf{u}}(t) - c)^2, \qquad (2)$$

where $l_i(t)$ is an independent Boolean discontinuity function (line process) at each scale. Discontinuities are progressively inserted at each scale to satisfy a smoothness criterion.

For each value of the tuning parameter, a slightly different solution curve $\mathbf{u}(t)$ is produced that reflects structure. This combines smoothing of the input data akin to that of active contours models (i.e. *snakes* [11]), with model fitting at multiple scales although the process can also be used directly on a parameterized input curve (i.e. with $\alpha = 0$) [6].

The use of multiple alternative stabilizers for curvature-tuned smoothing leads to selecting not only various structures at different curvatures, but also structures with different spatial extents. Low curvature segments are components of circles with large radii. Conversely, the segments selected when the curvature tuning is large must also have large curvatures. As a result, differently tuned stabilizers lead to different sets of discontinuities that decompose the curve into different segments.

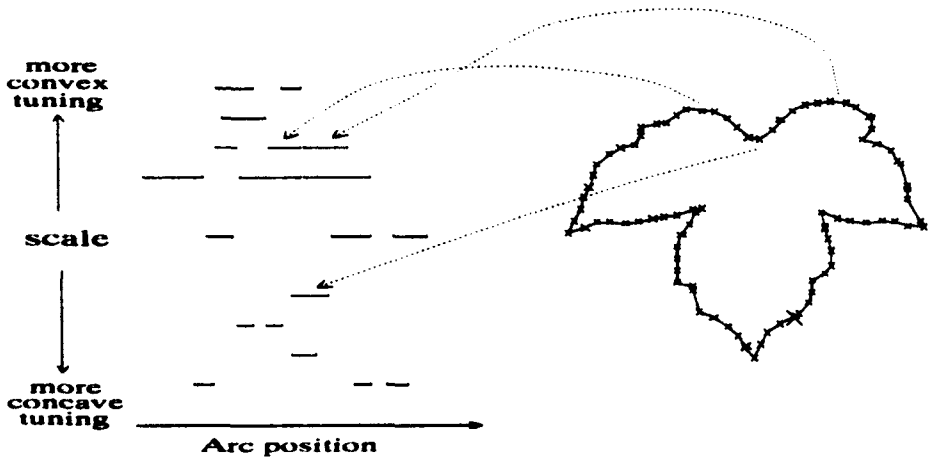


Fig. 1. Poison sumac leaf and scale-space.

The description of the poison sumac leaf (object s1) extracted using curvature-tuned smoothing. Segments corresponding to certain features on the leaf illustrated.

2.1 Abstraction into Segments

From the set of arc-like segments produced by the minimization operations it is possible to extract a small subset of the segments with high smoothness as a simplified description [7, 6]. These are the segments that best match the input data since their low energy implies that they had to deform least to suit the data (such a description is shown in Fig. 1). The segments themselves are sections of *approximately* uniform curvature, yet together they capture most of a curve's structure. The structure of each segment is so simple that it is unnecessary to retain all the internal point locations. As a coarse description the curve segments can be encoded only by their initial and final positions (t_j^I and t_j^F) and the curvature tuning c used to extract them. This encoding will be referred to as the *segment descriptor* for a segment j:

$$s_j = (t_j^I, t_j^F, c_j). \tag{3}$$

The set of segment descriptors for an object o constitutes its description $\mathcal{S}(o)$:

$$\mathcal{S}(o) = \bigcup_j s_j. \tag{4}$$

3 Matching with Deformation

Dynamic programming is one of the techniques used to match curves based on a sequence of extracted primitives such as those described above [10]. By constructing a matching function that ensures that matched curves have the same *sequence* of (multi-scale) primitives, matching is made insensitive to local deformations in a curve. For two segments s_1 and s_2 the mismatch is measured as $\langle s_1, s_2 \rangle_s = w_1 |\log c_1 - \log c_2| + |l_1 - l_2|$, where w_1 is a constant and l_1 and l_2 are the segment lengths ($|t_j^I - t_j^F|$). Note that logarithmic weighting is applied to the curvature components to impose a preference for coarse-scale information [25].

Curve matching can be formulated as a dynamic programming problem in terms of matching an increasingly long subsequence of segments from one curve to a series of segments from the other. Invariance to the initial position on either curve can be achieved by doubling the series of tokens and looking only for a substring of half the total length [6]. This has been demonstrated using an algorithm that constructs an incrementally expanded table of costs such that for two curves composed of segments, entry $C(i, j)$ in the cost table reflects the match the first i segments from one curve makes with the first j segments from the other. The process of matching one contour with another is then a process of executing the dynamic program for an observed data set against the set of models.

This procedure has been shown to be appropriate for matching curves that are noisy versions of one another or that have undergone a limited amount of deformation [6]. For pairs of curves that have significant structural variations

with respect to one another, there will be substantial mismatch error. For many natural processes structural variations may be present at a global level while sub-parts and local structures are similar. It has been suggested that one way in which this can occur is when local generative processes at different scales are combined in a pseudo-random or non-rigid manner [17, 25]. In such cases the alternative approach described below may be appropriate for shape recognition.

4 Statistical Measurement

Conventional approaches to curve recognition using local characteristics, such as the one described above, are based on determining the position of features on a curve and then using the position or spatial topology of these features for recognition. The approach described here as *scale-space statistics* is an alternative to using the relative locations of features on a curve for object recognition or classification.

At a given scale, the ease with which a curve can be described as having a given curvature c can be considered as a one-dimensional signal similar to a goodness-of-fit and will be denoted by

$$\phi(t, c) \in 0, 1 \tag{5}$$

that varies along a curve. A simple form of $\phi(t, c)$ is a binary function that indicates whether any segment descriptor having curvature c spans point t:

$$\phi(t, c) = \begin{cases} 1 \text{ iff } \exists\ s_j = (t_j^I, t_j^F, c) \ \in \ \mathcal{S}(o) \text{ and } t_j^I \leq t \leq t_j^F, \\ 0 \text{ otherwise .} \end{cases} \tag{6}$$

By observing the mean value $\overline{\phi(c)}$ of this function, we can describe "how much" of a contour can be well-approximated at the given curvature.

The similarity between the one-dimensional functions $\phi(\cdot, c)$ for different values of c indicates the interrelationship between the different-scale substructures that make up the curve at each point. As noted above and in the texture literature, specific statistical interrelationships are characteristic of many shapes including a variety of natural forms. Common examples include the trunks of trees, typified by a large-scale cylindrical curve combined with fine-scale bark patterns, geological formations, or the way the bumps and ridges on the leaves of a tree are combined. Note also that many objects are recognizable even though the sequence of sub-curves that compose them may be highly variable (Fig. 2).

The cross-correlation matrix C has elements defined by

$$C_{ij} = \int \frac{(\phi(t, c_i) - \overline{\phi(c_i)})(\phi(t, c_j) - \overline{\phi(c_j)})}{\sigma_{\phi(c_i)}\ \sigma_{\phi(c_j)}} dt \tag{7}$$

between this value at one curvature and the value of this function at another curvature. It provides a measure of what types of substructure in curvature space occur within a structure at another scale. This corresponds to taking a slice of

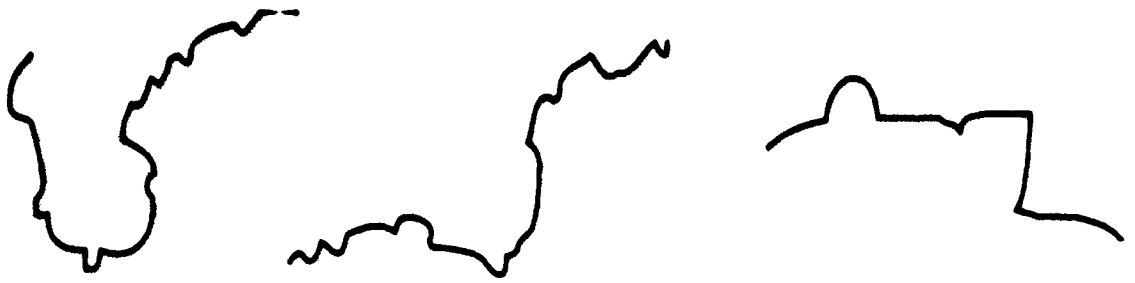


Fig. 2. Statistically similar objects. The first two coastal curves are similar in a structural or statistical sense, yet they cannot be globally deformed into one another easily; the third is different.

Fig. 3. Sample input curves. Left to right, top to bottom: r1 (raspberry), m1 (maple), a9 (unknown), r2 (raspberry).

the scale-space for a fixed position and measuring the statistical likelihood of features at one scale given the presence (or absence) of features at another scale.

Together, the vector $\overline{\phi}$ and the matrix C provide a statistical description of a curve which is similar to a texture measure for an intensity pattern. Whereas texture is often measured by decomposing a signal into different components such as bandpass channels [23, 21], the statistical shape measure presented here relates texture to the goodness-of-fit of shape operators at different curvature-based scales.

For appropriate classes of shapes, these statistical scale-space measures can be used directly for shape matching. The simplest such shape measure for two shapes o_1 and o_2 being compared is

$$\mathcal{M}(o_1, o_2) = \frac{C_1 \cdot C_2}{||C_1||||C_2||}, \tag{8}$$

where $\cdot$ denotes the dot or inner product. Shapes with identical scale-space statistics thus match with value 1, while unrelated shapes have a match score of zero.

Since C_1 and C_2 have uniform diagonals caused by autocorrelation, $\mathcal{M}$ has a positive offset. Cross-talk between the responses at different scales leads to a consistent positive bias for near diagonal elements as well. This off-diagonal coupling across scales, however, cannot readily be estimated *a priori* and depends

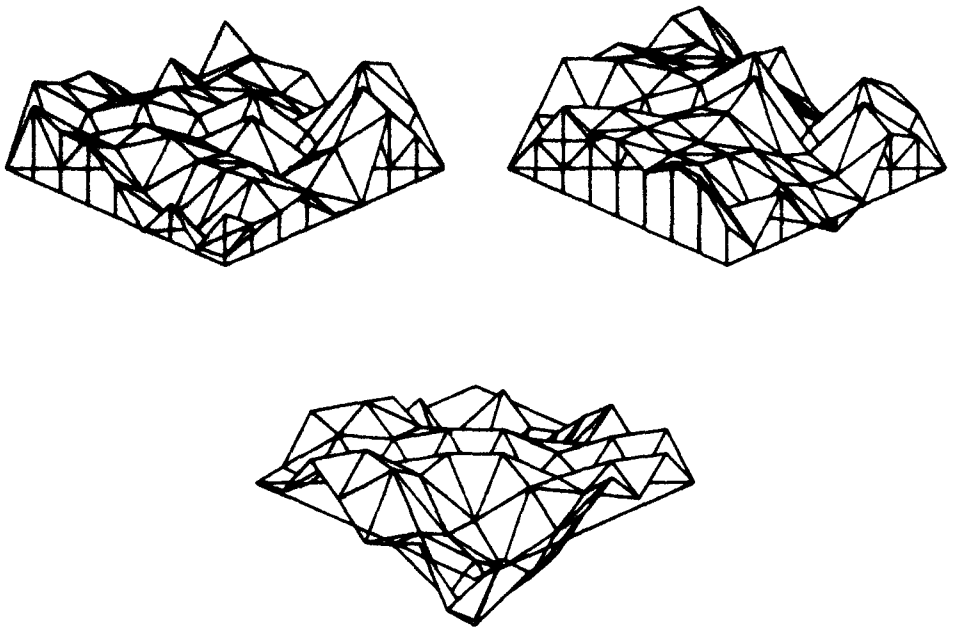


Fig. 4. Scale-space correlation surfaces. Three scale-space correlation surfaces for three different curves from the previous figure (leaf silhouettes). Curvature tuning (or scale) varies along each axis and the amplitude at any point reflects the correlation between ϕ signals for the two curvatures. The top two surfaces are from two different leaves of the same type (examples r1 and r2 at different orientations in depth). The lower surface is from a different type of leaf (example a9); note its qualitatively different profile.

on non-linear discontinuity effects in the original solutions. Hence, an additional heuristic is of utility: elements (correlations) of C that are well off the diagonal, corresponding to correlations between signals well separated in scale, can be more heavily weighted. This is further grounded in the observation that, in general, structures at different scales are independent except where non-accidental processes lead this to be otherwise; hence such structural correlations are especially salient [26, 17].

Hence, we have a refined measurement of the form:

$$\mathcal{M}^w(o_1, o_2) = \frac{C_1 \odot W \cdot C_2 \odot W}{||C_1 \odot W|| * ||C_2 \odot W||}, \tag{9}$$

where $\odot$ denotes the Hadamard product ($C(i,j) * W(i,j)$) and W is a weighting matrix of the form:

$$W(i,j) = 1 - e^{-|i-j|}. \tag{10}$$

In this way, an improved signal-to-noise ratio for the matching task is obtained.

5 Results

The results of matching particular contours (e.g. object *m1* of Fig. 3) to several others are tabulated below, each to two significant figures (the first letter

indicates the leaf species, the numerical suffix indicates the example; m and r species are intuitively similar):

Curve	$\mathcal{M}^w(m1,\cdot)$	$\mathcal{M}^w(t1,\cdot)$	$\mathcal{M}^w(a9,\cdot)$
m1	1.0	0.65	0.29
m2	0.79	0.72	0.44
r1	0.73	0.87	0.24
r2	0.71	0.78	0.17
s1	0.65	0.54	0.21
t1	0.65	1.0	0.086
a9	0.29	0.086	1.0

Note that the m1 and r1 contours and their deformed versions are rated similar to one another while other contours have much lower scores.

The statistical representations C and the matching function $\mathcal{M}^w$ describe the relationship between structures of different types without regard for the precise spatial arrangement of the structures. For example a large bump may equivalently contain several concavities without regard for the positions of the concavities with respect to one another. This form of view invariance has both advantages and shortcomings. A disadvantage of this coarse abstraction of a curve's shape is that it is insensitive to a large variety of possible variations in the object, in particular those that are obtained by reordering the major sections of the shape. On the other hand, this abstraction permits measurement of the the similarity between different shapes that have the same cross-scale structure because they are composed of the same building-block parts, but in different numbers or arrangements. For example, various natural forms such as cloud types are typified by the combination and co-occurrence of particular forms at multiple scales, for example lobes with serrations, whereas the specific spatial arrangement of the forms is highly variable. In essence, this simple shape measurement is best suited to classes of objects where a small number of interacting generative processes are responsible for each object, and each of these processes can be typified as creating subshapes at a particular scale but with random or hard-to-typify spatial arrangements. This characterization appears to be appropriate for many types of natural form such as rocks, leaves, microscopic particles, and clouds.

6 Conclusion

The use of a collection of curvature-based minimizing operators, which are termed collectively *curvature-tuned smoothing*, has been previously developed to address several difficulties with existing approaches to smoothing, interpolation, segmentation, and curve description. Using this representation as input, techniques for describing and recognizing objects via the sequencing of descriptors along the curve and via the correlation statistics of the descriptors in this space have been outlined.

The statistical method provides a notion of recognition based on structural regularities in shape rather than direct point-to-point similarity. As such, it allows objects to be recognized or deemed alike even when they have no identical sub-contours. Because the primitive elements in this description (bumps and valleys) are perceptually and functionally salient, the shape-similarity space can be described in intuitive or generative terms (for example it can be related to processes that produce bumps and valleys). Statistical shape description can also be formulated in terms of alternative primitive shapes if this is appropriate to specialized domains. This particular class of similarity appears well suited to the recognition of certain classes of biological and geological forms where the structural characteristics are common to the class, but individual members vary in terms of their particular layout.

The two matching techniques presented illustrate very different positions along a proposed continuum for the classification of shape matching methods.

References

1. Ansari, N., Delp, E. J. (1990). Partial shape recognition: A landmark-based approach, IEEE Trans. Pattern Analysis and Machine Intelligence 12(5), pp. 470–483.
2. Attneave, F. (1954). Some informational aspects of visual perception, Psychological Review 61, pp. 183–193.
3. Cass, T. A. (1988). A robust parallel implementation of 2-d model based recognition, Proc. Conf. on Computer Vision and Pattern Recognition Ann Arbor, MI., pp. 879–884.
4. Crowley, J. L., Parker, A. C. (1984). A representation for shape based on peaks and ridges in the difference of low-pass transform, IEEE Trans. Pattern Analysis and Machine Intelligence 1(2), pp. 156–170.
5. Dudek, G., Tsotsos, J. K. (1989). Using curvature information in the decomposition and representation of planar curves, Proc. NATO Advanced Study Institute of Robotics and Active Vision, Maratea, Italy.
6. Dudek, G., Tsotsos, J. K. (1991). Shape representation and recognition from curvature, Proc. of the 1991 Conference on Computer Vision and Pattern Recognition Maui, Hawaii. IEEE Press, pp. 35–41.
7. Dudek, G. L. (1990). Shape representation from curvature, PhD Thesis, Dept. of Computer Science, University of Toronto, Toronto, Canada.
8. Fischler, M. A., Bolles, R. C. (1983). Perceptual organization and the curve partitioning problem, Proc. Int. Joint Conf. on Artificial Intelligence, pp. 1014–1018, Karlsruhe, Germany.
9. Fogel, I., Sagi, D. (1989). Gabor filters as texture discriminators, Biol. Cybern. 61, pp. 103–113.
10. Gorman, J. W., Mitchell, O. R.,, Kuhl, F. P. (1988). Partial shape recognition using dynamic programming, IEEE Trans. Pattern Analysis and Machine Intelligence 10(2), pp. 257–266.
11. Kass, M., Witkin, A., Terzopoulos, D. (1988). Snakes: Active contour models, Int. J. of Computer Vision 1(4), pp. 321–331.

12. Kimia, B. B., Tannenbaum, A.,, Zucker, S. W. (1990). Toward a computational theory of shape: An overview, Proc. First European Conf. on Computer Vision, Antibes, France.
13. Koenderink, J. J., van Doorn, A. J. (1980). Photometric invariants related to solid shape, Optica Acta 27 (7), pp. 981–996.
14. Milios, E. (1988). Recovering shape deformation by an extended circular image representation, Proc. 2nd Int. Conf. on Computer Vision, Tarpon Springs, FL., IEEE Press, pp. 20–29.
15. Mokhtarian, F., Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes, IEEE Trans. Pattern Analysis and Machine Intelligence 8 (1), pp. 34–43.
16. Pentland, A. P. (1984). Fractal-based description of natural scenes, IEEE Trans. Pattern Analysis and Machine Intelligence 6 (6), pp. 661–674.
17. Pentland, A. P. (1985). Perceptual organization and the representation of natural form, technical note 357, SRI International.
18. Pentland, A. P. (1987). Perceptual organization and the representation of natural form. In: Fischler, M. A., Firschein (eds.), Readings in Computer Vision (also in SRI TR-357 1985), Morgan Kaufman Publishers, Los Altos, California, pp. 680–698.
19. Rektorys, K. (1980). Variational Methods in Mathematics, Science and Engineering, Reidel, Dordrecht, Holland.
20. Solina, F. (1987). Shape recovery and segmentation with deformable part models, PhD Thesis, Dept. of Computer and Information Science, Univ. Pennsylvania.
21. Subirana-Vilanova, J. B. (1991). On contour texture, Proc. Conf. Computer Vision and Pattern Recognition 1991, Maui, HA, IEEE Computer Society, pp. 753–754.
22. Terzopoulos, D. (1986). Regularization of inverse visual problems involving discontinuities, IEEE Trans. Pattern Analysis and Machine Intelligence 8 (4), pp. 413–424.
23. Turner, M. R. (1986). Texture discrimination by gabor functions, Biol. Cybern. 55, pp. 71–82.
24. Warrington, E. K., Taylor, A. M. (1978). Two categorical stages of object recognition, Perception 7, pp. 695–705.
25. Witkin, A. P. (1983). Scale-space filtering, Proc. 3rd Int. Joint Conf. on Artificial Intelligence, Vol. 2, Karlsruhe, West Germany.
26. Witkin, A. P., Tenenbaum, J. M. (1983). On the role of structure in vision. In: Rosenfeld, J., Beck. B., Hope. B. (eds.), Human and Machine Vision, Academic Press.

Learning Shape Classes

Stanley M. Dunn and Kyugon Cho

Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey 08855, USA

Abstract. This paper describes a shape representation technique for learning shape classes. This representation technique is based on the notion of representing categorical shape knowledge; shape itself is represented by so-called conjunctions of local properties (CLP). Shape concepts are learned by a technique called *property-based learning*, an incremental learning method that inductively selects properties crucial for classification. Unlike other classification methods based on distances or similarities, classification performance does not degrade as the number of classes increases and classification can be done correctly with only partial information of instances.

Using this shape representation, shape prototypes can be learned and shapes can be classified in the presence of viewpoint changes, local movements (such as moving handles of pliers or fingers) and occlusion.

Keywords: shape classification, shape representation, shape learning system, conjunctions of local properties, property-based learning.

1 Introduction

A class is a set of instances; a *shape class* is a set of *similar* shape instances. *Shape classification* is the process of labelling *shape* instances with their correct class names based on a representation of the shape. Usually, a shape classifier consists of a shape representation and a classifier model. Building such a shape classifier manually is tedious, and sometimes impossible. *Learning* is a good strategy for building a shape classifier. A shape learning system, that is, a shape classifier with learning capability, has a learning module that modifies its classifier based on its output and the correct class name provided by a supervisor.

The problem of learning shape classes is the same as that of *learning from examples* except that the instances are specified by shapes. This different assumption about the instance description raises many different research issues; some of the important issues are touched on in this paper.

Shape classification is difficult because shapes in a class may vary by many different factors: sensor and digitization noise, viewpoint, moving or flexible parts and occlusion.

These different sources of shape variation may be mixed in a shape class. Depending on the assumptions about shape variation, different classification approaches are possible. However, it is difficult to model arbitrary deformations where only some similarity is preserved. Image processing operations can also introduce variation since imperfect selection of a region of interest will introduce artifacts similar to occlusion. All of these contribute to the variation within a class, making classification difficult.

Existing systems that can learn shape differ (1) in their assumptions about input and output, (2) in their shape representation methods, (3) in their classifier models, and (4) in their learning methods.

The shape representation used in [11] consists of moments of a compact region; it has limitations since the features are subject to change by occlusion. The learning algorithm ACLS is an example of a decision-tree learning method [8] where each leaf node represents a class. A class can have multiple paths and the paths represent disjunctions of conjunctions.

The shape representation system in [3] is one example of shape representations by structural description. 2-D shapes are decomposed into subshapes using smoothed local symmetries [1] and a semantic network is computed from the subshapes. To learn shapes represented by semantic networks, a learning method that is a modified version of ANALOGY [14] is described. This algorithm incrementally updates the description of a concept with training examples; the concepts are generalized or specialized depending on the type of example. Unlike [14], the system in [3] allows disjunctive concepts.

Paper [4] describes a system for automatic generation of object class descriptions. The learning method is based on the INDUCE algorithm [6], where instance descriptions are generalized by a set of rules. The *generalization rule* set which generalizes two (less general) descriptions is extended into a general description. The control strategy to generate concept descriptions using generalization rules is a generate-and-test mechanism with a modification of an *external disjunction rule* that generalizes every pair of descriptions by their disjunction.

There is another group of approaches known as *feature hierarchies* [7] in shape representation techniques. The basic idea is to detect the most fundamental parts of shapes (such as straight line segments or points) and then these parts can be combined to form higher-level parts. *Layered graphs* [9] is an example of this approach, in which curvature extrema are the basic level parts. Layered graphs represent shape instances and *probabilistic layered graphs* represent shape concepts. The probabilistic layered graph of a class is constructed incrementally from its instances. Classification is to find a concept that maximally simplifies the description of the instance. In [10], Segen extended the original method [9] to handle disjunctive concepts by allowing multiple probabilistic layered graphs for a concept.

In Sect. 2, *prototypical* representations of shapes are used for classification and it is shown that a conjunction of local properties is a good categorical representation for classification purposes. In Sect. 3, a learning paradigm called property-based learning is proposed as a learning method for concepts that are described by conjunctions of many properties. In Sect. 4, some experimental results are presented.

2 Shape Representation for Reasoning Tasks

This section is an overview of the shape representation that was found to be most amenable for cognitive tasks such as the task of learning to be discussed later. The criteria that this representation must satisfy are:

Input criterion: The input to shape representation should be computable and general enough to cover any shape concept, including skeletons and boundaries that are not closed.

Output criterion: The output of the representation must be easily usable in classification processes, thus constraining the syntax of representation.

Uniqueness criterion: Shape representations should be unique for distinct shapes and invariant to some geometric transformations. This criterion is an interpretation of Marr's *scope* vs. *uniqueness* criteria [5].

Consistency criterion: Similarity between shapes should be preserved in representations. A representation should be changed *locally* by any *geometrically local* shape changes which are regarded as non-noise, making the representation *sensitive* and *stable* [5] at the same time.

The shape representation which satisfies all four of these criteria is called the *conjunction of local properties* (CLP). By conjunction we mean a logical conjunction; a local property is a feature of the shape that characterizes local structure. The regions where local properties are defined are not mutually disjoint but overlapping. The places where the local properties are computed are robust.

2.1 Local Properties of Straight Line Segments

CLP1 is a specific representation based on local properties defined on straight line segments. The input is a line drawing, curves are approximated with straight line segments first and local properties are computed for each pair of straight line segments. The *conjunction of all properties computed from* all straight line segments is the representation of the shape. The details of the straight line approximation method using corner detection and the iterative endpoint algorithm are explained in [2]. This method selects features that are stable through scale for endpoints of the line segments.

Figure 1 illustrates representing a local property of 2 straight line segments. The local properties of a pair of line segments are the *position* and *orientation* of the second line segment with respect to a coordinate system established by the first line segment. The local properties of straight line segments are the *position* (r and θ) of the centre of the line encountered and the *slope* (ϕ) of the line encountered. They are invariant to 2-D translations, rotations, and scale. Local properties are computed whenever the directional search encounters another line segment.

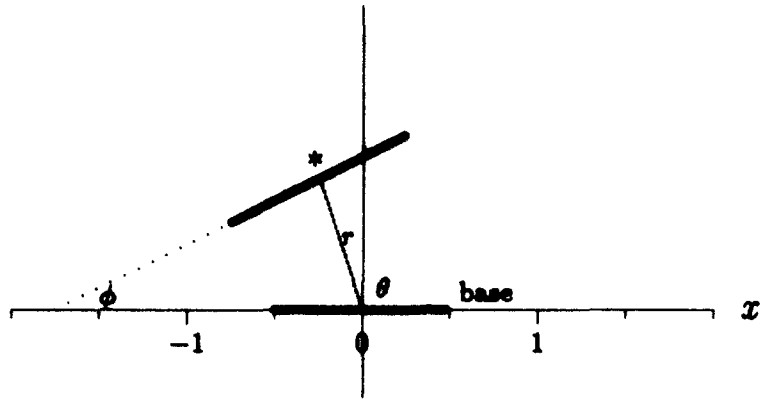


Fig. 1. Three parameters; r, θ, and ϕ

Implementation constraints on quantizing length and angle have the effect of constraining the locality of the properties. This fact has been exploited by Stein and Medioni [12], where the number of segments in a supersegment are limited. The smaller the number of segments, the more local is the information in supersegments.

In [2], it is shown (the uniqueness theorem of CLP1) that the straight line local property representation of a distinct shape is weakly unique (that is, there is at least one different weight) if the number of directional searches is large enough. Furthermore, the representation of a distinct shape is strongly unique (that is, there is at least one distinct property) if the shape is not the part of an equilateral polygon and the number of directional searches is large enough. For more details on the notions of weak and strong uniqueness, see [2].

The consistency criterion is satisfied by a conjunction of local properties representation and specifically is satisfied by local properties of straight line pairs. Under a local shape change such as occlusion (see Sect. 5 for examples), the properties related to the occluded part will be changed while other properties remain consistent, since the representation is a conjunction of local properties and the local properties of the unoccluded portion have not changed. Local properties of straight lines are also consistent under an imperfect extraction of line segments, as shown in Fig. 2. Figure 2 (a) shows a shape with 2 straight line segments and Fig. 2 (b) shows the same shape where one line segment is broken. In CLP1, most of the properties computed from the (unbroken) base are the same even though the other segment is broken. Of course, the properties computed from the broken line segment vary. In other words, the representation is changed locally.

2.2 Approximating Curves with Straight Line Segments

The straight line approximation of a curve is a combination of corner detection and an iterative endpoint algorithm; full details are given in [2]. If the curve segments are closed contours, then the approximation using the classical iterative

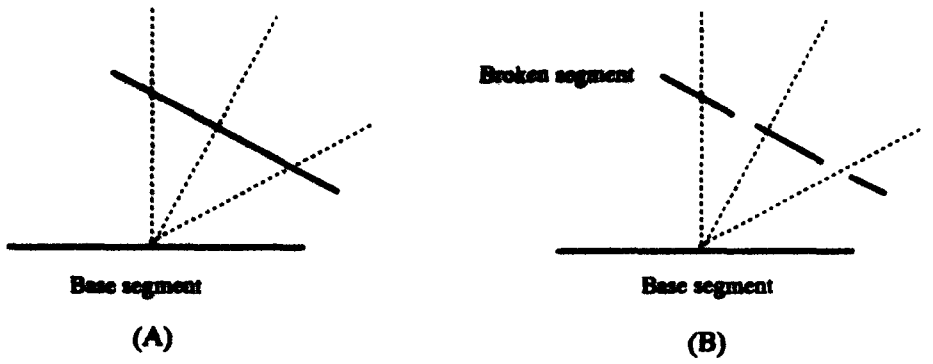


Fig. 2. (a) shows the local properties of two stable line segments. (b) shows the local properties of a stable base segment and a broken line segment. Most of properties remain the same.

endpoint algorithm is dependent on the starting point. This is undesirable for a robust approximation.

We apply a corner detector before the iterative endpoint algorithm to overcome this problem. The common definition of corner point is a point of high curvature. Many corner detectors use a definition of curvature in a single scale; here a corner is defined as a point of high curvature that is stable through scale-space. Only corners that are detected from small to large scales consistently are selected as real corners.

The straight line approximation procedure using corner detection and iterative endpoint approximation for the straight lines and the iterative endpoint algorithm are described below.

Straight Line Approximation Algorithm

```
with a given threshold for curvature
all points on a curve segment are corner candidates
from a small scale to a large scale
    for each candidate point
        calculate the curvature
        if the curvature is above the threshold
        or the curvature is not the largest within the scale
            exclude the point from candidates
for each curve segment defined by 2 consecutive corners
    call IEP
```

Iterative Endpoint Algorithm

```
with a given curve segment
if the the error between the curve and the straight line
connecting the endpoints is larger than the threshold
    select a point where the error is largest
```

split the line at that point
call IEP recursively for the split segments

To evaluate the fitting error, the distance between a point and a straight line segment is defined as the shortest distance between the point and any points on the line segment. The distance between a line segment and a curve segment is defined as the average of the distances between each point on the curve segment and the line segment, to suppress noise effects.

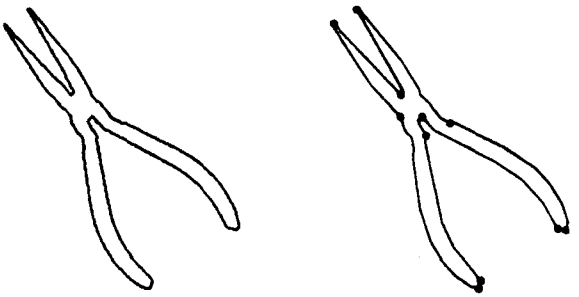


Fig. 3. Boundary of a pair of longnose pliers and its approximation by straight lines.

The approximation algorithm was tested on a boundary of a pair of longnose pliers. In Fig. 3 the top shows the original boundary and the bottom shows the corner points (small circles) found and the approximation with the straight line segments.

3 Property-based Learning

In this section, a new learning paradigm for CLP representations, called *property-based learning* (PBL), is presented. In both the learning and recognition procedures we will take advantage of two observations from cognitive psychology: namely, that recognition is composed of two phases of hypothesis formation and verification; and also, that the similarity of an object to a class is not necessarily the same as the similarity of the class to the the object. The shape representation described in Sect. 2 is not only robust with respect to possible changes in shape, but is also a facile shape representation for these cognitive tasks. This section shows how CLP (specifically, CLP1) can be used in learning and recognition where machine recognition is based on these two observations of cognitive psychology.

Classification in property-based learning is achieved by an indexing and a matching algorithm. Indexing is based on the similarity of instances to concepts; it is represented by a property-centred table (PCT). The property-centred table can be accessed by a property and the entry is a list of classes that have the property. Matching is based on the similarity of concepts to instances; it is

represented by the concept-centred table (CCT). The CCT can be accessed by a class and it has a list of property-weight pairs.

In PBL as in "concepts as prototypes" of cognitive psychology, *similarity* is an organising principle by which individuals classify instances, form concepts, and make generalizations. In property-based learning a *set-theoretical* approach to similarity is used rather than metric and dimensional approaches. Our similarity model is a special case of the ratio model [13]. In property-based learning, this asymmetry in similarities are exploited in the two step model of classification, consisting of *indexing* and *matching*. Indexing is based on the similarity of an instance to concepts, and matching is based on the similarities of concepts to the instance.

The learning part includes a *reinforcement* algorithm and a *garbage-collection* algorithm. The reinforcement algorithm acquires new properties or adjusts weights of properties for each class when the classification makes an error. The garbage-collection algorithm removes insignificant properties.

The goal of indexing is to retrieve some hypotheses with an instance. If a property is found solely in a single class, the property is a good clue for recalling the class. If a property can be found in several classes but not all, the property can be used to constrain the set of hypotheses. With an instance that is a list of properties, there can be many constraints. The similarities of instances to classes score the constraints systematically for each class. The similarity of instance I to a concept C is

$$s(I, C) = \frac{f(I \cap C)}{f(I)}$$

where the function $f(I)$ is the number of properties in I and the function $f(I \cap C)$ is the number of properties both in I and C.

The *matching* algorithm orders classes by the similarities of indexed classes (concepts) to the instance. The similarity of a concept C to an instance I is

$$s(C, I) = \frac{f(C \cap I)}{f(C)}$$

where $f(C \cap I) = \sum_i member(p_i, I)w_i$, $f(C) = \sum_i w_i$, and $member(p_i, I)$ is 1 if $p_i \in I$, otherwise 0, $C = ((p_1, w_1), (p_2, w_2), \cdots)$, $I = (p_1, p_2, \cdots)$.

Learning in PBL changes the property- and concept-centred tables so that a better classification result can be achieved. Two algorithms, called *reinforcement* algorithm and *garbage-collection* algorithm, form the basis of the learning; the PCT and CCT are changed only if the classifier's output is wrong.

Let an instance I be input to the classifier, assume that the classifier's output c_c is incorrect, and the correct class name c_s is given by a supervisor. Then, the learning algorithm is

for every p_i in I
    if p_i is in $CCT[c_s]$, $w_i \leftarrow w_i + 1$
    else insert $(p_i, 1)$ in $CCT[c_s]$
for every p_i in I
    if p_i is in $CCT[c_c]$, $w_i \leftarrow w_i - 1$

for every p_i in I
    if c_s is not in $PCT[p_i]$ insert c_s in $PCT[p_i]$
where $I = (p_1, p_2, \cdots)$, $PCT[p_i] = (c_1, c_2, \cdots)$,
and $CCT[c_c] = ((p_1, w_1), (p_2, w_2), \cdots)$.

Another important algorithm managing the PCT and CCT is *garbage-collection*. In a class, if a property has relatively small weight when compared to the others, it means the property is not important for identifying the class any more. The garbage-collection algorithm removes such properties. The algorithm is

let the maximum weight in $CCT[c_s]$ be w_{max}
for every p_i in $CCT[c_s]$ if $\frac{w_i}{w_{max}} < noise\text{-}ratio$
    delete the pair (p_i, w_i) from $CCT[c_s]$ and delete c_s from the list $PCT[p_i]$
    similarly for c_c

Let the number of properties of an instance be m, the average number of properties to describe a concept be n, and the average number of concepts that can be recalled by a property be k. Then, the complexity of indexing is $O(mk)$, the complexity of matching is $O(mn)$ and the complexity of classification is $O(mk + mn)$. Note that *this does not depend on the number of concepts learned.*

The complexity of the reinforcement algorithm is $O(mk + mn)$ since the cost of updating PCT is $O(mk)$ and the cost of updating CCT is $O(mn)$. The complexity of the garbage-collection algorithm is $O(nk)$. The cost of learning, which is necessary only if the system makes an error, is $O(mk + mn + nk)$.

4 Experiments with CLP and PBL

4.1 Classification of Tools

Figure 4 shows some of the sample data used in the shape learning and classification experiments. For each tool (longnose pliers, nippers, and pliers), 16 different images were taken changing the viewpoints. Four extra images for each tool were taken for testing and in some cases they are only partial views of the tools. The results of the shape classification (with learning) of the three experiments performed with this data are given in Table 1. In most of the cases, the system learned the training examples within 5 training cycles. The performance of the system varies with the order of examples because the learning is incremental.

Table 2 shows the effect of the garbage-collection algorithm. The table entries are the average number of properties in the CCT for each tool during experiment III. The next row shows the number of distinct properties of each tool. The garbage-collection algorithm reduces the size of property-weight pair lists in CCT roughly by half. This means that only half the distinct properties in each class were necessary for classification and the others are either erroneous or redundant. This explains why indexing schemes without any learning mechanisms are sensitive to noise.

The significance of these experiments is that the system can learn to classify instances of different viewpoints, instances with moving parts, and instances with occlusion.

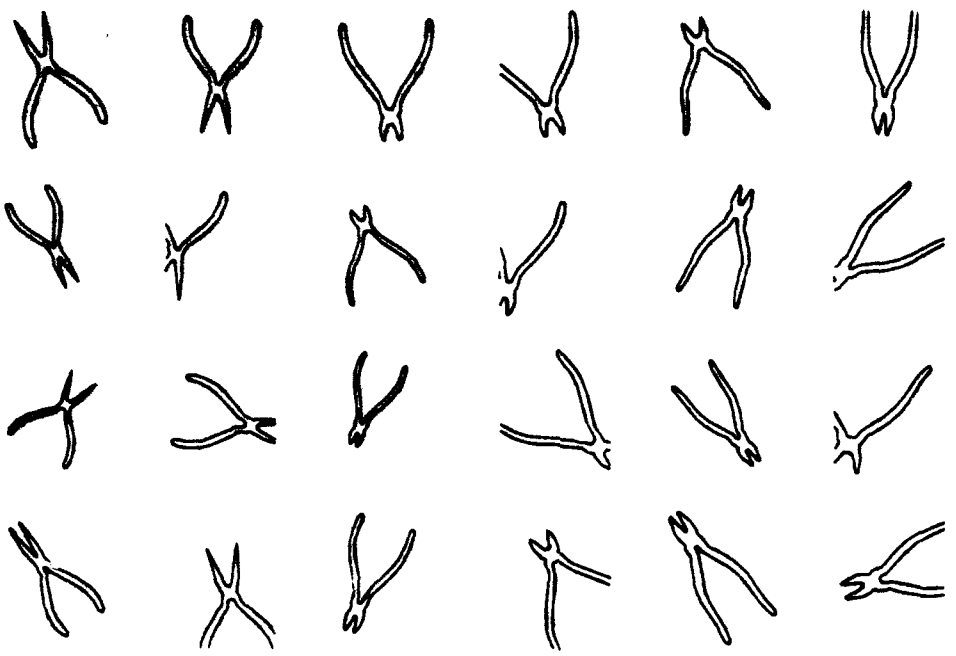


Fig. 4. Samples of the tool data.

Table 1. Success rates of 3 experiments

	1 2 3 4	Average
Trained on 48 (12)	12 10 11 11	91.67%
Leave One Out (48)	42 44 41 45	89.58%
Leave One Out (60)	52 59 56 58	93.75%

Table 2. Number of properties in CCT

	longnose	nippers	pliers
CCT	228.3	311.3	383.1
# properties	586	582	682

5 Conclusion

The main contribution of this research can be summarized as follows:

- As has been emphasized, learning is necessary because the shape variation of a class cannot always be defined.
- The criteria of shape representation for classification have been redefined and restated.
- A new paradigm of shape representation for classification, called CLP, was introduced. It was shown that representation by CLP can satisfy the shape representation criteria.

- A new learning paradigm called PBL was introduced. PBL is unique in that this is the only feasible learning method when instances are characterized by conjunctions of many properties. PBL is unique in that indexing and matching is combined so that there is no degradation in performance as the number of classes increases. PBL is also unique in that it can classify instances with partial information.

References

1. Brady M., Asada, H. (1984). Smoothed local symmetries and their implementation. In: Brady, M., Paul, R. (eds.), Robotics Research: The First International Symposium, pp. 331–354.
2. Cho, K. (1992). Learning Shape Classes, PhD thesis, Rutgers University.
3. Connell, J.H., Brady, M. (1987). Generating and generalizing models of visual objects, Artificial Intelligence, 34(2), pp. 159–183.
4. Cromwell, R.L., Kak, A.C. (1991). Automatic generation of object class descriptions using symbolic learning techniques, Proc. 9th National Conf. on Artificial Intelligence, AAAI, pp. 710–717.
5. Marr, D. (1982). Vision. W.H. Freeman and Company.
6. Michalski, R.S. (1980). Pattern recognition as rule-guided inductive inference, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2(4), pp. 349–361.
7. Milner, P.M. (1974). A model for visual shape recognition, Psychological Review, 81(6), pp. 521–535.
8. Quinlan, J.R. (1986). Induction of decision trees, Machine Learning, 1(1), pp. 81–106.
9. Segen, J. (1988). Learning graph models of shape, Proc. 5th Int. Conf. on Machine Learning, pp. 29–35.
10. Segen, J. (1990). Graph clustering and model learning by data compression, Proc. 7th Int. Conf. on Machine Learning, pp. 93–101.
11. Shepherd, B.A. (1983). An appraisal of a decision tree approach to image classification, Proc. Int. Joint Conf. on Artificial Intelligence, pp. 473–475.
12. Stein, F., Medioni, G. (1992). Structural Indexing: Efficient 3-D object recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, 14(2), pp. 125–145.
13. Tversky, A. (1977). Features of similarity, Psychological Review, 84(4), pp. 327–352.
14. Winston, P.H. (1975). Learning structural descriptions from examples. In: The Psychology of Computer Vision, McGraw-Hill.

Inference of Stochastic Graph Models for 2-D and 3-D Shapes

Jakub Segen

AT&T Bell Laboratories, Holmdel, NJ 07733, USA

Abstract. Shape interpretation methods that model a shape using stochastic graphs can recognize many classes of nonrigid objects, even if the objects are partially occluded, and interpret complete scenes composed of overlapping nonrigid shapes. These methods can also identify most parts of each shape. This paper describes the use of a stochastic graph as a model for a class of 2-D or 3-D shapes, and presents learning methods that infer stochastic graph models and their symbolic primitives from examples. These methods, as well as a graph-covering method used for scene interpretation, use a criterion of minimum description complexity which eliminates the need for subjective parameters. One practical application of this work is a trainable real-time system that recognizes hand gestures in 2-D images.

Keywords: shape description, relations, graph representation, stochastic graph, minimal description length, MDL.

1 Introduction

Structural descriptions composed of parts and relations are often used as models and representations of shape. Some examples of such use are in the work of Barrow *et al.*, [1] Shapiro [15], Nackman [6], Wong and Lu [20], Biederman [2], and Truve and Whitman [17]. Structural models have proved especially suitable in cases where recognition and interpretation are the main goals of modelling. It is likely that they will also find more extensive applications in other shape related fields, such as image compression or computer graphics. Advantages of these models are especially vivid when nonrigid shapes have to be recognized from their partial view (occlusion). Barring cases where shape can be adequately represented by a parametric form, such tasks are inherently difficult for recognizers based on a nonstructural description, such as a set of feature vectors or a neural net.

Despite their popularity, there has been relatively little work on two important aspects of structural models: the automatic construction of symbolic relations, and learning structural models from shape examples. Learning may

not be necessary if one attempts to model only rigid shapes, since then a relation, which represents the range of values of a numerical parameter such as distance, can be determined by error analysis. However, if a model is expected to describe a nonrigid shape, parameter ranges must be determined from data, either by a human or a machine.

One of the first structural learning methods was the work of Winston [19], which represented input data and the models in a semantic network. The applications of this method were restricted to block worlds and noise-free data. Connel and Brady [3] described a shape learning method, which allows disjunctive description of a class, and can accept some amount of noise. A method of learning structural descriptions in the form of an attributed random graph has been proposed by Wong and You [20], and they have shown an example of its application to synthetic scenes. The above methods rely on a predetermined set of symbolic types of parts and relations.

This paper presents an approach to structural modelling of shapes based on stochastic graphs. It extends and generalizes earlier work on planar shapes ([10], [11]). The presented approach consists of two parts. The first part automatically constructs symbolic properties and relations that are the elements of the structural description. The second part builds clusters of stochastic graphs that serve as models for shape classes. This part is almost free of user-settable parameters such as acceptance levels, or weights, which would limit its robustness. The resulting models can be used to recognize nonrigid shapes, interpret scenes composed of many overlapping shapes, and to identify shape parts. These methods are practical and fast, despite their use of graph matching. This is demonstrated by the current implementation in a 2-D vision system that recognizes hand gestures in real time.

1.1 Overview

The approach presented here attempts to capture similarity within a class of shapes represented as sets of primitive parts, by characterizing the range of values of displacement and rotation between parts, or a relative pose. In a rigid object these values are constant, but in a nonrigid object they vary between different object instances, and the range of variations can be different for different pairs of parts. The objective of learning is to describe a class of shapes using a structure of relations among parts, where each relation represents a specific range of displacement and rotation values.

Instead of analyzing each class separately, the entire set of training shapes is examined for natural groupings of part types and their relative poses in pairs of nearby parts. Each identified group represents a geometric relation, involving two parts and a distribution of parameters of their relative pose. A pair of primitive parts joined by one of the discovered relations is then treated as a higher-order part, and the grouping process repeats, giving rise to a hierarchy of parts and relations, similar to the hierarchy proposed by Marr and Nishihara [5]. While some true relations may be obscured by nearby groups or clouds

of random occurrences, and some identified relations can be coincidental, the grouping process usually discovers a large percentage of true relations.

The class membership of the training samples is examined only after constructing this hierarchy of relations. A shape is described as a graph, whose vertices are the relations. For each class of shapes the learning process constructs a model, which is a group of stochastic graphs. A stochastic graph is a probability model whose outcomes are graphs.

The key element of learning, recognition, and interpretation methods based on the graph representation is a fast heuristic for graph matching, that finds correspondences between vertices of two graphs.

2 Hierarchy of Parts and Relations

2.1 Primitive Parts

An instance of shape is initially represented as a collection of primitive parts. These parts do not have to be mutually exclusive, that is some parts may overlap. The collection of parts does not have to cover the entire shape. The literature on shape provides an abundance of algorithms and techniques for extracting parts from a two- or three-dimensional shape. Some methods partition the primary representation, e.g. a curve, surface, or volume into homogeneous regions, i.e. regions whose local properties are approximately constant. Other methods identify singularities, that is points or boundaries of the primary representation which are unique within their neighbourhood, such as edges, corners, local extrema of curvature, or critical points of a surface. Most of the published methods can be adapted to generate parts that satisfy the requirements of the approach described in this paper. This approach also allows one to mix together parts generated by different methods.

Definition 1. A primitive part p is a triple $[type, inv, var]$, where *type* is a symbol from a finite alphabet, *inv* and *var* are real valued vectors, such that if T is a rigid transformation, then $T(p) = [type, inv, T(var)]$

The *type* symbol specifies the format and interpretation for *inv* and *var* vectors, and it is used to distinguish parts generated by different extraction methods. The *inv* vector contains parameters that are invariant under rigid transformations of the coordinate system. These parameters are specific to the type of a part, which also remains constant under rigid transformations. Some parts, such as a single point, have no invariant parameters, that is the *inv* vector have dimension zero. The *var* vector consists of parameters that change with rigid transformations. This vector carries information related to the part's pose, that is part's position and orientation in space. In 2-D space pose consists of three values: two coordinates of position and the orientation angle. In 3-D space pose has six dimensions: three position coordinates and three angles. If the pose of a part can be determined based on the form of the part, then the *var* vector is the pose. However, for parts with symmetries pose cannot be determined completely, but it can be

restricted to a number of degrees of freedom (for continuous symmetries) or to a number of values (for discrete symmetries). The approach described here considers only parts with continuous symmetries such as line, or sphere. For such parts the *var* vector contains the maximal number of parameters that constrain the pose.

Definition 2. A representation by parts of shape S is a set of primitive parts $P(S) = \{p_1, p_2, \ldots\}$.

The function P in this definition stands for a method used to segment or to extract parts from S. The representation $P(S)$ should satisfy the following properties:

1. Invariance: Without presence of noise a rigid transformation should not change the representation. This means that for any rigid transformation T, $P(S) = \{p_1, p_2, \ldots\}$ implies $P(T(S)) = \{T(p_1), T(p_2), \ldots\}$.
2. Locality: A part description should not be affected by shape changes outside of the part's immediate neighbourhood.
3. Stability: The representation should not be significantly affected by random noise in the image. This means that if S_1 and S_2 are two noisy images of the same shape, then there exists a one-to-one mapping between large subsets of $P(S_1)$ and $P(S_2)$, where the corresponding parts have identical *type* symbols, and similar *inv* and *var* parameters.

2.2 Constructing New Parts

To build a structural description of shape one needs to identify symbolic relations among its parts. All pairwise relations might contain useful information, but examining all such relations in a large set of parts can be too costly. Therefore, direct pairwise relations are restricted only to pairs of nearby parts. Pairs of nearby parts are also used to construct new parts, called *composite* parts, that are treated just like the primitive parts. One can find binary relations between composite parts, and combine a pair of composite parts into a new composite part.

One can think of a composite part as a root of a binary tree. All non-leaf nodes of this tree are composite parts, and the leaves are primitive parts. This binary tree defines the composite part at its root, and determines the order of operations needed to construct it. Depth of this tree determines the *level* of the part at the root. The level of a primitive part is 0, a level-1 part is constructed from a pair of primitive parts, two level-1 parts give us a level-2 part, and so on. The current method combines parts with equal levels, so a composite part at level n has 2^n primitive parts as its leaves.

A composite part is represented the same way as a primitive part as $[type, inv, var]$. Composite parts are constructed bottom-up, one level at a time up to a preset maximal level. The construction method in the recognition mode is a bit different from the learning mode. In the learning mode the construction process repeats for each level the following sequence of steps:

1. Cluster the *inv* vectors of all parts at level k, separately for each part type, and assign a unique label to each cluster.
2. Assign to each part the label of its nearest cluster, or NIL if the nearest cluster is too far.
3. Terminate if k is equal to a preset maximal level.
4. Find pairs of neighbouring parts with non NIL labels, and construct $k+1$ level parts by applying a *composition* operation.

The recognition mode uses the same construction process, but step 1 (clustering) is omitted, and step 2 uses clusters computed in the learning mode. The key element of this process is the composition operation used in step 4 to form new parts.

2.3 Composition of Parts

The composition operation is applied to an ordered pair of parts. The result of this operation is a composite part represented as $[type, inv, var]$.

The type of the result of composition is a string obtained by concatenating the types of components, treated as strings. The *inv* and *var* vectors of the result are computed from the *var* vectors of the components. This operation depends on *symmetry types* of the component parts. A symmetry type specifies a group of rigid transformations which do not change the part's appearance. For example, an infinitely long cylinder has two continuous symmetries: rotation around, and translation along the axis. The approach proposed here is restricted to parts with continuous symmetries. Discrete symmetries, such as the symmetries of a cube may be treated in future extensions.

The symmetry type of a part is a function of the part type. The symmetry type of a primitive part has to be given defined for each primitive part type. Symmetry types of composite parts are determined by composition rules in Tables 2 and 4.

2.4 Parts in 3-D Space

Table 1 lists six symmetry types that are be used for three-dimensional parts. The first column shows a simple geometric form, that exhibits a given type of

Table 1. Part types in 3-D

PART	SYMBOL	*var* DIMENSION	SYMMETRIES	EXAMPLES
Point	P	3	3R	sphere
Line	L	4	1R+1T	cylinder
Plane	S	3	1R+1T	plate
Point on line	PL	5	1R	cone, paraboloid
Line on plane	LS	5	1T	infinite edge, ridge
Frame	F	6	None	edge segment

symmetry. The symmetry transformations are shown in fourth column using a symbolic notation, where nR which means rotation about n axes, and nT means an n-dimensional translation. All linear forms in the first column of Table 1 are assumed to be oriented, that is Line, Plane normal, and Frame axis are lines with sense.

Symmetry type of the composition of any two of types is always among the six types listed in Table 1. Table 2 shows the symmetry type and the number of invariant parameters of the result of composition, for all combinations of operand symmetry types.

Table 2. Parts composition in 3-D

COMPONENTS	RESULT	# OF INVARIANTS
P,P	PL	1
P,L	F	1
P,S	PL	1
P,PL	F	2
P,LS	F	2
P,F	F	3
L,L	F	2
L,S	F	1
L,PL,	F	3
L,LS	F	3
L,F	F	4
S,S	LS	1
S,PL	F	2
S,LS	F	2
S,F	F	3
PL,PL	F	4
PL,LS	F	4
PL,F	F	5
LS,LS	F	4
LS,F	F	5
F,F	F	6

In most cases, it is possible to define composition operation in several ways, using different formulas for the *inv* and *var* parts of the result. These alternative formulations are equivalent, that is from the composition result for one of the alternatives the result of any other formulation can be computed. The composition operation is defined below for four cases from Table 2. The following notation is used: $P(x)$ is the point specified by x, for symmetry types P and PL. $L(x)$ is the line and $r(x)$ the unit vector for symmetry types L, PL, and LS.

Case 1 Operands x, y of type P.
Result z of type PL:

var: $P(x)$ is the midpoint between $P(x)$ and $P(y)$, that is $\frac{1}{2}(P(x)+P(y))$. $L(z)$ is the line defined by $P(x)$ and $P(y)$.
inv: Distance between $P(x)$ and $P(y)$.

Case 2 Operands: x type P, y type PL.
Result z of type F:
var: The frame origin is the midpoint between $P(x)$ and $P(y)$. The orientation of the first axis is given by the unit vector

$$\mathbf{u} = \frac{P(x) - P(y)}{|P(x) - P(y)|} .$$

The remaining axes are obtained by a Gramm-Schmidt orthonormalization: Let $\mathbf{V} = \mathbf{r(y)} - (\mathbf{r(y)} \cdot \mathbf{u})\mathbf{u}$ then a unit vector $(v) = \mathbf{V}/|\mathbf{V}|$ defines the second axis; the third axis is $\mathbf{w} = \mathbf{u} \times \mathbf{v}$.
inv: Distance between $P(x)$ and $L(y)$, and signed distance from $P(y)$ to the projection of $P(x)$ on $L(y)$.

Case 3 Operands: x, y type PL.
Result z of type F:
var: Frame is computed as in Case 2, using $\mathbf{r(x)} + \mathbf{r(y)}$ instead of $\mathbf{r(y)}$ to find the second axis.
inv: Distance between $P(x)$ and $P(y)$, angle between $\mathbf{r(x)}$ and $\mathbf{u}$, angle between $\mathbf{r(y)}$ and $\mathbf{u}$ ($\mathbf{u}$ defined as in Case 2), and angle between the two planes defined by a line through $P(x)$ and $P(y)$ and vectors $\mathbf{r(x)}$, and $\mathbf{r(y)}$, or angle between normals to these planes.

Case 4 Operands: x, y type F:
Result z of type F:
var: Frame is computed as in Case 1, using the sum of the unit vectors from all six axes of x and y instead of $\mathbf{r(y)}$ to find the second axis. If this sum is 0 then any subset of the axes can be used.
inv: Six parameters of rigid transformation from x to y.

2.5 Parts in 2-D Space

Symmetry types in 2-D form a subset of symmetry types in 3-D. This subset consists of three symmetry types listed in Table 3. The composition operation

Table 3. Part types in 2-D

PART	SYMBOL	*var* DIMENSION	SYMMETRIES	EXAMPLES
Point	P	2	1R	circle
Line	L	2	1T	edge
Point on line	PL	3	None	edge segment

Table 4. Parts composition in 2-D

COMPONENTS	RESULT	# OF INVARIANTS
P,P	PL	1
P,L	PL	1
P,PL	PL	2
L,L	PL	1
L,PL,	PL	2
PL,PL	PL	3

in 2-D is defined in Table 4.

Each case of 2-D composition can be derived from a corresponding 3-D case, as shown in the example below.

Case 4 Operands: x, y type PL in 2-D.
Result z of type PL:
var: $P(z)$ and $L(z)$ correspond to the origin and the first axis in Case 3 above.
inv: Distance between $P(x)$ and $P(y)$, and the two angles between vectors $r(x)$ and $r(y)$, and the line through $P(x)$ and $P(y)$.

2.6 Ordering Pairs of Parts

The procedure for matching the structural descriptions requires the arguments of binary relations to be ordered. To order a pair of parts (x, y) the following three-step procedure is used. If x and y have different types then they are ordered according to a lexicographic order of their types. If the types are identical, but the parts have different labels, then they are ordered by their labels. If the labels are the same then an ordering function $f(x, y)$ is used. An ordering function must have the following properties:

1. There is a partial order relation $>$ defined on the range of f.
2. Generally, $f(x, y) \neq f(y, x)$.
3. For any rigid transformation T, $f(x, y) > f(y, x)$ implies that $f(Tx, Ty) > f(Ty, Tx)$. Of course this is satisfied if $f(x, y) = f(Tx, Ty)$. This property ensures that the order specified by f is invariant under rigid transformations.

With the aid of f one orders parts x and y as (x, y) if $f(x, y) > f(y, x)$, and vice versa. If neither $f(x, y) > f(y, x)$ nor $f(y, x) > f(x, y)$, then parts cannot be ordered.

An example of such a function for 2-D parts x and y and symmetry types L, or PL, is the signed angle between part orientations $r(x)$ and $r(y)$.

An example for 3-D parts with symmetry type PL is the first angle invariant in Case 3, Sect. 2.4.

Using an ordering function presents a minor problem. A natural cluster that intersects the hypersurface $f(x, y) = f(y, x)$ will be split into two clusters. This

is an undesirable feature, since such a split is purely artificial. Split clusters can be merged using the following consolidation procedure.

For each cluster C_i form an inverted cluster $-C_i$, then find a cluster C' in a set $C_i+1, C_i+2, \ldots$, which is nearest to $-C_i$. An inverted cluster $-C$ is a cluster formed by reversing the order of composition of elements of C. In most cases such an operation is a function of cluster parameters, so it does not require reprocessing the cluster elements. If $-C_i$ and C' are sufficiently close then merge $-C_i$ into C', provide a pointer from C_i to C', and delete all clusters that point to C_i. In addition, delete any cluster that is close to its own inverse.

If the above procedure is used then the labelling step is modified as follows: If a composite part is assigned to a cluster a that points to a cluster b, then the part is inverted and receives the label of the cluster b.

2.7 Extracting Relations

The label of a primitive part symbolically describes an invariant property (unary relation), such as size or curvature. The label of a composite part P describes a binary relation between its two component parts (children). It also describes a 4th-order relation among the part's grandchildren (if any), 8th-order relation among the great grandchildren, and so on, until it finally describes a 2^n-ary relation over a set of primitive parts.

After constructing the composite parts up to a preset level, one retains only their labels, and the parent-child links. The resulting structure is a graph with labeled vertices, that are grouped into layers according to their depth. The leaves of the graph represent the primitive parts; other vertices represent composite parts. This graph contains all the information about the shape, that is used for recognition and interpretation.

3 Graph Representation

A directed graph is a set of vertices V and a set of directed edges E. Beginning with a directed graph, a special case of a hypergraph called a *layered graph* is defined recursively. A vertex $v \in V$ will is called a level-0 vertex, or a *leaf*. Two level 0 vertices connected by an edge will be called a level-1 vertex. Given a set of directed edges between pairs of level-1 vertices, one acan similarly define a level 2 vertex, and generally define a level-$n+1$ vertex as an ordered pair (v_1, v_2), where v_1 and v_2 are level-n vertices. Their order corresponds to the direction of the edge between v_1 and v_2. Such a structure is called a layered graph.

A layered graph can be represented by a directed acyclic graph, whose vertices correspond to the vertices of the layered graph, and whose edges show the hierarchical dependency among vertices. The terms *parent, child, ancestor, descendant,* and *leaf*, are used for the vertices of a layered graph in the same sense as for a tree, except that a graph vertex can have multiple parents.

A layered graph has the following properties:

1. Each non-leaf vertex v has exactly two ordered *children*, distinguished as $left(v)$ and $right(v)$.
2. For every vertex v, each path between v and a leaf vertex has the same length. This length is called the *level* of v.
3. Two vertices can have no more than one common parent.

In addition, it is assumed that each vertex v of a layered graph has a label $L(v)$, which is a symbol from a finite alphabet, and there is a separate alphabet for each level. Such a graph is a special case of an attributed hypergraph: the leaves of the layered graph are vertices of a hypergraph, while the higher-level vertices are the hyperedges.

A group of layered graphs that are not identical can be described using a probability model whose outcome is a layered graph, or a *stochastic layered graph*. This model is defined just like a layered graph, except that each of its vertices is associated with a probability distribution over a set of labels, rather than a single label. A probability of finding the label l at a vertex v will be denoted by $p(l|v)$.

Further, a layered graph and a layered stochastic graph are called respectively a graph and a stochastic graph, providing it causes no confusion.

If T is a mapping from the vertices of a stochastic graph M onto the vertices of a graph H, one assigns to each mapped vertex of M the label of the corresponding vertex of H. The probability of H, given M and T, $P(H|T,M)$ is the probability of this set of assignments. Assuming independently distributed vertex labels, $P(H|T,M)$ is simply the product of the probabilities $p(L(Tv)|v)$, where v is a vertex of M, and Tv is the vertex of H assigned to v by T.

4 Graph-based Recognition and Learning

4.1 Minimal Representation Criterion

The minimal representation criterion ([14], [9]) was introduced to guide inference of models in cases when the maximum likelihood fails. Its formulation has been inspired by the pioneering work of Solomonoff [16]. This criterion seeks a minimal length program generating observed data X, among a given set of programs for a Turing machine. The Turing machine is treated here as a general purpose decoding device, where a program (with a null input) is a code for the output sequence. Mapping a family of probability distributions to a subset of programs and seeking a distribution corresponding to a minimal program results in an inference rule that select a distribution P that minimizes

$$C(P) - \log_2 P(X) \ ,$$

where $C(P)$ is the number of bits needed to specify the probability model P, or the *cost* of P. This criterion is equivalent to the minimal description length or MDL principle of Rissanen [7], and closely related to the information measure of Wallace and Boulton [18], but it was derived independently.

4.2 Matching Graphs

A key mechanism of a learning system based on a graph representation is a method for establishing a mapping between elements of two graphs, or graph matching. Based on the minimal representation criterion, the task of matching a stochastic graph M to a graph H, is formulated as a construction of a mapping between vertices of M and H, that maximally simplifies description of H, when H is represented relative to M. This formulation provides a natural decision rule for accepting a match, which does not rely on an arbitrary threshold. Intuitively, it works as follows: If part of M fits to a part of H, then representing H relative to the matching part of M should cost less than its *default* representation, independent of any model. However, the total cost of representing H with the aid of M includes an overhead associated with specifying the mapping. If the part of H represented by M is too small, the saving may not cover the overhead cost, and it will be cheaper to use the default representation, i.e., to reject the match.

A default representation of a graph is constructed as follows:

1. Specify N, the number of leaves in H.
2. Provide a list of leaf labels, ordered by leaf index.
3. Order pairs of leaves (e.g. lexicographically), and specify the label of the common parent for each pair (NIL if none).
4. For level 1, order pairs of non-null vertices. For every pair, specify the label of the common parent (or NIL).
5. Repeat the last step for each higher level, up to a level with all null vertices.

The cost of this representation $C(H)$ is

$$C(H) = C(N) - \sum \log_2 p(L(v)) \ ,$$

where $p(l)$ is a prior probability of the label l. The first term is the cost of representing N, the second term is the cost of specifying vertex labels (this value is within 1 bit from the length of Shannon block encoding).

A graph H can be described relative to a model M, given a one-to-one mapping $T : V_1 \rightarrow V_2$, where $V_1 \subseteq V(M)$ and $V_2 \subseteq V(H)$. To represent H under this mapping one uses the probability defined by M for labels of the mapped vertices in V_2, instead of their prior probability. The cost of representing H, given M and T, denoted $C(H|T, M)$ is:

$$C(H|T, M) = C(H) - \sum_{v \in V_1} \log_2 \frac{p(L(Tv)|v)}{p(L(Tv))} \ ,$$

where the second term is the sum of bits saved over V_2. The cost of describing H relative to M, $C(H, T|M)$ includes the overhead for specifying T:

$$C(H, T|M) = C(H|T, M) + C(T) \ ,$$

where $C(T)$ is the cost of T. To find this cost, notice that mapping T is completely determined from its submapping T' restricted to the sets of leaves $F(M)$

and $F(H)$. T' is a mapping from $F_1 \subseteq F(M)$ onto $F_2 \subseteq F(H)$, such that $T'(v) = T(v)$ if v is a leaf. The fact that T' determines T is implied by the Property 3 of the layered graph. Therefore, one needs only to specify the leaf mapping T', and $C(T) = C(T')$. If T' maps k out of n leaves of one graph, to k out of m leaves of the other, then knowing n and m, T' can be encoded using

$$C(T') = \log_2 \min(n, m) + log_2 \frac{n!}{(n-k)!k!} + log_2 \frac{m!}{(m-k)!}$$

bits. The first term is the cost of specifying $k > 0$, the second term is the cost of selecting k leaves of the first graph, and the third term is the cost of selecting k leaves of the second graph and specifying their permutation. This representation essentially assigns equal prior probability to each value $k > 0$. An alternative is to assume equal prior probability for each mapping, which corresponds to a constant cost

$$C(T') = \log_2 \sum_{k=1}^{\min(n,m)} \frac{n!m!}{(n-k)!(m-k)!k!} .$$

The first representation is used, since it is less penalizing to small values of k. The number of bits saved with respect to the default representation is

$$\begin{aligned} Q(H, T|M) &= C(H) - C(H, T|M) \\ &= \sum_{v \in V_1} \log_2 \frac{p(L(Tv)|v)}{p(L(Tv))} - C(T) . \end{aligned}$$

If this value is positive, the representation based on M and T should be used instead of a default representation, since it is less expensive.

The task of matching a graph H with a model M is formulated as a problem of constructing a mapping T, that maximizes the value of $Q(H, T|M)$. It is easy to show that the problem of finding a maximal isomorphic subgraph, which is NP-complete, is a special case of the above problem, so the maximization of $Q(H, T|M)$ is NP-hard. Therefore, one must compromise and accept a quasi-optimal solution obtained by heuristic search. The fast graph-matching heuristic described below is an improved version of the method from [10].

The procedure for matching H with M consists of two steps, called *map* and *refine*. *Map* finds an initial mapping T that maximizes an upper bound of Q, then *refine* iteratively edits T until Q reaches a local maximum.

The function *map* uses contextual similarity as a basis for the leaf assignment. One can think of a vertex v of a layered graph as a relation $l(f_1, f_2, \ldots, f_k)$, where l is the label of v, and $f_1, f_2, \ldots$ are recursively ordered leaf descendents of v. Similarly, a vertex v of a stochastic layered graph is associated with a relation $l(f_1, f_2, \ldots, f_k)$, which has the probability $p(l|v)$. The context of a leaf v is the set of relations associated with its ancestors. The context is described by a *support* of the leaf v.

Definition 3. A support $SPS(v)$ of a leaf v of a layered graph, is a set of pairs $\{(R, K)\}$, where R describes a relation by its label and the position of the vertex

v in its argument list, and K is the number of occurrences of R among ancestors of v. A support of a leaf v of a stochastic layered graph, is a set of pairs $\{(R, G)\}$, where R describes a relation, as above, and $G = (G_1, G_2, \ldots)$ is a list containing a value $2^{-level(u)} \log \frac{p(l|u)}{p(l)}$ for each occurrence of R in an ancestor u of v, for which $\frac{p(l|u)}{p(l)} > 1$. This list is sorted in a decreasing order. The factor $2^{-level(u)}$ divides equally the support of u among its leaf descendents.

Definition 4. A *similarity* $S(u, v)$ between a leaf u of a graph and a leaf v of a stochastic graph is

$$S(u,v) = \sum_{\substack{(R,K)\in SPS(u)\\(R,G)\in SPS(v)}} \sum_{i=0}^{\min(K,|G|)} G_i \ . \tag{1}$$

Similarity $S(u, v)$ provides an upper bound on an increase of Q resulting from adding an assignment (u, v) to T.

Map finds a mapping that maximizes the sum of similarities between mapped leaves, by solving an assignment problem. This mapping is iteratively improved by *refine*, which deletes or adds one assignment in each iteration, seeking a maximal increase of Q. The iteration stops when no single addition or deletion can further improve Q.

4.3 Recognition and Interpretation

Given a graph H and a library of stochastic graph models $ML = \{M_0, M_1, M_2, \ldots, M_n\}$, where M_0 is an empty graph, one might want to select one model M which is in some way nearest to H. This task is called *recognition*. Since the mapping of H to a single model may not exhaust all the vertices of H, different subgraphs of H could be mapped to different models. These mappings taken together form an *interpretation* of H. Interpretation can be considered an attempt to explain H using the models from the library as primitives. Since both recognition and interpretation can lead to a compressed representation of H, both tasks are formulated as problems of minimizing the cost of representation.

Recognition problem: Given a model library ML and a graph H, find a model m in ML, and a mapping T from m to H, that maximizes

$$Q(H, T, m) = Q(H, T|m) + \log_2 P(m) \ , \tag{2}$$

where $P(m)$ is the prior probability of model m. The empty graph M_0 is always associated with a null mapping, so that $Q(H, T|M_0) = 0$ for any H.

A brute-force recognition method seeks a maximum of this expression by matching H with each model in ML. The computational cost of this method

grows approximately linearly with the size of the model library. To speed up the search a *preclassification* method can be used based on a similarity measure, related to the measure defined in (1) that provides an upper bound on $Q(H, T, m)$. One first computes the similarity between H and each model in the library, then matches the models, and computes $Q(H, T, m)$ in order of their decreasing similarity to H. This search terminates upon reaching a model with a similarity value less than the highest match score $Q(H, T, m)$ found.

Interpretation problem: Given a model library ML and a graph H, construct a sequence of models $m_1, m_2, \ldots, m_k$, where $m_i \in M_1, M_2, \ldots, M_n$ for $i = 1, 2, \ldots, k-1$ and $m_k = M_0$, and a sequence of associated mappings $T_1, T_2, \ldots, T_k$, to maximize the value of

$$\sum_{i=1}^{k} Q(H_i, T_i | m_i) + \log_2 P(m_i) \ ,$$

where the graphs H_i, $i = 1, 2, \ldots, k$ are constructed recursively as follows:

- $H_1 = H$.
- Form H_{n+1} from H_n by removing all vertices of H_n mapped to m_n by T_n.

A heuristic method for graph interpretation solves the recognition problem, first for $H_1 = H$, then for H_2, etc., until some graph H_k is recognized as M_0. Since interpretation requires multiple recognitions, the preclassification method described above can be used to speed it up.

4.4 Learning Graph Models

Models used for recognition and interpretation are learned from a training set of shape examples represented as layered graphs, and their class assignments. The objective is to describe shape classes using stochastic graphs. This task can be approached in two ways. First, a single stochastic graph can be forced to represent all examples within a class. However, such a model may not perform well if the graphs in a class are not sufficiently similar. In such a case, attempt can be made to divide the class into subclasses consisting of mutually similar graphs, and to use a separate stochastic graph to describe each subclass. This task is similar to classical clustering problems, except the objects of grouping are graphs, so it is referred to as graph clustering. The remainder of this section describes a method for fitting a single stochastic graph to a set of graph examples, and two methods of graph clustering.

4.5 Forced Fit

This learning procedure incrementally constructs a stochastic graph from a sequence of graph examples. Given a sequence of graphs $H_1, H_2, \ldots, H_k$, the first graph H_1 is converted to a stochastic graph M_1, by assigning a probability value to each vertex label, using the formula in (3) below. The following graphs are then

used to update the model, which results in a sequence of models $M_1, M_2, \ldots, M_k$, using the following match-and-merge operation.

A graph H_{n+1} is matched with model M_n giving a mapping T. Since T maps a subset of leaves of M_n to a subset of leaves of H_{n+1}, generally some leaves of M_n and H_{n+1} remain unmapped. The mapping T and the graph M_n are extended to T' and M' in the following way: Initially M' is set to M_n. For each unmapped leaf f of H_{n+1}, a new leaf f' is added to M', and a pair (f', f) is added to T. Then new higher-level vertices are added to M', such that each vertex of M' has a corresponding vertex in H_{n+1}. The mapping of these vertices is determined by recursively following child-to-parent links. Finally, vertex-label statistics for the vertices of M' are updated, based on the labels of their matches in H_{n+1}, and new label probabilities are computed using the Bayes estimator:

$$P(i) = \frac{n(i) + 1}{n + k} \quad , \tag{3}$$

where $n(i)$ is the number of occurrences of label i among n observations, and k is the number of different labels. The final form of graph M' is then taken as M_{n+1}. The last graph M_k obtained this way is used as the model of a class.

4.6 Graph Clustering

Graph clustering is a task of dividing a set of graphs into groups, such that similar graphs are grouped together. Previously proposed methods for clustering graphs ([4], [20]) share the requirement for user-specified parameters or other subjective devices to control the number of clusters, or cluster separation. Here the clustering of graphs is formulated in a non ad hoc manner, as an optimization task void of any free parameters. The optimality measure is based on the criterion of minimal representation.

Graph clustering problem: Given a sequence of graphs $I = H_1, H_2, \ldots, H_N$, construct a set of stochastic models, $ML = M_1, M_2, \ldots, M_k$, such that when graphs in I are represented relative to models in ML, the total cost of representing ML and I is minimum.

The set of all graphs that are represented relative to the same model is called a graph cluster. To avoid the need for encoding label distributions for a stochastic graphs, a model is represented predictively using a small set of cluster members called *internal*. The remaining members of the cluster are called *external*. This approach is related to Rissanen's predictive minimal description length principle [8], but it does not depend on the ordering of data elements.

A model M_i, $i > 0$, will be represented by a *generating* sequence of n_i graphs from I, $I_i = H_{i1}, H_{i2}, \ldots, H_{in_i}$, (internal members). The forced-fit algorithm applied to this sequence results in a sequence of models $M_{i1}, M_{i2}, \ldots, M_{in_i}$. The last model in this sequence is then used as the model M_i of a cluster i. A default representation is used for the first graph in the generating sequence, then remaining graphs are encoded predictively, that is a graph H_{ij+1} is represented

relative to a model M_{ij}. In addition, for each graph in I_i, the position in I must be encoded to preserve the initial ordering of I.

The total representation for ML and I consists of the following parts:

1. The length N of I, and the number of clusters K.
2. The predictive encoding of a generating sequence for each cluster.
3. The position of each of $n = \sum n_i$ internal members in I.
4. For each external member H of some cluster, a cluster index i, and the representation of H relative to M_i.
5. For each *free* (not a member of any cluster) graph H, the index of the group of free graphs, and the default representation of H.

An incremental clustering method begins by forming a cluster from the first element of I, then it assigns each successive element to its nearest cluster, or creates a new cluster containing this element, based on the value of (2). The nearest cluster is the one which gives maximal value to this expression; a new cluster is formed if this value is not positive. After examining the last element of I, all singleton clusters are eliminated, and their members become free graphs.

While the above method is simple and fast, its results depend on the order of data, and it tends to find only well separated clusters. A second graph clustering method uses an agglomerative procedure, that does not dependent on the order of samples. While this method is more expensive than the previous one, it usually results in cleaner clusters and lower representation cost.

Beginning with a default representation for each graph in I, and 0 clusters, the program repeatedly applies one of the following moves, until there is a single cluster containing all elements of I:

1. Form a new cluster from a pair of free graphs.
2. Assign a free graph to one of the clusters, as an external member.
3. Merge two clusters by assigning members of the first cluster to the second cluster as external members, and removing the first cluster.

Each iteration selects the move, which results in a minimal representation cost after the move, even if this cost increases. When a new member is added to a cluster, the program attempts to reduce its cost by appending external members to the generating sequence. The final result is the best among the examined sets of clusters.

5 Final Remarks

The graph learning and recognition methods described above have been implemented in a Common Lisp system. The relation constructor has been programmed so far only for the 2-D case, using one type of a primitive part: a local extremum of curvature of the boundary of a planar shape. This part has one invariant parameter: the curvature. The *var* vector contains the position of the extremal point, and the direction of the curve normal at this point, that is its symmetry type is PL. The symmetry type of all the composite parts derived

from these primitives is also PL. The system has been trained to recognize three postures of a flexible toy and it achieves over 99% recognition accuracy on isolated cases. It also recognizes occluded objects, but its accuracy drops with the number of invisible parts.

The 2-D recognition method was also implemented in a real-time vision system GEST. The first version [13] used a network of 30 processors, and it processed 4 video frames/s. The current version [12] uses 3 processors (SUN 4 and Intel i860) and its throughput is 10 - 20 frames/sec, with latency of 100 - 150 ms. GEST has been trained to recognize 10 hand gestures. It is used as an input device that allows the user to interact with 3-D graphics programs using hand signs.

A first 3-D implementation will use corner-like parts of symmetry type PL, that are extracted by a structural stereo system.

References

1. Barrow, H. G., Ambler, A. P., Burstall, R. M. (1972). Some techniques for recognizing structure in pictures. In: Watanabe, S. (ed.), Frontiers of Pattern Recognition, Academic Press, New York, pp. 1–29.
2. Biederman, I. (1987). Matching of image edges to object memory, Proc. First Int. Conf. on Computer Vision, London, England: IEEE Comp. Soc. Press, pp. 384–392.
3. Connel, J. H., Brady, M. (1985). Learning shape descriptions, Proc. 9th Int. Conf. on Artificial Intelligence, Los Angeles, CA, Morgan Kaufmann, pp. 922–925.
4. Levinson, R. (1985). A self-organizing retrieval system for graphs. Ph.D. Thesis, University of Texas, Austin, TX.
5. Marr, D. Nishihara, H. K. (1978). Representation and recognition of spatial organization of three-dimensional structures, Proc. R. Soc. B 200, pp. 269–294.
6. Nackman, L. R. (1984). Two-dimensional critical point configuration graphs, IEEE Trans. on Pattern Analysis and Machine Intelligence, 6, pp. 442–449.
7. Rissanen, J. (1978). Modeling by shortest data description, Automatica 14, pp. 465–471.
8. Rissanen, J. (1987). Stochastic complexity. J. R. Stat. Soc. 49, pp. 223–239.
9. Segen, J. (1980). Pattern-directed signal analysis, Ph.D. Thesis, Carnegie-Mellon University, Pittsburgh, PA.
10. Segen, J. (1989). Model Learning and Recognition of Nonrigid Shapes, Proc. Conf. on Computer Vision and Pattern Recognition, San Diego, pp. 597–602.
11. Segen, J. (1990). Graph clustering and model learning by data compression, Proc. 7th Int. Conf. on Machine Learning, Austin, Texas, pp. 93–101.
12. Segen, J. (1993). GEST: A learning computer vision system that recognizes hand gestures. In: Michalski, R.S., Tecuci, G. (eds.) Machine Learning IV, Morgan Kaufmann.
13. Segen, J., Dana, K. (1991). Parallel symbolic recognition of deformable shapes. In: Burkhardt, H., Nuevo, Y., Simon, J.C. (eds.), From Pixels to Features II, North-Holland.
14. Segen, J., Sanderson, A. C. (1979). A minimal representation criterion for clustering. Proc. 12th Annual Symposium on Comp. Science and Statistics, Univ. of Waterloo, Canada.

15. Shapiro, L. G. (1980). A structural model of shape. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2, pp. 111–126.
16. Solomonoff, R. J. (1964). A formal theory of inductive inference I & II, Information and Control, 7, pp. 1–22 & 224–254.
17. Truve, S., Richards, W. (1987). From Waltz to Winston (via the connection table), Proc. First Int. Conf. on Computer Vision, London, England, IEEE Comp. Soc. Press, pp. 393–404.
18. Wallace, C. S. Boulton, D. M. (1968). An information measure for classification, Comp. Journal 11(2), pp. 185–194.
19. Winston, P. H. (1975). Learning structural descriptions from examples. In: Winston, P.H. (ed.), The Psychology of Computer Vision, Chapter 5, McGraw Hill, New York.
20. Wong, A. K. C., You, M. (1985). Entropy and distance of random graphs with application to structural pattern recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 7, pp. 599–609.
21. Wong, A. K. C., Lu, S. W. (1985). Recognition and knowledge synthesis of 3-D object images, Proc. Conf. on Computer Vision and Pattern Recognition, San Francisco, pp. 162–166.

Hierarchical Shape Analysis in Grey-level Images*

Annick Montanvert[1], *Peter Meer*[2], *and Pascal Bertolino*[1]

[1] Equipe RFMQ-TIMC-IMAG, Université Joseph Fourier Grenoble, Cermo BP 53X, 38041 Grenoble Cedex, France
[2] Department of Electrical and Computer Engineering, Rutgers University, P.O. Box 909, Piscataway, NJ 08855-0909, USA

Abstract. The problem of parallel, bottom-up construction of hierarchical structures adapted to the content of an input image is addressed. This principle is essential in image analysis and understanding. For each level of the structure, which is related to a resolution level, the region adjacency graph of a segmentation of the input image is defined. All segmentation and resolution reduction operations are local, and therefore a complete input-dependent hierarchy can be built in O[log(image-size)]. The region adjacency graph contraction is achieved by extracting a maximal independent set (MIS) of the graph. A new parallel probabilistic method for MIS computation is proposed and discussed. The representation uncertainty introduced by the probabilistic component of the hierarchy construction is reduced through consensus among an ensemble of outputs obtained from the same input image.

Keywords: graph representation, shape description, hierarchical processing, image segmentation, maximal independent set, probabilistic algorithms, consensus approach, stochastic symmetry breaking.

1 Introduction

Multiresolution image analysis is a widely used tool in computer vision. A hierarchical stack of decreasing resolution representations is derived recursively from the input image, as are defined image pyramids. Each resolution corresponds to a level of the hierarchical structure, starting at level 0 and ending at level n. The hierarchy is built bottom-up and in parallel, the values at level $l+1$ (called *parents*) are computed from the values at level l (called *children*). The reduction procedure assures the construction of a hierarchy in $O[\log(image-size)]$. Multiresolution image analysis exploits the hierarchical architecture and achieves fast global analysis by local processes defined at low-resolution levels. The traditional hierarchies have a regular structure and are known as image pyramids. In an image pyramid the resolution is frequently reduced fourfold between adjacent

* Peter Meer would like to acknowledge the support of the National Science Foundation under Grant IRI-9210861, and the support of the Rutgers Research Council.

levels. Numerous applications of image pyramids to feature extraction and image analysis have been developed [21, 6]. Image pyramids, however, do not have the right architecture for image analysis. This resolution reduction process is constrained by the rigidity of the pyramid and produces artifacts. Contour tracing with image pyramids is an example of a task where such problems appear and is discussed in [16]. The artifacts of rigidly structured hierarchical processing were recognized a long time ago [23] and algorithms in which the parent-children link weights and the neighbourhood size are iteratively changed were proposed (e.g., [8, 2]). However, the resources reallocation approach cannot eliminate completely the undesired effects [3].

An optimal approach requires that the structure of a hierarchy is adapted to the content of the input image. This can be achieved only if the structure of the different levels is described by graphs. Each level of the hierarchy is a graph adapted to the input; the whole hierarchy is a stack of graphs recursively reduced from the graph representing the structure of the input. In shape analysis the homogeneous regions must first be delineated. To segment an image into homogeneous regions, the region adjacency graph must be used when building the hierarchy. In Sect. 2 the graph formulation is given, and it is shown that the construction of the lower resolution representation is equivalent to finding the maximal independent set (MIS) of the adjacency graph of the current level. A fast probabilistic algorithm for MIS computation is described. In Sect. 3 the local operations adapting the structure of the hierarchy to the content of the input image are discussed. The application to image segmentation is presented in Sect. 4. In Sect. 5 a consensus-based methodology is presented which distinguishes reliable features from sporadically detected features.

2 Resolution Reduction with MIS

The hierarchy to be constructed is a stack of undirected graphs $(G[l])_{l=0\ldots n}$. The index l is called the level of G; $G[l] = (V[l], E[l])$, where $V[l]$ is the set of vertices and $E[l]$ is the set of edges (each edge connecting two vertices). $V[l+1]$ is an MIS of $V[l]$ if the subgraph induced by the vertices of $V[l+1]$ does not have any edge of $E[l]$. The graph $G[0]$ is defined on the *8-connected* square sampling grid of the input image. Each vertex at level l is linked with a connected subset of vertices at level $l-1$. Tracing top-down these parent-children links, each vertex at level l delineates a connected subset of vertices at level 0 called *receptive field.* When delineation of homogeneous regions is of interest, the hierarchy must be constructed such that the receptive fields correspond to homogeneous regions at the input. The edges at level l represent the adjacency relations between these receptive fields. Thus $G[l]$ is the region adjacency graph of the segmentation of the input as it is obtained at the resolution of level l.

The definition of $G[l+1]$, the reduced resolution version of $G[l]$ is a graph contraction problem. Three steps must be solved:

- extraction of $V[l+1]$, a subset of $V[l]$;
- creation of the parent-children links;

– computation of $E[l+1]$.

This section presents the extraction of $V[l+1]$.

To ensure efficient processing, the stack of reduced resolution representations must be derived recursively from the input through a parallel process.

The vertices retained for level $l+1$ are called the survivors of the resolution reduction process; they form a subset of the vertices of level l. The value associated with a survivor (parent) is the reduced resolution representation of a neighbourhood on level l (its children) and therefore of the concatenation of the receptive fields of the children. Parallel processing and significant resolution reduction between consecutive levels are required. In other words, the selected survivors for $V[l+1]$ have to be spread out on $G[l]$ uniformly while they are significantly less numerous than in $V[l]$. This kind of problem is well known in graph theory, in terms of dominating set and maximal independent set [5].

$V[l+1]$ is defined as a maximal independent set of $G[l]$, and is characterized by the two following properties:

G1: Two adjacent vertices in $G[l]$ cannot both be selected for $V[l+1]$.
G2: Any vertex in $V[l]$ not selected for $V[l+1]$ is connected to a vertex which is.

The constraint G2 ensures that the allocation of non-survivors to a survivor (definition of the parent-children relation) can be performed in one step. The MIS of a graph is not unique and neither is its cardinality. In Fig. 1 two examples of MIS of the same adjacency graph $G[l]$ are shown. The receptive fields of the vertices are also drawn.

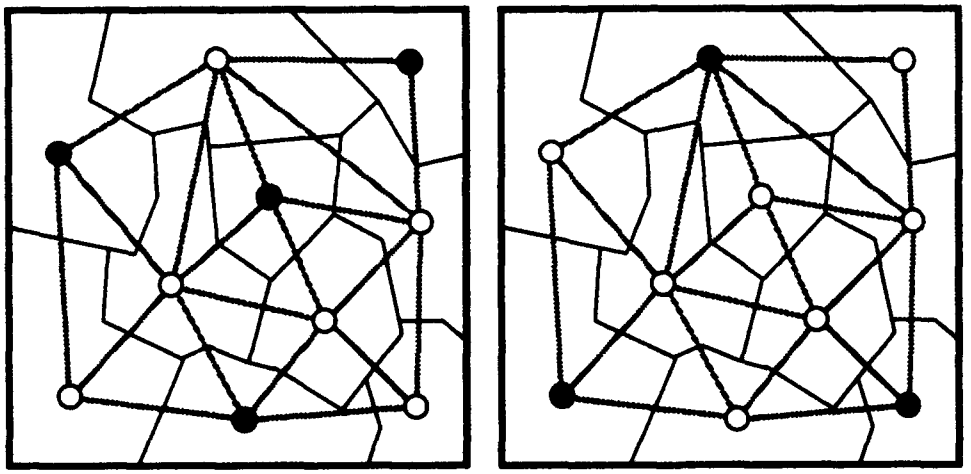


Fig. 1. Two examples of maximal independent set for the same adjacency graph. The solid circles are the selected vertices. The tessellation of the input by the receptive fields is also shown.

In order to build the hierarchy in parallel and preserve pyramidal concepts, parallel methods described in the literature were studied (e.g., [1, 12, 7]). All the parallel MIS algorithms are based on the same principle.

MIS Algorithm

```
begin
    V[l+1] := ∅; V := V[l];
    while V ≠ ∅ do
    begin
        select IS, an independent set of a subgraph induced by V on G[l];
        add IS to V[l+1];
        remove IS and all the vertices adjacent on G[l] to IS from V;
    end
end
```

The specific parallel MIS algorithms differ by the adopted model of computation and the way the intermediate independent set IS is selected. The Monte Carlo technique proposed by Luby [12] is typical.

Selection by Trial and Error

```
begin
    In parallel, for all v in V[l]
        compute d(v), the degree of vertex v;
        if d(v) = 0, set rand(v) = 1
        else set rand(v) = 1 with probability 1/2d(v);
        if rand(v) = 1, add vertex v to IS;

    In parallel, for all edges (v1, v2) ∈ E[l]
        if v1 ∈ IS and v2 ∈ IS, remove from IS the vi, (i = 1, 2)
            with min d(vi);
end
```

The Trial and Error selection procedure can be implemented in parallel on $O[|E[l]|.d_{max}]$ processors, where $|E[l]|$ is the number of edges and d_{max} is the maximum degree (that is the maximum number of edges connected to a vertex) in $G[l]$. The number of necessary iterations is $O[\log|V[l]|]$, proportional with the logarithm of the number of vertices in the graph [1].

The probability of a vertex to be selected into the intermediate independent set is reciprocal to its degree. The selection procedure is biased toward vertices with a smaller degree to increase the number of retained vertices, that is, to obtain an MIS of better quality. The degree-based selection also reduces the probability that two adjacent vertices are chosen simultaneously. When this happens, the algorithm must backtrack and remove one of the vertices before proceeding to the next iteration. Due to backtracking, the Trial and Error selection procedure is not the fastest possible parallel method for obtaining an MIS. In our approach, a different selection procedure in which all the selected vertices belong to $V[l+1]$ is used. The new procedure is based on a probabilistic

algorithm for symmetry breaking in parallel environments [13, 14].

Selection by Local Extrema

```
begin
    In parallel, for all v in V[l]
        choose rand(v) from the (0,1) uniform distribution;
        if rand(v) > rand(v_i) for all i such that (v, v_i) ∈ E[l]
            add vertex v to IS;
end
```

The vertices selected are those associated with local extrema and therefore two adjacent vertices can never be chosen in the intermediate independent set IS. No backtracking is necessary and the Local Extrema method is faster than the Trial and Error method. Only strict local extrema are considered; ties due to the finite machine precision are broken at the next iteration. Note that the procedure implicitly takes into account the degree of a vertex. The higher the degree, the less probable it is that the vertex has the highest outcome of $rand(v)$ in its neighbourhood. The number of processors needed to run the parallel MIS algorithm with Local Extrema selection is $V[l]$, the number of vertices in the graph. Thus implementation of this procedure requires a minimum number of processors.

The performances of the two selection procedures were compared by simulations. The first three levels of a hierarchy were built recursively. The *8-connected* graph of a 64×64 sampling lattice was used as the adjacency graph $G[0]$ for the base of the hierarchy. The extracted MIS $V[1]$ induces the adjacency graph of the next level, $G[1]$. (The procedure to define the new adjacency relations that is, $E[1]$, will be discussed in Sect. 3.) From $G[1]$ its MIS $V[2]$ was extracted and the adjacency graph $G[2]$ of the second level obtained. From $G[2]$ the set of vertices retained for the third level $V[3]$ was extracted. Fifty hierarchies were built with each selection procedure.

The rate of convergence of the MIS algorithms was monitored through the number of iterations required to extract an MIS. Each iteration corresponds to a cycle of the loop in the MIS algorithm. Note that due to backtracking an iteration with the Trial and Error procedure requires more steps than with the Local Extrema method. The quality of the obtained MIS was measured by the number of vertices retained in the set. The statistics computed: mean, standard deviation, minimum and maximum value of the range, are given in Tables 1–6.

Both selection procedures generate MIS of similar sizes, the Trial and Error procedure yielding slightly larger sets. The Trial and Error procedure requires about $\log(number - of - vertices)$ iterations as predicted by the theory [1]. The Local Extrema procedure selects the surviving vertices based on their immediate neighbourhood on the graph, and its rate of convergence depends only weakly on the number of vertices in the graph. The number of iterations is also more stable than that of the Trial and Error procedure.

Table 1. Trial and Error procedure, first level

	Mean	Std.dev.	Min	Max
No. iterations	13.28	2.01	11	21
No. vertices	789.3	7.87	775	806

Table 2. Local Extrema procedure, first level

	Mean	Std.dev.	Min	Max
No. iterations	3.98	0.31	3	5
No. vertices	779.7	8.34	759	797

Table 3. Trial and Error procedure, second level

	Mean	Std.dev.	Min	Max
No. iterations	9.44	1.70	6	14
No. vertices	194.3	4.77	183	205

Table 4. Local Extrema procedure, second level

	Mean	Std.dev.	Min	Max
No. iterations	3.32	0.47	3	4
No. vertices	188.5	5.30	176	199

Table 5. Trial and Error procedure, third level

	Mean	Std.dev.	Min	Max
No. iterations	7.92	2.24	4	15
No. vertices	54.8	3.07	47	62

Table 6. Local Extrema procedure, third level

	Mean	Std.dev.	Min	Max
No. iterations	2.94	0.37	2	4
No. vertices	52.4	3.39	43	59

3 Adaptive Hierarchical Structures

To define the hierarchy completely, parent-children links must be created, as well as $E[l+1]$.

The graph $G[l+1]$ cannot be defined without creating the parent-children links from which $E[l+1]$ can be obtained. By *using* the MIS for the graph contraction, a non-survivor vertex is always connected with an edge to a survivor. The parent-children links can thus be established in parallel with local processes. After the non-survivors are allocated, the edges in $E[l+1]$ are simply defined by the adjacency between the receptive field of $V[l+1]$ on $G[l]$. Note that all the necessary information is available locally. In Fig. 3 the adjacency graph representing the structure of the next level in the hierarchy is shown for the example in Fig. 2. The receptive field of a parent is the concatenation of the receptive fields of its children.

The procedures described earlier do not take into account the information available at the bottom of the hierarchy. To adapt the structure of the hierarchy to the content of the input:

- $G[l]$, the adjacency graph of the current level, must be transformed into a similarity graph $S[l]$, and
- the parent-children links must depend on the values assigned to the vertices, that is, the representations of the receptive fields.

The similarity graph is defined in parallel using only local processes.

Let $g(v)$ be the value associated with vertex v (for instance the average grey-level). This value is computed from the values of the children of the vertex. Extraction of the similarity graph $S[l]$ from the adjacency graph $G[l]$ means definition of arcs (directed edges) between two adjacent vertices whenever a

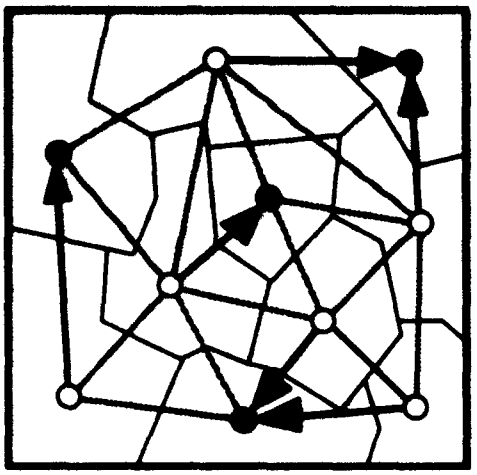


Fig. 2. Allocation of non-survivors.

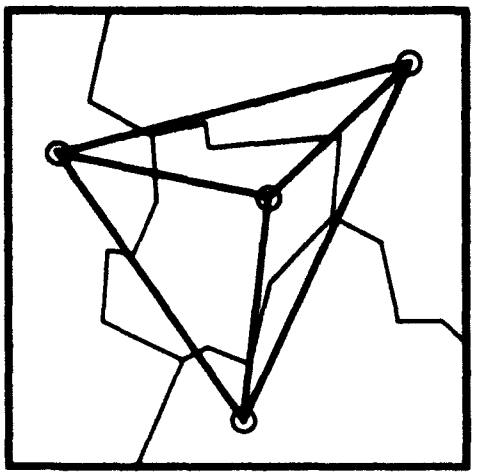


Fig. 3. Reduced resolution representation.

distance measure of the values associated with the two vertices is less than a predefined threshold. The similarity graph is obtained with the following parallel procedure.

Similarity Graph Definition

```
begin
    In parallel, for all v in V[l]
        while (v, v_i) ∈ E[l] do
        begin
            if ||g(v_i) − g(v)||_H ≤ T(v), set λ_i = 1;
            if ||g(v_i) − g(v)||_H > T(v), set λ_i = 0;
            if λ_i = 1, retain the arc (v, v_i) in S[l];
        end
end
```

The Similarity Graph Definition procedure allocates a class membership to every vertex connected to the vertex v. The procedure is executed in parallel for all the vertices and thus can have at most d_{max} steps. The distance measure $\|.\|_H$ is contingent upon the employed criterion H for homogeneity. All the arcs starting from v are now weighted with the binary variable λ_i and the neighbourhood of v is dichotomized into two classes. If the distance between the values associated with the two vertices v and v_i is less than a threshold, it is concluded that the receptive fields of the two vertices can be fused, that is, under the homogeneity criterion they belong to the same region. The decision threshold $T(v)$ is neighbourhood specific and in general $\|g(v_i) - g(v)\|_H \leq T(v)$ does not imply that $\|g(v) - g(v_i)\|_H \leq T(v_i)$, since the two thresholds are computed based on neighbourhoods that only partially overlap: an arc from v to v_i may not be reciprocated by an arc from v_i to v (Fig. 4). Thus, the class memberships define a directed similarity graph $S[l]$.

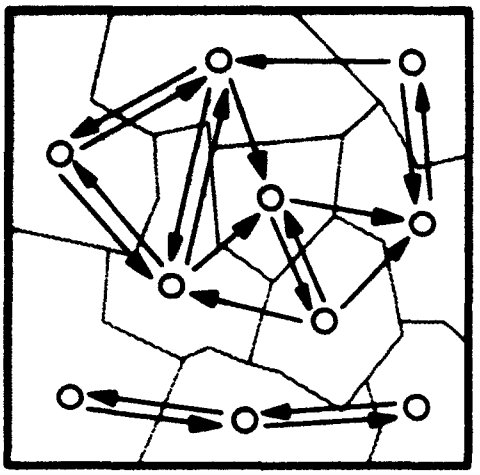


Fig. 4. Local class memberships as arcs.

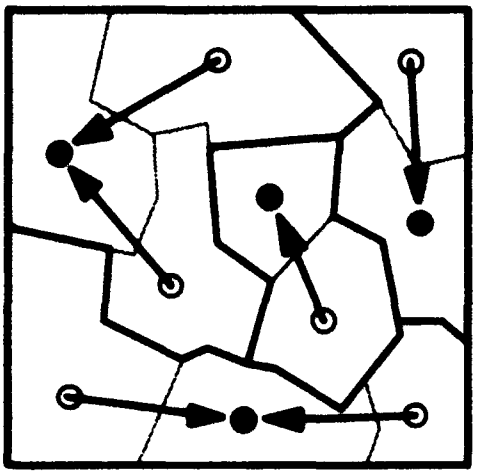


Fig. 5. Resolution reduction and allocation of non-survivors.

To take into account the information available at the input, the selection condition in the Local Extrema procedure has to be modified to: if $rand(v) > \lambda_i rand(v_i)$ for all i as $(v, v_i) \in E[l]$, add vertex v to I. The random value of the vertex v is compared only with the random values of those vertices in its neighbourhood which were classified as being in the same class. These vertices are connected to v by an arc. The graph contraction condition G1 is satisfied only for the similarity graph but not for the adjacency graph of level l. Two survivors may now be neighbours on $G[l]$ if they are not connected by an arc in the similarity graph. $V[l]$ is a dominating set of the directed graph $S[l]$ (it verifies properties (G1) and (G2)).

Note that a non-survivor vertex of $V[l]$ can be connected in $G[l]$ to vertices retained for $V[l+1]$ (survivors). The vertex (child) will be allocated to the vertex in $V[l+1]$ (parent) from which it is at minimum distance. After the survivors are allocated, the edges of the adjacency graph $G[l+1]$ can be obtained using the graph $G[l]$ and the receptive fields of $V[l+1]$ delineated on it. In Fig. 5 the resolution reduction and the allocation of non-survivors are shown for the example in Fig. 4.

4 Image Segmentation

The presented adaptive hierarchy construction method starts from an input image and recursively generates the coarser representations. The value $g(v)$ assigned to each vertex which is the average grey-level of its receptive fields, can also be computed recursively. For every homogeneous region of the input a separate hierarchy is built having approximately $\log(region\_size)$ levels. The resolution reduction ratio is no longer fixed at four as in the case of rigidly structured image pyramids, although it is close to it. The receptive field of a vertex at the

apex of a hierarchy, the root of that hierarchy, represents the entire homogeneous region. The root adjacency graph is defined by the apexes of the different hierarchies. This graph describes the spatial relations among the delineated homogeneous regions. Hierarchies built over different homogeneous regions have different heights. The value associated with an earlier detected region may become within threshold distance of a neighbour on the adjacency graph of a higher level. Thus the region may disappear at subsequent levels, its receptive field being fused into a larger region. In our segmentation applications, this procedure gave better results than removing the region from subsequent adjacency graphs or making it an automatic survivor. For tasks with more complex homogeneity criteria, however, a careful analysis is needed to assess which of the region preservation strategies is adapted.

Additional local information can also be incorporated into the construction of a hierarchy. Jolion and Montanvert [9] proposed a variant in which the chances of a vertex becoming a survivor are improved if its neighbourhood is more homogeneous. Montanvert and Bertolino [18] defined a quality measure for the contours separating adjacent receptive fields. This measure is recursively derived from discontinuity information available at the input. The Similarity Graph Definition procedure then uses both contour quality and grey-level difference.

Fig. 6. Original image.

As an example, the segmentation of the 64 × 64 grey-level image shown in Fig. 4 was performed using these adaptive hierarchies. The receptive field of each root vertex delineates a constant valued region at the input. The results for two different adaptive hierarchies (constructed with different outcomes of the random variables in the selection procedure) are shown in Figs 7–10. The receptive fields are coloured with their average grey-level. The number of vertices in the root adjacency graph were 112 and 94 respectively. The large number of

Fig. 7. Output of the first segmentation: Average grey-level coloured receptive fields.

Fig. 8. Output of the first segmentation: Randomly coloured receptive fields.

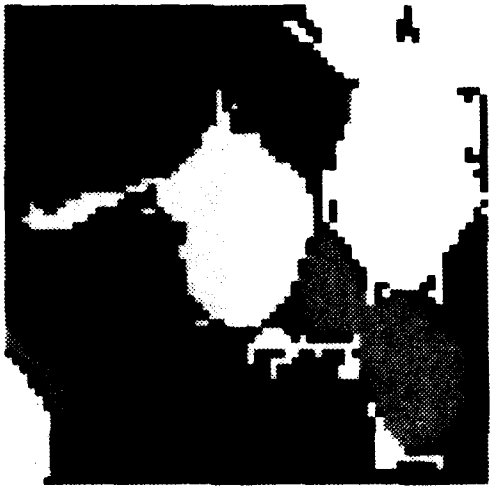


Fig. 9. Output of the 2nd segmentation: Average grey-level coloured receptive fields.

Fig. 10. Output of the 2nd segmentation: Randomly coloured receptive fields.

vertices in the root adjacency graph is due to the many vertices having small receptive fields. The quality of the segmentations is given in Figs 8 and 10: the receptive fields (homogeneous regions) are coloured with arbitrary grey-levels.

High contrast features (like the black blobs) have similar shapes in both segmentations but regions with blurred boundaries present significant changes. The variations are due to the probabilistic survivor selection procedure. A different set of survivors will have slightly different neighbourhoods and thus different $g(v)$ values. The cumulative effect of these small differences may lead to significant changes in the segmented image. The structural uncertainty of the hierarchies

is inherently embedded into the method and will always yield shape uncertainty for the delineated regions. Employing robust local operators to dichotomize a neighbourhood cannot yield significant improvement since the size of the neighbourhood is too small for any estimator. Graph theoretical analysis of hierarchical structures obtained by recursive application of the MIS algorithm is given in [11]. Montanvert *et al.* [17] discuss the practical issues in constructing the adaptive hierarchies and the influence of the probabilistic selection procedure on the outcome of a task.

The shape uncertainty can be reduced, however, when several output images are combined together into a consensus image as proposed in Sect. 5.

5 Performance Improvement by Consensus

Assume that several output images are derived from the same input image. For example, several segmentations are obtained as in Figs 7–10 for the same input image. When the output images differ only in the uncertainty of the representations, a consensus approach can reduce this uncertainty. Consensus means

Fig. 11. Consensus for segmenting the original image combining 1 output image.

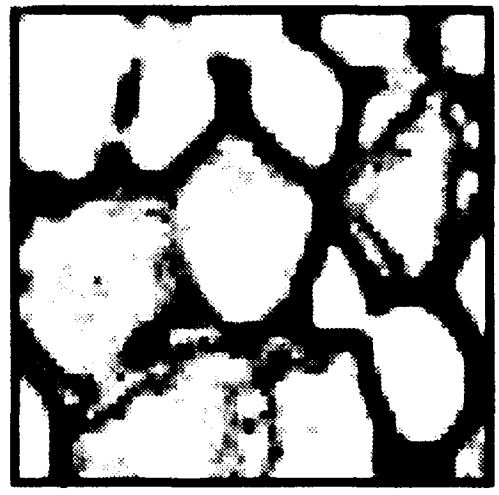


Fig. 12. Consensus combining 5 output images.

"general (all or most) agreement in opinion". A formal description of the consensus paradigm is given in [15]. The necessary conditions for successful application of the consensus approach are:

C1: Representations in different output images are outcomes drawn from the same distribution.
C2: The mode of the distribution is the optimal representation.

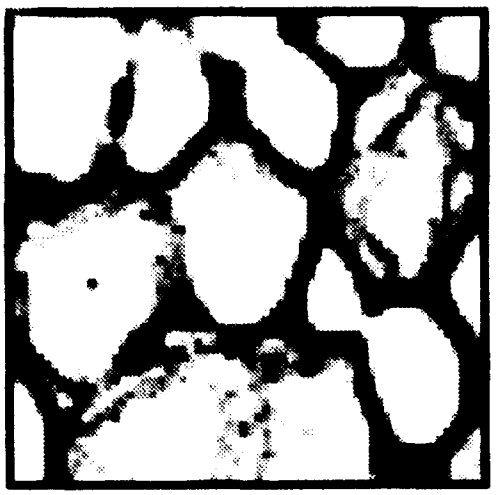


Fig. 13. Consensus combining 10 output images.

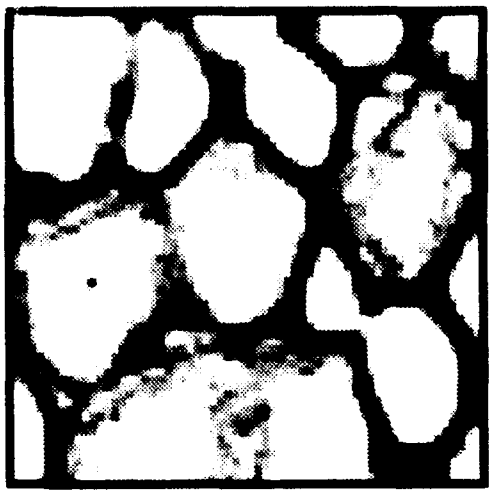


Fig. 14. Consensus combining 30 output images.

For example, a low contrast region extracted as homogeneous from the image in Fig. 4 has different shapes in the two segmentations shown in Figs 7 and 9. The two delineations of the region are regarded as the outcomes of a random process. The mode of the distribution governing the process then defines the best possible delineation of the region, given the input and the employed homogeneity criterion (piecewise constant). *The condition C1 implies that the most probable* delineation of the homogeneous region can be found by combining many output images.

To compare the different segmentations, the output images (segmented images) are transformed first. For every pixel in an output image the number of its *8-connected* neighbours having the same label (that is which are in the same region) are counted. The transformed output image thus retains only the boundaries of the delineated regions. An example of a transformed output image is shown in Fig. 11. The image is scaled between 0 and 255. Higher pixel values correspond to more homogeneous neighbourhoods. When several transformed output images are combined the scores are cumulated pixelwise in a consensus image.

In Figs 12–14 consensus images taking into account an increasing number of outputs are shown. As more segmentations are used, regions not present in the first image (Fig. 11) are revealed. The region boundaries at the bottom-right are a good example. The consensus process practically converges after combining 10 outputs and no significant changes appear later. In the consensus image every pixel carries a measure for the homogeneity of its neighbourhood.

The consensus paradigm is a powerful approach. Using the simplest local homogeneity criterion (difference of grey-levels) all the important features (discontinuities relative to a piecewise constant model) of a grey-level image are delineated. The probabilistic nature of the adaptive hierarchy construction was

exploited to reduce the uncertainty of the output and to eliminate the artifacts of the segmentation. The consensus approach was also applied successfully to eliminate the block effects in a pyramid-based image smoothing [20], and to reduce the importance of cluster position on pyramidal delineation of compact dot patterns in an image [22]. Consensus can become useful when constructing a hierarchy to obtain robust local decision thresholds for the Similarity Graph Definition procedure. The distribution of local thresholds is analysed at a higher level of the hierarchy for several different constructions. The most probable threshold is then broadcast top-down to all the vertices in the receptive field.

6 Conclusion

A system to perform shape delineation for grey-level images has been described. All the procedures are executed in parallel and in $O[\log(image - size)]$. Some of the principles can be compared with some parallel techniques for image segmentation [4]. The approach was also used for curve pyramids and contour approximations [10, 19].

Several outputs (segmented images) are derived from the same input and are combined to yield a more accurate result. This representation can then be used to guide further processing in a top-down fashion. The method of recursively building the hierarchy is a general one. Any (scalar or vector) value computed from a receptive field can be associated with its vertex as long as the computation can be performed recursively. Thus, extraction of shape descriptors for the delineated regions is immediate and is obtained in logarithmic time.

The principles behind the described technique recall biological systems where a large uncertainty of individual representations (neuronal signals) is compensated by a high degree of parallelism and redundancy as well as by efferent pathways.

References

1. Alon, N., Babai, L., Itai, A. (1986). A fast and simple randomized parallel algorithm for the maximum independent set problem, J. of Algorithms 7, pp. 567–583.
2. Baronti, S., Casini, A., Lotti, F., Favaro, L., Roberto, V. (1990). Variable pyramid structure for image segmentation, Computer Vision Graphics Image Processing 49, pp. 346–356.
3. Bister, M., Cornelius, J., Rosenfeld, A. (1990). A critical view of pyramid segmentation algorithms, Pattern Recognition Letters 11, pp. 605–617
4. Chantemargue, F., Popovic, M., Canals, R., Bonton, P. (1991). Parallelization of the merging step of the region segmentation method, Proc. 7th Scandinavian Conf. on Image Analysis, Aalborg (Denmark), August 13–16, 1991, pp. 933–940.
5. Christofides, N. (1975). Graph theory: An algorithmic approach, Academic Press, London.
6. Dyer, C.R. (1987). Multiscale image understanding. In: Uhr, L. (ed.), Parallel Computer Vision, Academic Press, Boston, pp. 171–213.

7. Goldberg, M., Spencer, T. (1987). A new parallel algorithm for the maximal independent set problem, Proc. 28th Symposium on Foundations of Computer Science, pp. 160–165.
8. Hong, T.H., Rosenfeld, A. (1984). Compact region extraction using weighted pixel linking in a pyramid, IEEE Trans. Pattern Anal. Machine Intell. 6, pp. 222–229.
9. Jolion, J.M., Montanvert, A. (1992). The adaptive pyramid: A framework for 2D image analysis, CVGIP: Image Understanding 55 (3), pp. 339–348.
10. Kropatsch, W. (1987). Curve representation in multiple resolutions, Pattern Recognition Letters, 6, pp. 315–322.
11. Kropatsch, W.G., Montanvert, A. (1991). Irregular versus regular pyramid structures. In: Eckhardt, U., Hubler, A., Nagel, W. , Werner, G. (eds), Research in Informatics 4, Proc. 5th Workshop on Geometrical Problems of Image Processing, Georgenthal, Germany, March 1991, Akademie Verlag, Berlin, pp. 11–15.
12. Luby, M. (1986). A simple parallel algorithm for the maximal independent set problem, SIAM J. of Computing 15, pp. 1036–1053.
13. Meer, P. (1989). Stochastic image pyramids, Computer Vision Graphics Image Processing 45, pp. 269–294.
14. Meer, P., Connelly, S. (1989). A fast parallel method for synthesis of random patterns, Pattern Recognition 22, pp. 189–204.
15. Meer, P., Mintz, D., Montanvert, A., Rosenfeld, A. (1990). Consensus vision, Proc. AAAI-90 Workshop on Qualitative Vision, Boston, Mass., July 1990, pp. 111–115.
16. Meer, P., Sher, C.A., Rosenfeld, A. (1990). The chain-pyramid. Hierarchical processing of contours, IEEE Trans. Pattern Anal. Machine Intell. 12, pp. 363–376.
17. Montanvert, A., Meer, P., Rosenfeld, A. (1991). Hierarchical image analysis using irregular tessellations, IEEE Trans. Pattern Anal. Machine Intell. 13, pp. 307–316.
18. Montanvert, A., Bertolino, P. (1992). Irregular pyramids for parallel image segmentation, Bischof, H., Kropatsch, W.G., (eds.), Proc. 16th AGM Meeting, Vienna, Austria, May 5–9, Oldenbourg Verlag 1992, pp. 13–35.
19. Nacken, P., Toet, A. (1993). Candidate groupings for bottom-up segmentation, this volume, pp. 549–558.
20. Park, R.-H., Meer, P. (1991). Edge-preserving artifact-free smoothing with image pyramids, Pattern Recognition Letters 12, pp. 467–475.
21. Rosenfeld, A. (1984). Multiresolution Image Processing and Analysis, Springer Verlag, Berlin.
22. Sher, C.A., Rosenfeld, A. (1991). Pyramid cluster detection and delineation by consensus, Pattern Recognition Letters 12, pp. 477-482.
23. Tanimoto, S. (1976). Pictorial feature distortion in a pyramid, Computer Graphics Image Processing 5, pp. 333–352.

Irregular Curve Pyramids*

Walter G. Kropatsch and Dieter Willersinn

Technical University of Vienna, Institute for Automation 183/2, Department for Pattern Recognition and Image Processing, Treitlstr. 3, A-1040 Wien, Austria
Email: krw@prip.tuwien.ac.at

Abstract. Regular $2 \times 2/2$ curve pyramids are hierarchical symbolic representations of curves that can be constructed and processed in logarithmic time. The rigidity of the regular structure causes drawbacks that were overcome by extending the concept to irregular pyramids. These have a structure that adapts to the image data by deriving control information from curve relations. The algorithm that builds the irregular curve pyramid goes far beyond merely solving the shift variance problem. It allows the definition of rules for the control that image data holds on the structure of the pyramid. These rules can be used to reflect the importance of local elements of shape with respect to a given application.

Keywords: discrete curve representation, dual graph, curve relation, decimation, irregular tesselation, bottom-up construction.

1 Introduction

Digital images represent objects from the real world in a discrete structure. Such a structure consists of a large set of atomic cells and certain neighbour-relations. Through analog-to-discrete mapping the properties of real objects are transformed into relations among the atomic cells of the discrete representation.

The shape of an object may be described either by its boundary or by its axis, a curve in the middle of the shape. Curves are connected sets of points. Since the connectivity of curves is their most important property and since this property is represented only implicitly in grey-level or binary images we have introduced the scheme of curve relations [3]. It is based on the idea that a curve crossing the region of an atomic cell (e.g. a square) intersects the cell boundary exactly twice. To check the connectivity of the curve segments distributed in the image cells, it is sufficient to verify that the intersection points of adjacent cells match. The model can be further simplified by identifying only the side of the cell on which the intersection point is located and not memorizing its precise coordinates. In this way a curve crossing an atomic cell creates a binary curve relation between two sides of the cell.

* This work was supported by the Austrian Science Foundation under grant P 8785.

In Sect. 2 we summarize the properties of the regular $2\times2/2$ curve pyramid [4] which can be constructed from an image, the content of which is represented by binary curve relations. In this structure long thin objects can easily be extracted. The major drawback of the $2 \times 2/2$ curve pyramid is the position variance of the object representation (Sect. 2.1). Section 3 summarizes the principles of irregular pyramids and introduces the idea of representing curves in irregular sampling grids. The algorithm to construct the stochastic curve pyramid is presented in Sect. 4. It is based on a stochastic process which selects vertices that survive during a decimation. In Sect. 5 we show how the construction of the irregular curve pyramid can be controlled by image data.

2 The $2 \times 2/2$ Curve Pyramid

In image processing the classical arrangement of cells is a square grid. The $2\times2/2$ pyramid stacks square grids of successively lower resolution in a regular way: cells of overlapping 2×2 reduction windows are SPLIT by their diagonals, deriving curve relations in the resulting triangles. Groups of four triangles are then merged into one cell at the next level, building the transitive closure of the curve relations stored in the triangles (MERGE) [5]. Figure 1 gives an example of how the representation of a curve is simplified when a cell of the new pyramid level is constructed.

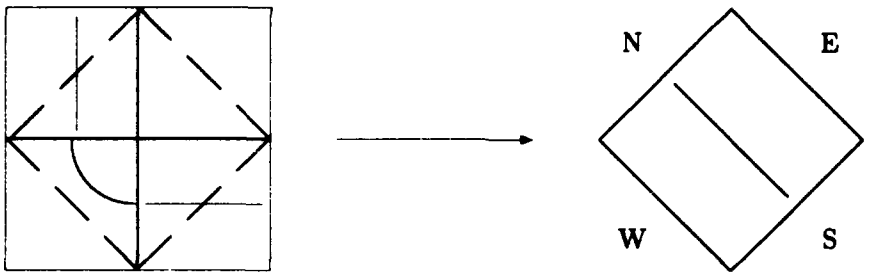


Fig. 1. Reduction in the $2 \times 2/2$ curve pyramid

The grid is rotated by 45 degrees, and the number of cells is divided by 2 from level to level. Figure 2 illustrates the structure of such a stack.

A $2 \times 2/2$ pyramid recursively built with operations SPLIT and MERGE has been shown to possess the *length reduction property* [6]. It states that short curves remain at the lower (i.e. high- resolution) levels of this "curve pyramid" whereas long curves survive up to higher (i.e. low-resolution) levels. The connectivity is preserved in the bottom-up building process.

Two applications have demonstrated the efficiency of the concept: structural filtering of short curves produced by noise [7, 9, 10] and preserving the contrast of boundaries in the concept of dual pyramids [16, 8].

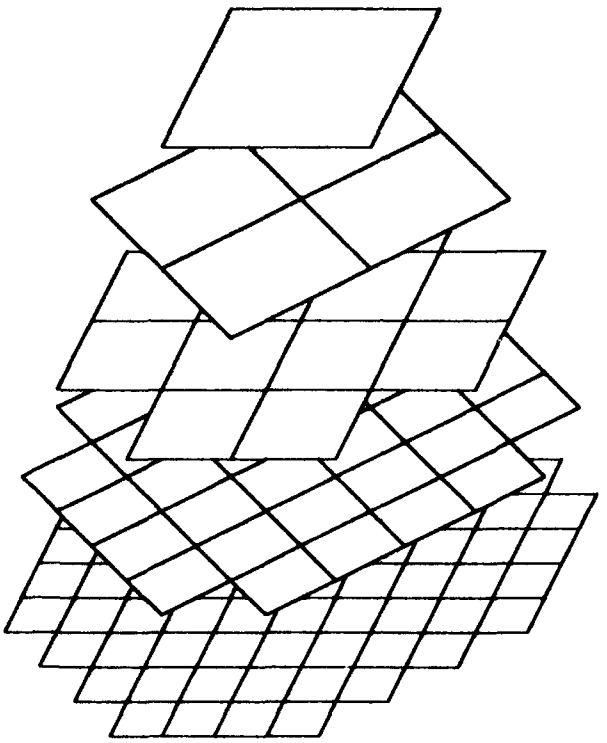


Fig. 2. Structure of a $2 \times 2/2$ pyramid

2.1 Drawbacks

Unfortunately the $2 \times 2/2$ curve pyramid has also some drawbacks related to its rigid structure. These drawbacks are similar in nature to those described by Bister *et al.* [1]. The first problem appears as an exception in the "length reduction" theorem and concerns *isolated blobs*, i.e. short curves surrounding a vertex. Such a blob survives the reduction until the vertex it surrounds disappears. This may be as high as the apex of the pyramid and depends heavily on the position of the vertex in the pyramid. Blobs that may lead to an exception of the length reduction property of the curve pyramid cover only a 2×2 window. These can be removed during reduction by an additional local filter.

The second drawback of the $2 \times 2/2$ curve pyramid involves the representation of *parallel lines*. It is not so easy to remedy. Consider two parallel lines at 45 degrees and at a distance of one pixel (Fig. 3b). The corresponding curve relations form two series of opposite left-right turns. If the vertices (•) between the two lines survive (Fig. 3c), both lines will be represented at the next two lower resolutions yielding the same situation as before. However if the vertices between the parallel lines do not survive, the two lines are merged into one (double) line (Fig. 3a).

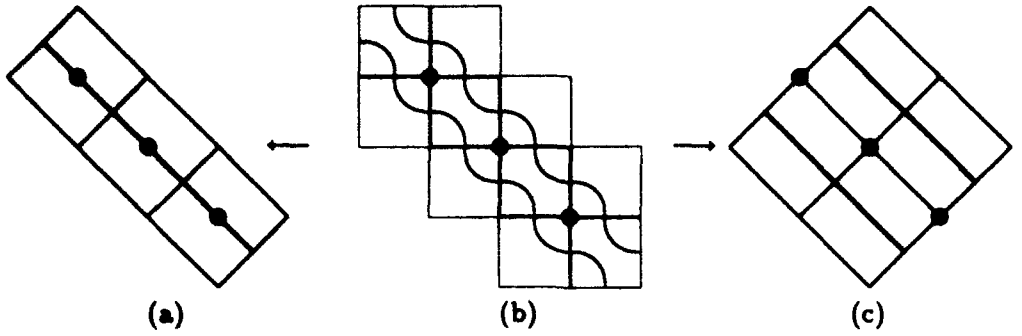


Fig. 3. Parallel lines may or may not be merged

3 Irregular Pyramids

Pyramidal structures that are flexible enough to adapt their structure during construction have been introduced by Meer [13]. The levels of the stochastic pyramid are not regular (square) grids but general graph structures. The bottom-up construction first assigns random numbers to all vertices and then selects vertices with a local maximum of this variable (surviving vertices). Then non-surviving vertices are assigned to the survivors. All non-survivors that are assigned to a surviving vertex form the *receptive field* of this survivor. These receptive fields determine the neighbour-relations of the reduced graph. Jolion and Montanvert have related the selection of survivors to the image data in the adaptive pyramid [2, 15, 14].

This scheme overcomes many of the drawbacks of rigid pyramids because the structure of the pyramid reflects the structure of the data. Since these pyramids are characterized by the fact that their neighbour-relations are not regular they are called irregular pyramids.

3.1 Curves in Irregular Pyramids

Combining the curve relations with the adaptive pyramid seems to solve the problems with isolated blobs and parallel lines. However, if we decimate the cells with the curve relations, the number of sides of the larger cells may grow from level to level and hence also the space required to store the curve relations.

This problem can be solved using the notion of the dual graphs. Let an edge $e = (v_1, v_2) \in E$ connect two vertices $v_1, v_2 \in V$ of the neighbourhood graph $G(V, E)$. Then the dual of G, $\overline{G}$, consists of faces F and of sides $\overline{E}$, $\overline{G} = (F, \overline{E})$. A side $\bar{e} \in \overline{E} \subset F \times F$ separates two adjacent faces and, at the same time, it corresponds one-to-one to an edge $e \in E \subset V \times V$ which connects two neighbouring vertices. Sides and edges are differentiated here for formal reasons; their graphical representation (e.g. in figures) is the same line segment. We proved that the degree of faces in the dual graphs of irregular pyramids does not increase [11]. Hence if we decimate the vertices of $G(V, E)$ and not the atomic cells (= faces) and if we further store the reduced curve relations in faces formed

by the surviving vertices and their surviving neighbours, the number of sides of a face does not increase. Starting with a square grid we need to extend our curve relations only to triangular cells because cells with two straight sides are not possible. The sides of the reduced grid connect two neighbouring survivors by a straight line. If this line crosses grid cells of the lower level an extended SPLIT operation must be applied to find the intersections with the new grid side. The MERGE operation is also generalized and derives the reduced curve relations by building the transitive closure of all cells covered by the new grid cell. With these modifications an irregular curve pyramid can be built. It has the freedom of selecting which vertices should survive. A simple scoring can control this process and can force vertices not to survive in the cases that caused the problems in the $2 \times 2/2$ curve pyramid.

4 Constructing the Irregular Curve Pyramid

The bottom-up construction of a pyramid assumes image data to be given at the highest resolution (= lowest pyramid level 0).

A level $n+1$ of an irregular pyramid is built on level n by decimating a neighbourhood graph $G_n(V_n, E_n)$. Decimation produces a subset of surviving vertices $V_{n+1} \subset V_n$, the *parents*, and by the use of receptive fields in G_n also the neighbourhood structure E_{n+1} of $G_{n+1}(V_{n+1}, E_{n+1})$ [13].

In the curve pyramid, the image data are curve relations expressing the fact that a curve connects two sides of a cell (face) in a graph $\overline{G_n}$ dual to G_n. We therefore assume that a pair of dual graphs $G_n(V_n, E_n)$ and $\overline{G_n}(F_n, \overline{E_n})$ and the curve relations $CR_n \subset \overline{E_n} \times (\overline{E_n} \cup \{stop\})$ for every face in F_n are given. $(\bar{e}, stop)$ denotes a curve end. The recursive construction involves three independent steps:

1. G_n is decimated yielding G_{n+1}.
2. The structure implied by decimation is propagated to the dual graphs $\overline{G_n}$ and $\overline{G_{n+1}}$.
3. Curve relations CR_n are reduced from $\overline{G_n}$ to $\overline{G_{n+1}}$ yielding $CR_{n+1} \subset \overline{E_{n+1}} \times (\overline{E_{n+1}} \cup \{stop\})$.

Decimation creates relations between graphs G_n and G_{n+1}. We propagate these relations to the dual graphs $\overline{G_n}$ and $\overline{G_{n+1}}$ by means of labels. If we assign to every vertex $v \in V_{n+1}$ a unique label $l(v)$ and if the parents in G_{n+1} transmit their labels to their children in G_n then the children in G_n form a segmentation that reflects the structure of the decimation.

A face $f \in F$ is surrounded by vertices. Let $L(f)$ be the set of labels of the surrounding vertices. Several faces $f_n \in F_n$ correspond to a single face $f_{n+1} \in F_{n+1}$ after decimation and label propagation. We formally collect them in the *merge area* $MA(f_{n+1})$ of a face $f_{n+1} \in F_{n+1}$ as follows:

$$MA(f_{n+1}) := \{f_n \in F_n \mid L(f_n) \subseteq L(f_{n+1})\} \ . \tag{1}$$

Figure 4 shows an example of a merge area. Faces within the merge area are outlined in bold. The propagation of labels from parents (•) to children (o) is indicated by arcs, small characters indicate the labels.

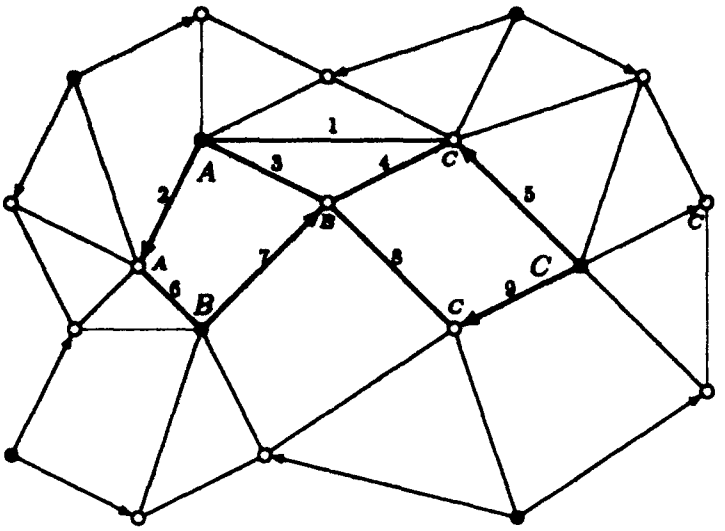


Fig. 4. Merge area and boundary of the triangle $(A, B, C) \in F_{n+1}$

4.1 The Extended MERGE Operation

We consider the merge area $MA(g)$ of a face $g \in F_{n+1}$. The *boundary of a merge area* g is the set $B(g)$ of all sides that separate faces belonging to $MA(g)$ from faces not belonging to $MA(g)$:

$$B(g) := \{\overline{e} = (f_i, f_j) \in \overline{E_n} \mid L(f_i) \subseteq L(g) \wedge L(f_j) \not\subseteq L(g)\} \ . \tag{2}$$

New curve relations are created in a face $g \in F_{n+1}$ by building the transitive closure (see Fig. 5) of all curve relations in the merge area $MA(g)$. Among those are relations that connect distant sides on the boundary $B(g)$.

Sides at level $n+1$ are spanned by surviving vertices $v \in V_{n+1}$. Consequently, we consider a segmentation of the merge area's boundaries. Every such boundary segment will correspond to one side of $\overline{E_{n+1}}$. *Boundary paths* of a merge area

$$P(p, q) \subset B(g) \tag{3}$$

are defined by two surviving vertices $p, q \in V_{n+1}$ with no other survivor lying on a connection (p, q) and built entirely of sides $\overline{e} \in B(g)$.

The transfer of the curve relations from level n to level $n+1$ by boundary unification is based on Jordan's curve theorem. A curve relation $(\overline{e_a}, \overline{e_b}) \in CR_n$ connecting a side $\overline{e_a}$ of path $P_A := P(p_A, q_A) \subset B(g), \overline{e_A} := \overline{(p_A, q_A)} \in \overline{E_{n+1}}$, with a side $\overline{e_b}$ of any boundary path $P_B := P(p_B, q_B) \subset B(g), \overline{e_B} := \overline{(p_B, q_B)} \in \overline{E_{n+1}}$ creates a curve relation $(\overline{e_A}, \overline{e_B})$ in face g:

$$\forall(\overline{e_a}, \overline{e_b}) \in CR_n, \overline{e_a} \in P_A, \overline{e_b} \in P_B \Longrightarrow (\overline{e_A}, \overline{e_B}) \in CR_{n+1} \ . \tag{4}$$

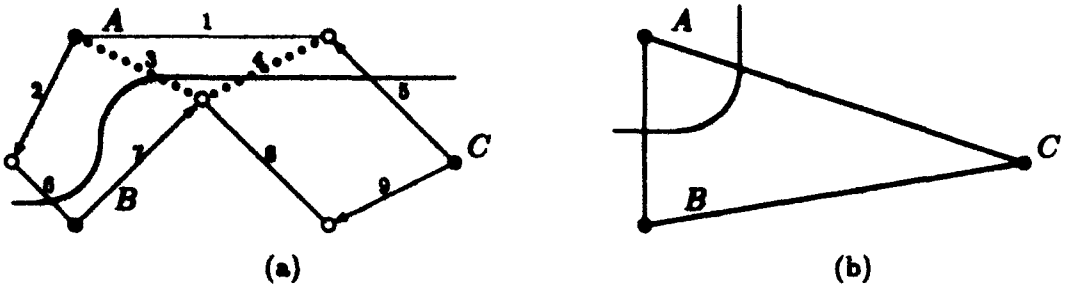


Fig. 5. Steps of the merge process

Example. The boundary in Fig. 4 consists of the following sides:

$$B(face(A, B, C)) = \{1, 5, 9, 8, 7, 6, 2\}.$$

Figure 5 shows a curve traversing the merge area of Fig. 4. It is represented by the curve relations (6,3), (3,4), and (4,5). New curve relations are created between two sides on the boundary by *transitive closure*:

- $(6, 4) := (6, 3) \oplus (3, 4)$;
- $(3, 5) := (3, 4) \oplus (4, 5)$;
- $(6, 5) := (6, 4) \oplus (4, 5) = (6, 3) \oplus (3, 5)$.

After elimination of inner sides 3 and 4 the curve connects sides 6 and 5. The three boundary paths connecting the surviving vertices (A, B, C) are

- $P(A, B) = \{2, 6\}$,
- $P(B, C) = \{7, 8, 9\}$,
- $P(C, A) = \{5, 1\}$.

Since 6 is on path $P(A, B)$ and 5 on path $P(C, A)$, the curve connects sides $\overline{(A, B)}$ and $\overline{(C, A)}$ in $\overline{E_{n+1}}$ which is expressed by a curve relation $(\overline{(A, B)}, \overline{(C, A)}) \in CR_{n+1}$ ('unification') in $face(A, B, C) \in F_{n+1}$.

4.2 Isolated and Central Survivors

Unfortunately survivors $v_s \in V_{n+1}$ need not always be on the boundary $B(g)$ of their merge areas. Depending on the neighbours of v_s, two exceptions can be differentiated, *isolated* (Fig. 6a) and *central vertices* (Fig. 6b):

$$v_s \text{ is } \begin{cases} \text{isolated, if } card\{v \in BV(g) \mid (v, v_s) \in E_n \vee (v_s, v) \in E_n\} = 1 \ , \\ \text{central, if } card\{v \in BV(g) \mid (v, v_s) \in E_n \vee (v_s, v) \in E_n\} > 1 \ , \end{cases} \tag{5}$$

where $BV(g) := \{v_i \mid (v_i, v_j) \in B(g) \vee (v_j, v_i) \in B(g)\}$ define the *boundary vertices* of g. In both cases these *non-boundary survivors* cannot be connected to any of their neighbours in face $g \in F_{n+1}$, using only sides of $B(g)$.

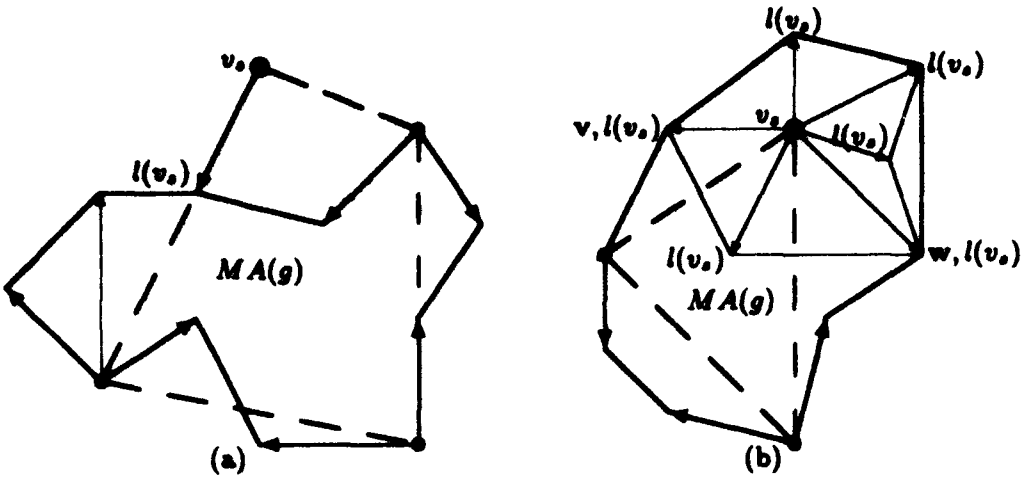


Fig. 6. Isolated and central vertex of a merge area $MA(g)$

A *central vertex* v_s is completely surrounded by faces g carrying only the label $l(v_s)$. Since $l(v_s) \in L(g)$, $BV(g)$ contains vertices with label $l(v_s)$ but not v_s. The subsequence of vertices in $B(g)$ with label $l(v_s)$ is bounded by two vertices v and w (see Fig. 6 b). Both v and w are neighbours of v_s in G_n and allow the following modifications of BV, B, and the label sets of sides in $B(g)$:

1. $BV'(g) = \{v_i \in BV(g) \mid l(v_i) \neq l(v_s)\} \cup \{v, w, v_s\}$;
2. $B'(g) = \{(f_1, f_2) \in B(g) \mid \overline{(v_1, v_2)} = (f_1, f_2), v_1 \in BV'(g) \wedge v_2 \in BV'(g)\}$;
3. Remove $L(g)$ from $B(g) - B'(g)$.

By this modification of the boundary we exclude several faces from merge area $MA(g)$. It can be shown that the curve relations stored in these faces are included for transitive closure in another merge area [12].

The edge $e_s \in E_n$ between an *isolated survivor* v_s and its child is the only connection to merge area $MA(g)$. Consequently, boundary paths to its neighbouring survivors must contain this edge. If included in the extended boundary $B'(g)$, side $\overline{e_s}$ must be traversed twice to close the extended boundary. Since $\overline{e_s}$ corresponds to two different sides in $\overline{E_{n+1}}$ we introduce an *auxiliary curve relation* between the two sides if $\overline{e_s}$ is crossed by any curve. This operation is necessary to preserve the connectivity of the curve in CR_{n+1}.

4.3 The Extended SPLIT Operation

Adjacent merge areas may overlap, $MA(f) \cap MA(g) \neq \emptyset$. Curve relations in such overlap areas contribute to both merge areas. Jordan's curve theorem satisfies the consistency if the curve crosses the overlap area from $MA(f)$ to $MA(g)$. However a false U-turn may be generated (e.g. in face $g = face(B, C, D)$

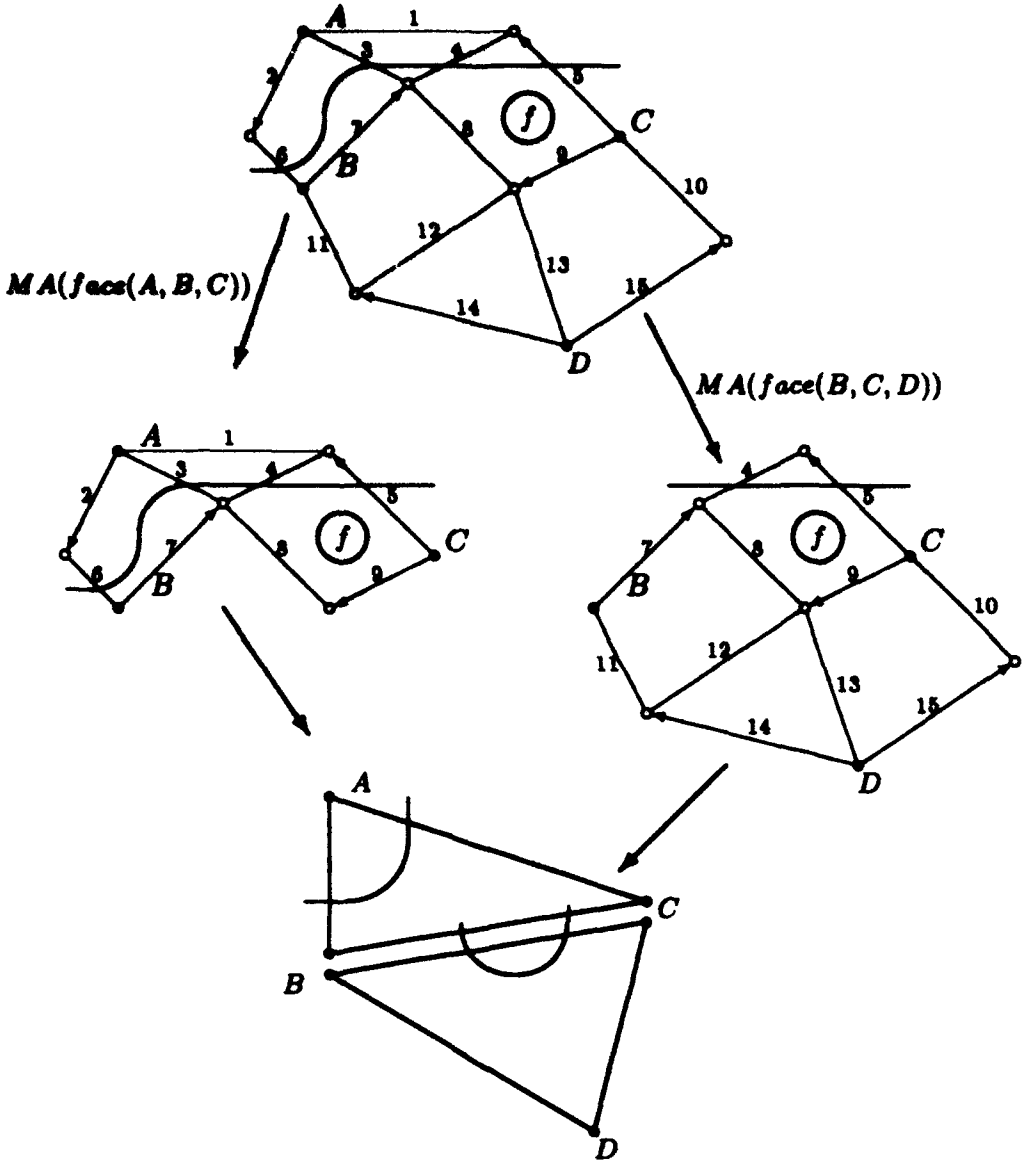


Fig. 7. Overlapping merge areas may create false U-turns

of Fig. 7) if curves remain completely within one merge area, e.g. $MA(f)$ ($f = face(A, B, C)$ in Fig. 7).

An overlap area is limited by two boundary paths $P_f \subset B(f)$ and $P_g \subset B(g)$. Since adjacent faces $f, g \in F_{n+1}$ have one side $\overline{(p,q)} \in \overline{E_{n+1}}$ in common, both $P_f(p,q)$ and $P_g(p,q)$ connect also $p \in V_n$ and $q \in V_n$. A curve creating a false U-turn would cross P_f twice without leaving the overlap across P_g. We find

this property in two steps. First we initialize all sides $\bar{e} \in \overline{E_{n+1}}$ with an empty label set, $LS(\bar{e}) := \emptyset$, and accumulate labels on the sides $\overline{e_b}$ of all boundaries $B(z), z \in F_{n+1}$ with label set $L(z)$, $LS(\overline{e_b}) := LS(\overline{e_b}) \cup \{L(z)\}$. Second we propagate the label sets LS of the sides to the curve relations and accumulate them during transitive closure:

$$LS(r) := LS(\overline{e_1}) \cup LS(\overline{e_2}) \quad \text{for } r = (\overline{e_1}, \overline{e_2}) \in CR_n \tag{6}$$

$$LS(r \oplus s) := LS(r) \cup LS(s) \quad \text{for } r = (\overline{e_1}, \overline{e_2}) \in CR_n, s = (\overline{e_2}, \overline{e_3}) \in CR_n \tag{7}$$

After the *transitive closure-boundary unification* cycle (MERGE) described above, any consistent curve relation carries at least two different label sets. We can therefore delete U-turns which carry only one label set because the curve enters and leaves an overlap area without connecting P_f and P_g.

5 Control Information from Curve Relations

The representation of an object in the regular $2 \times 2/2$ curve pyramid depends on its position with respect to surviving vertices. In the stochastic decimation vertices survive depending on a random variable. The value of this variable provides the means to overcome the position variance problem: depending on the image data we can add scores to vertices. The rules according to which scores are distributed depend on the recognition problem to be solved. Scores on vertices can be accumulated and represent thus the vertices' importance with respect to a given application.

Our strategy to remove the drawbacks of the regular curve pyramid is double: eliminating isolated blobs and preserving parallel curves. From this strategy we derive rules for scoring vertices.

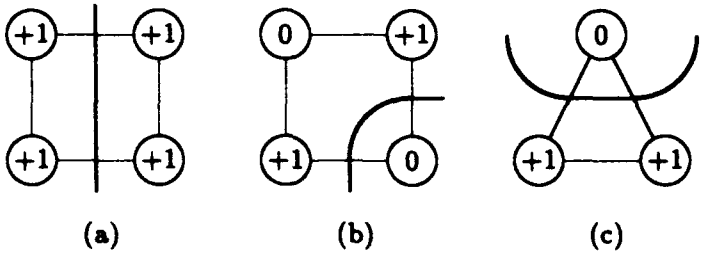


Fig. 8. Scores are distributed in faces with a curve relation

Isolated blobs are surrounded by a closed curve. They disappear as blobs if the vertex surrounded by the curve is a non-survivor, and they are preserved if it survives. Closed curves that surround a single vertex cannot be reduced in length any more. To preserve the length reduction property of the curve pyramid, these curves, i.e. the vertices they surround, must be eliminated. A closed curve

is represented in faces surrounding a vertex. In any of these faces the curve relation is a turn around this vertex, i.e. the related edges have that vertex in common. Since it is desirable that it does not survive during decimation, we add 1 to its neighbouring vertices in the face (Fig. 8b,c).

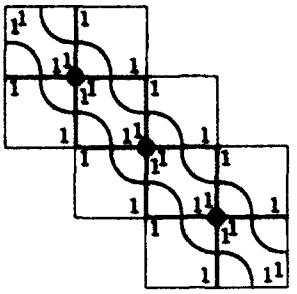


Fig. 9. Score distribution for the example of the paraliel lines

Parallel curves are represented in adjacent faces in F_n. To preserve those curves, vertices between them should survive. Therefore scores are also distributed in faces where the related sides do not form a corner. All vertices that span the connected sides are therefore incremented by 1 (Fig. 8a). Thus if two adjacent faces hold a curve relation each, the vertices that these two faces have in common accumulate the scores distributed in the faces. For the example of the parallel lines, the resulting score distribution is reported in Fig. 9.

The random numbers added to the scores before the decimation are within $[0, 1)$. Since they do not exceed the quantization step of the score distribution, the selection of survivors implied by the curve relations remains unchanged. However, they are used to break the remaining ties, e.g. in areas without curve relations.

6 Conclusion

We have shown that by means of scores one can control the survival of vertices in a sampling grid during decimation. We derived the scores from curve relations stored in the dual graph $\overline{G_n}$ of the sampling grid.

The position-variance problem mentioned in Sect. 2.1 has been solved: by means of the scores we can avoid the merging of two parallel lines, and also isolated blobs can now be eliminated independent of their position in the sampling grid. However blobs cannot be eliminated in every case, whatever decimation is chosen. Consider, for example, an image plane where every vertex is surrounded by a closed curve. Then surviving vertices will also be surrounded by closed curves. Consequently, we cannot guarantee the annihilation of blobs in all cases, but images where blobs are forced to survive must show a very high density of

blobs. This characterizes rather a texture than a representation of objects. To be sure of identifying objects correctly in the curve pyramid it is necessary that objects are distinct from their surrounding in all scales considered.

Although motivated by the goal of finding long connected curves in digital images the scope of the presented approach goes beyond the original goal. Putting together curve segments that meet in a common point has many analogies in image analysis as in other fields of computer science. For example, a complex object is often composed of several parts which are themselves composed of smaller parts. These object parts have different physical properties and, hence, may be recognized as individual image parts. In order to reassemble an object from its parts, pairs (or triples ...) of parts must satisfy certain constraints, e.g. they meet at a given angle. Checking all possible combinations of parts is a problem of high computational complexity ("combinatorial explosion"), as is the problem of building the transitive closure of a set of relations. This approach considerably reduces this complexity by

1. embedding the set into a discrete partition of a (geometrical) space;
2. building a hierarchy of partitions using only local processes to aggregate small parts to larger parts;
3. adapting the structure of this hierarchy to the image data to overcome certain problems arising in rigid (regular) structures.

References

1. Bister, M., Cornelis, J., Rosenfeld, A. (1990). A critical view of pyramid segmentation algorithms, Pattern Recognition Letters 11, pp. 605–617.
2. Jolion, J.M., Montanvert, A. (1992). The adaptive pyramid, a framework for 2D image analysis, Computer Vision, Graphics, Image Processing: Image Understanding 55, pp. 339–348.
3. Kropatsch, W.G. (1985). Hierarchical curve representation in a new pyramid scheme, Technical Report TR-1522, University of Maryland, Computer Science Center.
4. Kropatsch, W.G. (1985). A pyramid that grows by powers of 2, Pattern Recognition Letters 3, pp.315–322.
5. Kropatsch, W.G. (1986). Kurvenrepräsentation in Pyramiden. In: Kropatsch, W.G., Mandl, P. (eds.), Mustererkennung'86, OCG-Schriftenreihe, B36. Österr. Arbeitsgruppe für Mustererkennung, Oldenbourg, pp. 16–51.
6. Kropatsch, W.G. (1987). Curve representations in multiple resolutions, Pattern Recognition Letters 6, pp. 179–184.
7. Kropatsch, W.G. (1987). Elimination von "kleinen" Kurvenstücken in der 2 × 2/2 Kurvenpyramide. In: Paulus, E. (ed.), Mustererkennung 1987, Informatik Fachberichte 149, Springer-Verlag, Berlin, pp. 156–160.
8. Kropatsch, W.G. (1988). Preserving contours in dual pyramids, Proc. 9th Int. Conf. on Pattern Recognition, Rome, Italy, IEEE Comp. Soc., pp. 563–565.
9. Kropatsch, W.G. (1990). Digitales Sehen mit Bildpyramiden, Elektronik-Journal: Maschinelles Sehen - Industrielle Bildverarbeitung, pp. 93–102.

10. Kropatsch, W.G. (1990). Hierarchical methods for robot vision. In: Jordanides, T., Torby, B. (eds.), Expert Systems Robotics, NATO ASI Series F, Vol. 71, Springer-Verlag, Berlin, pp. 63–109.
11. Kropatsch, W.G., Montanvert, A. (1991). Irregular versus regular pyramid structures. In: Eckhardt, U., Hübler, A., Nagel, W., Werner, G. (eds.), Geometrical Problems of Image Processing Georgenthal, Germany. Akademie Verlag, Berlin, pp. 11–22.
12. Kropatsch, W.G., Willersinn, D. (1992). Representing curves in irregular pyramids. In: Kropatsch, W.G. Bischof, H. (eds.), Pattern Recognition 1992, Vienna, Austria, Oldenbourg, pp. 333–348.
13. Meer, P. (1989). Stochastic image pyramids, Computer Vision, Graphics, Image Processing 45, pp. 269–294.
14. Montanvert, A., Bertolino, P. (1992). Irregular pyramids for parallel image segmentation. In: Kropatsch, W.G., Bischof, H. (eds.), Pattern Recognition 1992, Vienna, Austria, Oldenbourg, pp. 13–34
15. Montanvert, A., Meer, P., Rosenfeld, A. (1992). Irregular tesselation based image analysis, Proc. 10th Int. Conf. on Pattern Recognition, Atlantic City, New Jersey, USA, IEEE Comp. Soc. Vol. I, pp. 474–479.
16. Paar, G., Kropatsch, W.G. (1988). Hierarchical cooperation between numerical symbolic image representations. In: Mohr,R., Pavlidis,T., Sanfeliu, A. (eds.), Structural Pattern Analysis, World Scientific Publ. Co., pp. 113–130.

Multiresolution Shape Description by Corners

Cornelia Fermüller[1,2] *and Walter Kropatsch*[1]

[1] Department for Pattern Recognition and Image Processing, Institute for Automation, Technical University of Vienna, Treitlstraße 3, A-1040 Vienna, Austria
[2] Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3411

Abstract. A robust method for describing planar curves in multiple resolution using curvature information is presented. The method is developed by taking into account the discrete nature of digital images as well as the discrete aspect of a multiresolution structure (pyramid). The main contribution of this paper lies in the robustness of the technique, which is due to the additional information that is extracted from observing the behaviour of corners in the pyramid. Furthermore the resulting algorithm is conceptually simple and easily parallelizable. Theoretical results showing the behaviour of curvature extrema under varying scales are developed based on the analysis of curvature of continuous curves in scale-space. These results are used to eliminate any ambiguities that might arise from sampling problems due to the discreteness of the representation. Finally, experimental results demonstrate the potential of the method.

Keywords: 2-D shape description, scale-space, reduction of curvature extrema, corner detection, multiresolution representation.

1 Introduction

The aim of this work is to introduce a curve description suitable for many higher-level visual tasks, such as matching used in problems related to stereo, motion, or object recognition. This description uses the corners or curvature extrema of curves, since they provide a natural means of segmenting boundaries [2].

A description of curves should clearly be robust under rotation, scaling and translation. Further criteria of importance for a reliable computer description are: the computability of the representation by using only local support, the representation of the description at varying levels of detail and its stability, defined in the sense of numerical analysis; that is, small changes in the input should cause only small changes in the representation.

The curve pyramid [11] is first used in order to obtain a representation of the curve at varying levels of detail. Different resolutions of curves in digital images are calculated by reducing a small number of curve segments at higher resolution to one segment at lower resolution. The images (the levels of the

pyramid) are superimposed on each other in such a way that there exists a geometrical relationship between their elements.

Then, a method for calculating corners in parallel is introduced. It is based on the idea of deciding whether a pixel represents a corner by looking only at the pixel itself and a few of its neighbors. Continuous curves in scale-space are considered to analyze the behaviour of curvature extrema under smoothing. The results obtained are used to define measures for the description of a curve in the pyramid. These measures form the basis of a stable description.

Previously published methods dealing with descriptions of planar curves that are based on points of interest along the curve can be broadly classified as those performing corner detection at one scale and those dealing with descriptions at different scales. The latter are further classifiable into methods that deal with the problem in a continuous manner, in scale-space [1, 15], and methods that represent the data in a discrete way by employing multiresolution structures (e.g. pyramids [9]) or using symbolic representations of features at multiple scales [17].

Techniques that operate at just one level of resolution may suffer from the disadvantage of finding many unimportant details while at the same time missing large rounded corners. Techniques that operate in scale-space on a continuous representation are quite elaborate, involving considerable overhead. Therefore, various discretization schemes have been introduced [3]. Although scale-space methods have produced interesting results, they may be problematic in practical applications, since they must employ either 1-D [15] or 2-D smoothing [8, 16]. In the first case important large scale structures may be lost, while in the second case the topological properties may be destroyed. Finally, techniques that operate on a discrete pyramid, where the number of grid points is reduced from one level to the next, are limited to a finite number of resolutions and may suffer from the problem of undersampling.

The method introduced in this paper, which employs "syntactic smoothing" (to be explained later), works on a discrete pyramid [11], but takes advantage of mathematical relationships among curves in scale-space and can thus be considered as a hybrid algorithm. Although it is of a discrete nature, it is supported by scale-space information. Furthermore, it combines the advantages of 1-D and 2-D smoothing, since local 2-D smoothing is performed, but the context information inherent in the curve is considered.

2 Curvature Points of Continuous Curves in Scale-space

Planar curves are described by points where the curvature has a local extremum or has the value zero.

Definition 1 Curvature points. $C(t)$, with t any parameterization, is an oriented, planar, closed curve. The maxima, minima, and zero-crossings of the curvature are called *curvature points*. Among the extrema there is a further distinction as to whether the value of the curvature at these points is positive or negative. Therefore there exist five classes of points: positive maximum

(Max$^+$), negative maximum (Max$^-$), positive minimum (Min$^+$), negative minimum (Min$^-$) and inflection point (0) (Fig. 1).

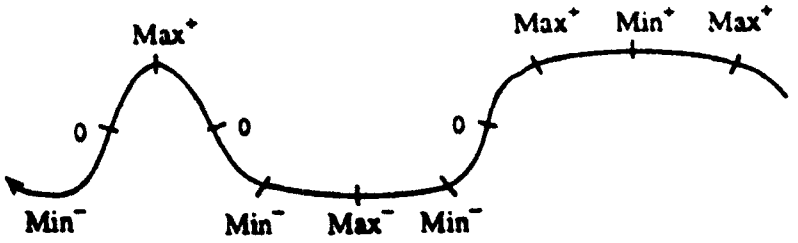


Fig. 1. Curvature points of a curve

This classification depends on the orientation of the curve. If it changes, positive maxima become negative minima, negative maxima become positive minima and vice versa.

Our method of curve description using curvature employs both the size (value) and scope of the curvature.

Definition 2 Scope of curvature. Let $C(t)$ be an oriented, planar, closed curve. The *scope of curvature* $(B(t_k))$ with centre at point C_{t_k} consists of the right and left scope $B(t_k) = \{BR(t_k), BL(t_k)\}$. The right scope is the length of the curve's arc from C_{t_k} to the next curvature point in the positive orientation, and $BL(t_k)$ is the arc's length from C_{t_k} to the next curvature point in the negative orientation.

Using only the corners detected at one scale is often not useful for creating a representative description. A resolution that is too fine may show many unimportant details and large rounded corners may not be detected, while at too coarse a resolution important corners may be missed. Therefore, a description at different scales seems to be desirable. To relate the descriptions at different scales we need to analyze the behaviour of curves under progressive smoothing. This method is called scale-space filtering [18]. It is a way of describing a curve $C(t, \sigma)$ under smoothing with a kernel of width σ, where σ is treated as a continuously increasing parameter. $C(t, \sigma)$ is the convolution $(\star)$ of a curve $C(t)$ with a kernel $g(t, \sigma)$:

$$C(t, \sigma) = C(t) \star g(t, \sigma).$$

In principle, there are several possibilities for choosing $g(t, \sigma)$. Babaud *et al.* [4] and Yuille and Poggio [19] proved that when filtering a one-dimensional function $x(t)$ with a Gaussian, no generic zero-crossings and no curvature extrema are created as the scale increases. Bergholm [5] showed that when blurring with a two-dimensional Gaussian, any closed curve turns into a circle.

An equivalent way of generating the family of signals in scale-space is by solving the diffusion equation [10]. Lindeberg [13] has analyzed the nature of

smoothing kernels when dealing with discrete signals, which led to the development of a discrete analog of Gaussian kernels and to a discretized version of the diffusion equation.

The previous studies show that the number of maxima, minima, and zero-crossings of curvature, for curves that are smoothed in these ways, decreases. In this paper curves are characterized by their corners; therefore the behaviour of maxima and minima of curvature under smoothing is analyzed. The following theorem makes this explicit.

Proposition 3. *There are ten possible combinations of three successive curvature points, when the middle one is an extremum as listed below. The local reduction of these triples under smoothing are shown below:*

$$
\begin{array}{lcccl}
(R1) & Min^+ & Max^+ & Min^+ & \longrightarrow Min^+ \\
(R2) & 0 & Max^+ & Min^+ & \longrightarrow 0 \\
(R3) & Min^+ & Max^+ & 0 & \longrightarrow 0 \\
(R4) & 0 & Max^+ & 0 & \longrightarrow Max^- \\
(R5) & Max^+ & Min^+ & Max^+ & \longrightarrow Max^+ \\
(R6) & Min^- & Max^- & Min^- & \longrightarrow Min^- \\
(R7) & Max^- & Min^- & Max^- & \longrightarrow Max^- \\
(R8) & Max^- & Min^- & 0 & \longrightarrow 0 \\
(R9) & 0 & Min^- & Max^- & \longrightarrow 0 \\
(R10) & 0 & Min^- & 0 & \longrightarrow Min^+
\end{array}
$$

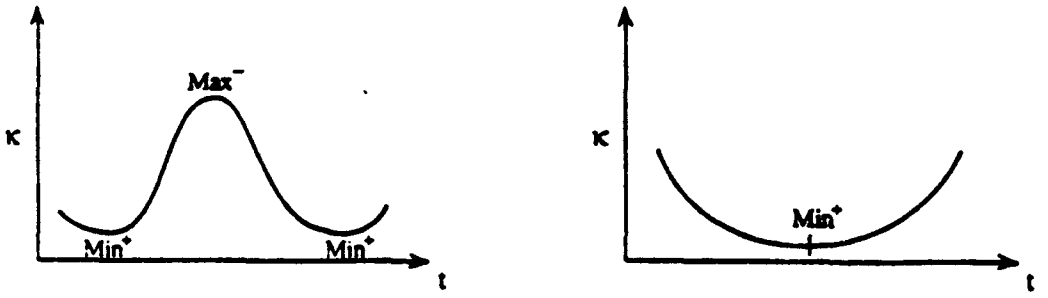


Fig. 2. Reduction of curvature points

To prove the validity of each reduction one needs to plot the curve in 2-D space, with the x-axis representing the arc length and the y-axis the curvature and observe in this space when the middle extremum disappears [6]. For example R1 can be established by comparing Fig. 2a and Fig. 2b.

At this point we need to emphasize that the results of the theorem are of a syntactic nature and do not involve the smoothing parameter σ. The value of σ at which the middle extremum in any of the rules (R1–R10) disappears depends on the size and scope of curvature at the points under consideration.

3 Digital Representation of Curves

In this study the curve code [12] is used for encoding curves. A digital image is overlaid with a grid. Curves are represented by their intersections with the sides of the square grid cells. This information about intersections is stored in the cells.

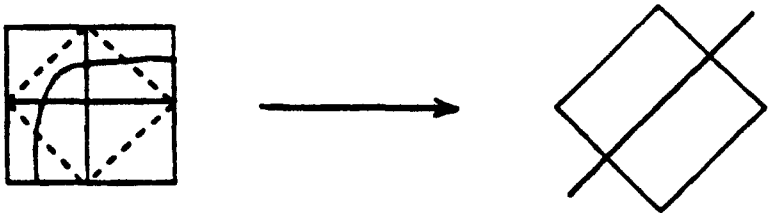


Fig. 3. Reduction of curve code in the pyramid

The curve pyramid is obtained by merging the contents of the cells and producing in this way a stack of images of different cell sizes and different resolutions, where the cells' areas at each resolution are twice as large as those at the next lower resolution. The lattices are rotated by 45° from level to level [12]. In the first step the squares are divided by a diagonal into two triangles (operation split) and in the second step groups of four triangles are merged and their contents are reduced to the content of one cell at the lower resolution (operation merge) [11] (Fig. 3). Since there are two possibilities for splitting a curve, there are 2^n possibilities for building a pyramid of n levels.

3.1 RULI-Chain Code

In order to explain our theoretical research the RULI-chain code has to be introduced. The curve is followed from a starting point to an end point, so the coding of spatial position within the segments can be dropped. The relative movement within the cell can be described by one of four different symbols: R for curves that enter the cell at one border and leave it at the border on the right; L for curves that turn to the left; I for curves that pass straight through; and U for curves that enter and leave at the same side (see Fig. 4). A sequence of such code elements is called a RULI-chain.

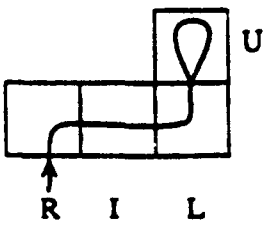


Fig. 4. RULI-chain code

The reduction of a RULI-chain can be done by a formal grammar [12], and therefore a description of a curve through a RULI-chain code and its formal reduction becomes equivalent to a description in the curve pyramid.

This syntactic method of reducing the resolution offers an alternative to one- and two-dimensional smoothing. As mentioned earlier, both of these methods have disadvantages: two-dimensional smoothing does not necessarily maintain the topological properties and one-dimensional scale-space descriptions, such as the analytical one described above, perform less well with spike-like, highly convex or concave features [14]. With the proposed method these problems can be overcome because the reduction used is a smoothing of curves in both dimensions, which takes into account 1-D context information.

4 Corners of a RULI-chain

The corner detection algorithm introduced here is a parallel one; just three or five picture elements are considered in order to decide if a point represents a corner and in this way two types of curvature points are considered: Max^+ and Min^-. Curvatures whose estimation need more than five pixels will not be detected in this step.

A code is recognized as originating from curvature if it cannot be created by a straight line.

The following proposition shows what corners look like in the RULI-chain code.

Proposition 4. *If a corner is detected from a code sequence of up to three (or five) elements, it must consist of one of the following sequences:*

Corners of three elements:	*Corners of five elements:*
1. $\hat{U}$	*1.* $\hat{U}$, $\widehat{RR}$ *or* $\widehat{LL}$, $I\hat{R}I$ *or* $I\hat{L}I$
2. $\widehat{RR}$ *or* $\widehat{LL}$	*2.* $R\hat{I}R$ *or* $L\hat{I}L$, $R\hat{\Pi}R$ *or* $L\hat{\Pi}L$
3. $I\hat{R}I$ *or* $I\hat{L}I$	*3.* $IR\hat{L}RI$ *or* $ILR\hat{L}LI$
4. $R\hat{I}R$ *or* $L\hat{I}L$	*4.* $RL\hat{R}II$ *or* $LR\hat{L}II$, $IIR\hat{L}R$ *or* $IIL\hat{R}L$
	5. $RI\hat{\Pi}IR$ *or* $LI\hat{\Pi}IL$

The proof follows directly from the contradiction of the corner sequences to the conditions of straightness. The procedure using only three elements is called the *Three-element method* and the one using five elements is called the *Five-element method*.

5 Detectability of Corners

5.1 Necessary Corner Conditions

The particular choice of the curve code out of the 2^n possibilities is determined by its position in the pyramid. Since the description should be independent of the

specific reduction the *size* of the *angles* in the corner sequences and the *distance* between two neighboring corners are examined and *necessary corner conditions* are developed. These conditions have to be satisfied by corners in order for them to be detectable under all possible reductions.

For the Three-element method (Five-element method) the straight lines forming a corner must enclose an angle of 63.4° (108.4°) and two neighbouring corners have to be at one receptive field (a receptive region of three code-elements) distance. A receptive field of a cell at a level k in the pyramid is defined as the region from which this cell obtains its information and a receptive region is the union of receptive fields corresponding to neighbouring cells. If we detect corners with the Three- or Five-element method and exclude those that do not satisfy the necessary corner conditions, we detect only corners which are detectable under all reductions.

5.2 Corner Detection in the Pyramid

Since the description is a discrete one, the problems caused by undersampling have to be considered. It will now be shown how this might affect the description and a remedy will be presented.

In comparison with an analytical description of curves, where all curvature points that appear at a low resolution must also be detected at a high resolution, in pyramids curvature points may appear for the first time at a low resolution. These points correspond to curvatures which cannot be detected with the proposed parallel method because of the angle's size (for example, corner 15 or 18 in Fig. 5). Because of discrete sampling, corners may appear at level E_{i-1} and E_{i+1} but not at level E_i (e.g., corner 3 is not detected at level 3 in Fig. 5). Furthermore, it is possible for a corner to be detected as two adjoining corners at the level above, but in these cases the necessary corner conditions are not satisfied. However, the proposed algorithm does not suffer from such problems because information about corners is complemented with knowledge about the scale-space behaviour of curves.

5.3 Robustness Measurements

Measures that reflect the size and scope of the curvature in the pyramid are defined next.

Three properties are important, namely the lowest and highest levels at which a corner is detected and at how many levels it is detected.

The last appearance (moving from the bottom to the top) of a corner in the pyramid gives information about the scope of the curvature; we therefore call it the *measure of scope* (S). The first appearance reflects the size of curvature; it is called the *measure of curvature-approximation* (C). The sharper the enclosed angle, the earlier the curvature point will be detected. The number of levels at which a corner is detected is called the *measure of importance* (I). For individual corners, it was proved that they must be detected when they satisfy the corner

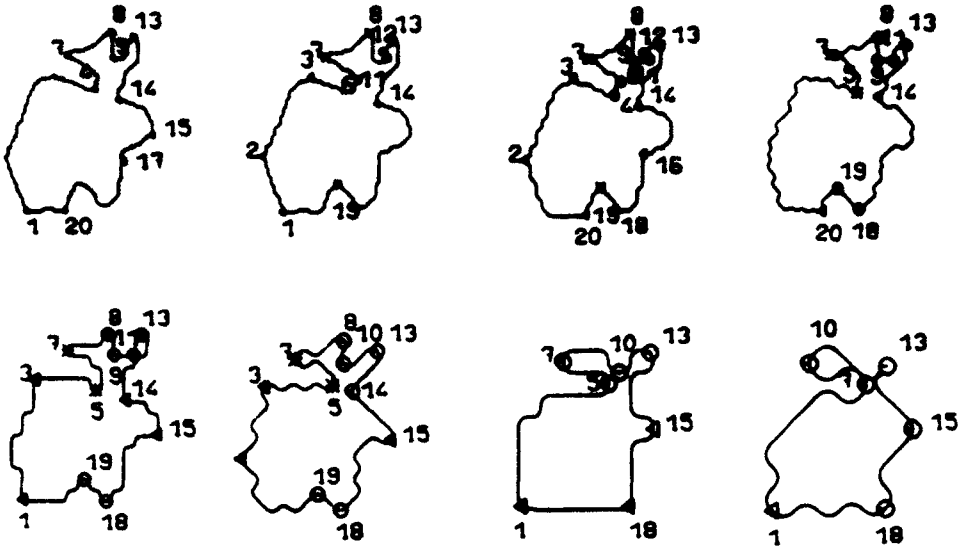


Fig. 5. Corners of a closed curve in the pyramid: $*$ ... corners that satisfy the necessary corner conditions; $\oplus$ ... corners which do not satisfy the distance condition; $\triangle$ ... corners not satisfying the condition of the angle's size, but satisfying the distance condition.

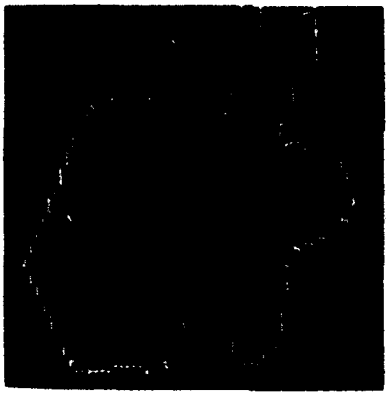


Fig. 6. Results of curve partitioning method by Fischler and Bolles.

conditions. The measure of importance describes the possibility of the appearance of a corner in the pyramid. It is introduced to make the description usable for more corners.

The three measurements stabilize the description in the sense that small changes in the input result in small changes in the description.

6 Experimental results

The results of applying our method to a curve which was also used by Fischler *et al.* [7] are shown in Fig. 5. The curve is plotted at eight successive levels of

resolution. At every level corners are extracted using the three-element method and classified in relation to size of angle and distance to the next corner.

The method of Fischler *et al.* for curve partitioning is based on the arc-chord distance. For every point on the curve they decide whether the arc stays close to the chord or makes excursions away from it and partitions the curve at the points of single excursion that are farthest away from the chord. Points must be at least a predefined distance apart in order to be detected as different points. They chose a quarter of a chord length as the threshold. In their paper they show the partitioning of the curve at one resolution; their results are shown in Fig. 6. Comparisons to our method show that with their relatively large threshold they do not extract small corners (6,11,12,17). They also do not detect point 2 and none of the points between 8 and 13.

The advantage of our method lies in the fact that not only corners at different resolutions are detected, but the descriptions at different resolutions are combined into one description. Therefore, corners can be differentiated by adding attributes to them.

7 Conclusions

A multiresolution description of planar curves using corners and the curve pyramid has been presented. Continuous curves under smoothing have been examined and the results used to define measures that stabilize the description. A method has been developed for detecting corners of digital curves in parallel. This local method has been analyzed. It was found that corners are detected in all cases when the straight lines enclose an angle of at least 63.4° (108.4°) and the distance from one corner to the next is a receptive field (a receptive region of three cells).

A possible application of this description is multiple-resolution contour matching. Starting at a low resolution, the description can gradually be refined by adding the information stored at the next higher level of resolution of the pyramid.

References

1. Asada, H., Brady, M. (1986). The curvature primal sketch, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(1), pp. 2–14.
2. Attneave, F. (1954). Some informational aspects of visual perception, Psychological Review, 61(3), pp. 183–193.
3. Aviad, Z. (1987). A discrete scale-space representation, Proc. Int. Conf. on Computer Vision, pp. 417–422.
4. Babaud, J., Witkin, A. P., Baudin, M., Duda, R. O. (1986). Uniqueness of the gaussian kernel for scale-space filtering, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(1), pp. 26–33.
5. Bergholm, F. (1987). Edge focusing, IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-9(6) pp. 726–741.

6. Fermüller, C. (1989). Hierarchisches Vergleichen von Konturen. Master's thesis, Technical Report 42, Institute for Image Processing and Computer Graphics, Joanneum Research, Graz.
7. Fischler, M. A, Bolles, R. C. (1986). Perceptual organisation and curve partitioning, IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-8(1), pp. 100–105.
8. Goshtasby, A. (1986). Multiple-scale segmentation and representation of solid plane shapes, Proc. Conf. on Computer Vision and Pattern Recognition, pp. 351–355.
9. Hartley R., Rosenfeld, A. (1983). Hierarchical line linking for corner detection, Technical Report CS-TR-1288, Center for Automation Research, University of Maryland.
10. Koenderik, J. J., Doorn, A. van (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
11. Kropatsch, W. G. (1985). A pyramid that grows by powers of 2, Pattern Recognition Letters, 3, pp. 315–322.
12. Kropatsch, W. G. (1987). Curve representation in multiple resolution, Pattern Recognition Letters, 6(3), pp. 179–184.
13. Lindeberg, T. (1991). Discrete Scale-Space Theory and the Scale-Space Primal Sketch, PhD thesis, Computational Vision and Active Perception Laboratory, Royal Institute of Technology, Stockholm, Sweden, 1991.
14. Meer, P., Baugher, E. S., Rosenfeld, A. (1988). Extraction of trend lines and extrema from multiscale curves, Pattern Recognition, 21(3), pp. 217–226.
15. Mokhtarian, F., Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8 (1), pp. 34–43.
16. Richards, W., Dawson, B., Whittington, D. (1986). Encoding contour shape by curvature extrema, J. Opt. Soc. Am., 3 (3), pp. 1483–1491.
17. Saund, E. (1990). Symbolic construction of a 2-d scale-space image, IEEE Trans. on Pattern Analysis and Machine Intelligence, 12(8), pp. 817–830.
18. Witkin, A. P. (1983). Scale-space filtering, Proc. 7th Int. Joint Conf. on Artificial Intelligence, pp. 1019–1022.
19. Yuille, A. L., Poggio, T. A. (1986). Scaling theorems for zero crossings, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8 (1), pp. 15–25.

Model-based Bottom-Up Grouping of Geometric Image Primitives

Peter Nacken * *and Alexander Toet*

T.N.O. Institute for Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

Abstract. A new bottom-up technique for grouping geometric image primitives is presented. In this scheme, each pair of adjacent primitives is compared to a model. The outcome of this evaluation is used to select pairs of primitives for merging. The technique is applied in a hierarchical graph context using algorithms which perform in parallel and use only local information. These algorithms inherently have a stochastic nature. Limitations imposed by the scheme's bottom-up character can be remedied by the introduction of a top-down flow of information.

Keywords: graph representation, hierarchy of graphs, segmentation, grouping, region adjacency graph, maximal independent set.

1 Introduction

Low-level image analysis frequently involves *grouping* of geometric primitives [13]. For grey-level image segmentation, pixels must be grouped in such a way that (*i*) the regions which they represent satisfy some homogeneity condition and (*ii*) adjacent regions have distinct properties. For polygonal curve approximation, pixels must be grouped such that (*i*) the curve segments they represent can be approximated by a line segment and (*ii*) adjacent groups (segments) have different orientations.

Every segmentation algorithm must (implicitly or explicitly) adopt a *model* for homogeneous image regions. The most simple model adopts a constant grey-level value for each region. In some cases (e.g. for the segmentation of textured images) more intricate models may be required. Segmentation also requires an *error measure* for determining the deviation of an image region from the model.

* This research was supported by the Foundation for Computer Science in The Netherlands (SION) with financial support from The Netherlands Organisation for Scientific Research (NWO). It was performed in a joint project of TNO Institute for Human Factors, the Centre for Mathematics and Computer Science (Amsterdam) and the Faculty of Mathematics and Computer Science of the University of Amsterdam.

For the piecewise-constant grey-level model one often uses the root-mean-square value of the residues.

Many techniques for image segmentation have been proposed. They can be divided into two groups: those computing the homogeneous *regions* in the image and those computing the *boundaries* between such regions. Region-oriented techniques compute connected groups of pixels which satisfy the region model. Edge-detection methods detect points in the image which satisfy a discontinuity model. It is not possible to evaluate all groupings. Groups of pixels that satisfy the model are typically found by splitting or merging groups repeatedly until the result fits the model within a given error.

Two different approaches to bottom-up grouping can be distinguished:

1. Region-*merging* methods [3] first consider each pixel as an individual region. Two regions are replaced by their union if the latter satisfies the model. Merging continues until no union of adjacent regions satisfies the model.
2. Region-*growing* methods [15] first select a special set of pixels called *seeds*. Regions are grown by aggregating pixels to the seeds. This growth process continues until the image plane is covered by regions which satisfy the model. Each region in the final segmentation contains exactly one seed. The selection of appropriate seeds is a difficult problem, which requires procedures that are adapted to a particular class of images (e.g. [10]).

Bottom-up techniques use only *local information*, that is, information from a restricted area. The regions over which information is collected increase progressively as the grouping process continues. When global information becomes available, a clustering performed in the early stages may prove incorrect. Relinking methods [4] can be invoked to revise incorrect clusterings.

In contrast to stochastic pyramid schemes [9, 12], in which an arbitrary number of adjacent regions can be replaced by their union, the bottom-up grouping scheme performs *pairwise region merging*. For each pair of adjacent regions the fit of their union to the region model is calculated. This information is used to select the pairs that are actually merged. In this sense, the method is also related to merging schemes like the one described by Beveridge [1].

The rest of this paper is organized as follows. In the next section grouping is presented in a hierarchical-graph context and some related work is described. Section 3 describes the grouping method. In Sect. 4 some results in grey-scale image segmentation are presented. In Sect. 5 the grouping method is applied to the problem of polygonal approximation of curves. Section 6 presents some concluding remarks.

2 Grouping with Hierarchical Graph Structures

A partition of the image plane in a number of regions can be represented by a *region adjacency graph*. The vertices of this graph represent the image regions and its edges represent the adjacency relations between the image regions. The region adjacency graph is an important tool for image segmentation methods

based on region merging. Henceforth, a graph is indicated by the symbol G and its vertex and edge sets by the symbols V and E respectively. If $G = (V, E)$ is a graph, a subset H of V is called *connected* (with respect to G) if, for all $x, y \in H$, there is a sequence $x = x_0, \ldots, x_n = y$ such that $(x_i, x_{i+1}) \in E$ for all i. The graphs representing the results of two successive steps of the iterative bottom-up grouping procedure are called the *child graph* and the *parent graph* respectively. The parent graph represents the result of the grouping procedure applied to the child graph.

An iteration step in a bottom-up grouping procedure involves three stages.

1. *Group selection.* Groups of vertices in the child region adjacency graph are selected to be merged (note that groups may consist of a single element). These groups are chosen such that
   - each group of vertices is a connected subset of G, and
   - the image regions obtained by merging the regions corresponding to the individual nodes in the group (i) closely fit the region-model and (ii) do not overlap.
2. *Vertex construction.* The region adjacency graph is transformed such that each group of vertices (the *children*) in the child graph maps to a single vertex (the *parent*) in the resulting parent graph. A parent represents an image region which equals the union of the regions corresponding to its children.
3. *Edge construction.* Vertex pairs of the parent graph that represent adjacent image regions are joined by edges. Two parent vertices are adjacent if and only if they have adjacent child vertices.

Formally, the iteration step can be defined as follows:

Definition 1. Let $G = (V, E)$ be a graph and let $H_1, \ldots, H_k$ be connected subsets of V such that $H_i \cap H_j = \emptyset$ for $i \neq j$ and $\cup H_i = V$. A graph $G' = (V', E')$ is said to *result from a grouping step in G according to the groups $H_1, \ldots, H_k$* if there is a function $\phi : V \rightarrow V'$ such that

1. $\phi(V) = V'$;
2. $\phi(x) = \phi(y)$ for each $x, y \in V$ if and only if there is a group H_i which contains x and y;
3. two vertices x' and y' in G' are neighbours ($(x', y') \in E'$) if and only if there are two vertices x and y in V such that $\phi(x) = x'$, $\phi(y) = y'$ and $(x, y) \in E$.

The first condition implies that all vertices in the parent graph are derived from the vertices of the child graph. The second condition implies that each vertex in the parent graph corresponds exactly to one of the groups H_i. The third condition describes adjacency between vertices in the parent graph.

Starting with the initial graph representing the input image, a hierarchy of graphs is built by recursive application of the grouping process.

Definition 2. A *hierarchy of graphs* is a sequence $(G_0, \ldots, G_n)$ of graphs and a sequence $(\phi_0, \ldots, \phi_{n-1})$ of mappings $\phi_i : V_i \rightarrow V_{i+1}$ such that:

1. for $i = 0, \ldots, n-1$, $\phi_i(V_i) = V_{i+1}$;
2. for each $i = 0, \ldots, n-1$ and each $x \in V_{i+1}$, $\phi_i^{-1}(x)$ is a connected subset of G_i;
3. for each $i = 0, \ldots, n-1$ and $x, y \in V_{i+1}$, $(x, y) \in E_{i+1}$ if and only if there are $x' \in \phi_i^{-1}(x)$ and $y' \in \phi_i^{-1}(y)$ such that $(x', y') \in E_i$.

For $x \in V_i$, the vertex $\phi_i(x) \in V_{i+1}$ is the parent of x; the vertices in $\phi_{i-1}^{-1}(x)$ are the children of x.

Montanvert *et al.* [12] and Jolion and Montanvert [8] presented merging schemes which use a transformed version of the region adjacency graph. This adapted graph is constructed from the region adjacency graph by deleting edges between regions if the difference between their average grey-levels exceeds a given threshold. In the adapted graph, a grouping is performed. Their method requires the selection of a subset of vertices. The vertices in this subset are called *surviving* vertices, because each vertex in the parent graph corresponds to one of these vertices. Different selection criteria [11, 12] result in different image segmentations. Clusters are formed by assigning the non-surviving vertices to surviving ones. A merging step is then performed by taking the union of the regions in each cluster. In terms of Definition 1, the merging scheme of Montanvert *et al.* [12] and Jolion and Montanvert [8] transforms a graph (V, E) to a graph (V', E') such that V' equals the set of survivors; the mapping ϕ as described in Definition 2 satisfies $\phi(v) = v$ if $v \in V'$.

A large number of groups is indeed merged in each step and the assignment of non-survivors to survivors can be performed through local computation if the set V' satisfies the following two properties:

1. no two vertices in V' are adjacent;
2. each non-surviving node has at least one surviving node as a neighbour.

In graph theory, a set satisfying these properties is called a *maximal independent set* [6].

The following method [12] can be used for the selection of a maximal independent set in a graph. Each vertex is given some random label ξ from the interval $[0, 1]$. Vertices which have a larger label than all their neighbours are selected as members of the maximal independent set. Their neighbours are rejected. As a result, two neighbouring vertices of the graph cannot both be members of the maximal independent set. It is possible that there are unselected nodes which are not adjacent to a selected node. In this case, all vertices that have been neither selected nor rejected are attributed a new random label, and the selection procedure is repeated until each vertex has been either selected or rejected. The selection process usually converges after a few iteration steps.

In the segmentation methods of Montanvert *et al.* [12] and Jolion and Montanvert [8], the merging process continues until the grey-level difference between every pair of adjacent nodes exceeds the threshold. These methods are related to region growing, because the surviving vertices in each stage act as seeds in a merging step.

3 Evaluating Candidate Pairs

This section presents a grouping scheme in which each pair of adjacent vertices is considered for merging. For each candidate pair, the fit of the union of the two vertices to the region model is computed. This information is used to select those pairs that are actually grouped, as described in the previous section.

The selection of pairs for merging poses a transitivity problem. If, for example, three vertices x, y and z are mutually adjacent, both (x, y) and (y, z) are candidates for merging. However, they cannot both be selected, because y is a member of both groups, and only two nodes can be merged at a time. Each choice of a number of pairs to be grouped corresponds to the selection of a subset of edges in the region adjacency graph. An admissible choice of pairs, for which no transitivity conflicts occur, corresponds to a set of edges such that no vertex in the graph lies on more than one edge in the subset. In graph theory, such a set of edges is called a *matching* [6].

The procedure which selects a set of pairs must not only find a matching, but it must also select pairs that give rise to a segmentation that is optimal in a certain sense. The *line graph* [6] is used to perform the selection.

Definition 3. Let $G = (V, E)$ be a graph. The *line graph* $L(G) = (V', E')$ is the graph for which each vertex $v' \in V'$ corresponds to an edge $(v_1, v_2) \in E$ such that two vertices v' and $w' \in V'$ are connected by an edge if and only if the corresponding edges $\{v_1, v_2\}$ and $\{w_1, w_2\} \in E$ share a common point.

A matching in the graph G corresponds to an independent set in the line graph $L(G)$. The selection of a suitable independent set is partially performed by the following deterministic algorithm:

1. Each vertex in the line graph is associated with a *merge score*, which is equal to the error of the associated candidate pair with respect to the region model.
2. Vertices which have a merge score above a threshold value t are not selected.
3. Of the remaining vertices, those having a smaller merge score than their neighbours are selected; their neighbours are all rejected.

This procedure never selects a pair of adjacent vertices. The selected pairs are used in the region-merging scheme of Beveridge [1]. There may be large parts of the graph in which no local extrema occur, such that no vertices are selected. For these regions, the random symmetry breaking method of Montanvert *et al.* [12] is used.

The merge score associated with each vertex in the line graph is a measure for the quality of the union of the corresponding regions with respect to the homogeneity model. Vertices with a low merge score should have a large chance of being selected. This is achieved by attributing to each vertex a random label, using a distribution that depends on the error associated with that vertex. The distribution is chosen such that vertices with a small error value have a large chance of drawing a high number, and therefore have a large chance of being

selected. Suppose that a vertex is labelled with error e and the rejection threshold is t. Then that vertex draws a random number from $[0, 1]$ from a distribution with a probability density function of the form $p(x) = ax + b$ for $0 \leq x \leq 1$. The slope a is chosen to be $2e/t - 1$ and the constant b is used for normalization.

Montanvert *et al.* [12] used random variables that do not depend on local image properties. Jolion and Montanvert [8], on the other hand, used labels which are derived deterministically from a local image property. The first strategy does not use the available image information at each stage in the grouping process, while the second one may cause problems, for example in homogeneous regions. The strategy proposed here is a compromise between these two strategies.

4 Grey-level Image Segmentation

This section presents the application of the grouping procedure described in the previous section to grey-level image segmentation. The input image is represented by a 4-connected graph in which each vertex represents an individual pixel.

Two different region models have been tested:

1. In the first model, regions have a *homogeneous grey-value*. The deviation of the data from this model is defined as the root-mean-square error from the best fitting constant, which equals the standard deviation.

2. The second model adopts a *linear function* of the x- and y-coordinates to approximate the grey-level value in each region. Here, the error is defined as the root-mean-square error with respect to the best fitting linear function. It is not possible to fit a unique optimal linear function through one or two pixels. Also, the best fit for a small number of pixels is noise sensitive. Therefore, the model adopts a homogeneous grey-level value for regions containing less than 10 pixels.

Segmentations based on the constant grey-value model appear less cluttered than segmentations based on the linear model, even when they are calculated using the same thresholds. This seems remarkable, since the linear model can provide a closer fit to a given region than the constant grey-level model. This results from incorrect decisions made in the initial stages of the segmentation process. If there is a weak step edge present in the image, it is not possible to find a constant grey-level region which contains the step edge and fits the data reasonably well. It is possible, however, to find a region with a linear grey-level which contains the step edge and still fits the data reasonably well. This is especially true if the step edge is blurred. Therefore, step edges are contained in the regions that are formed in the early stages of the grouping process. As a result, in the later stages it is no longer possible to find the "true" regions in the image.

The chance that pairs which contain a boundary will be selected is reduced by increasing the merge score of pairs that correspond to linear functions with

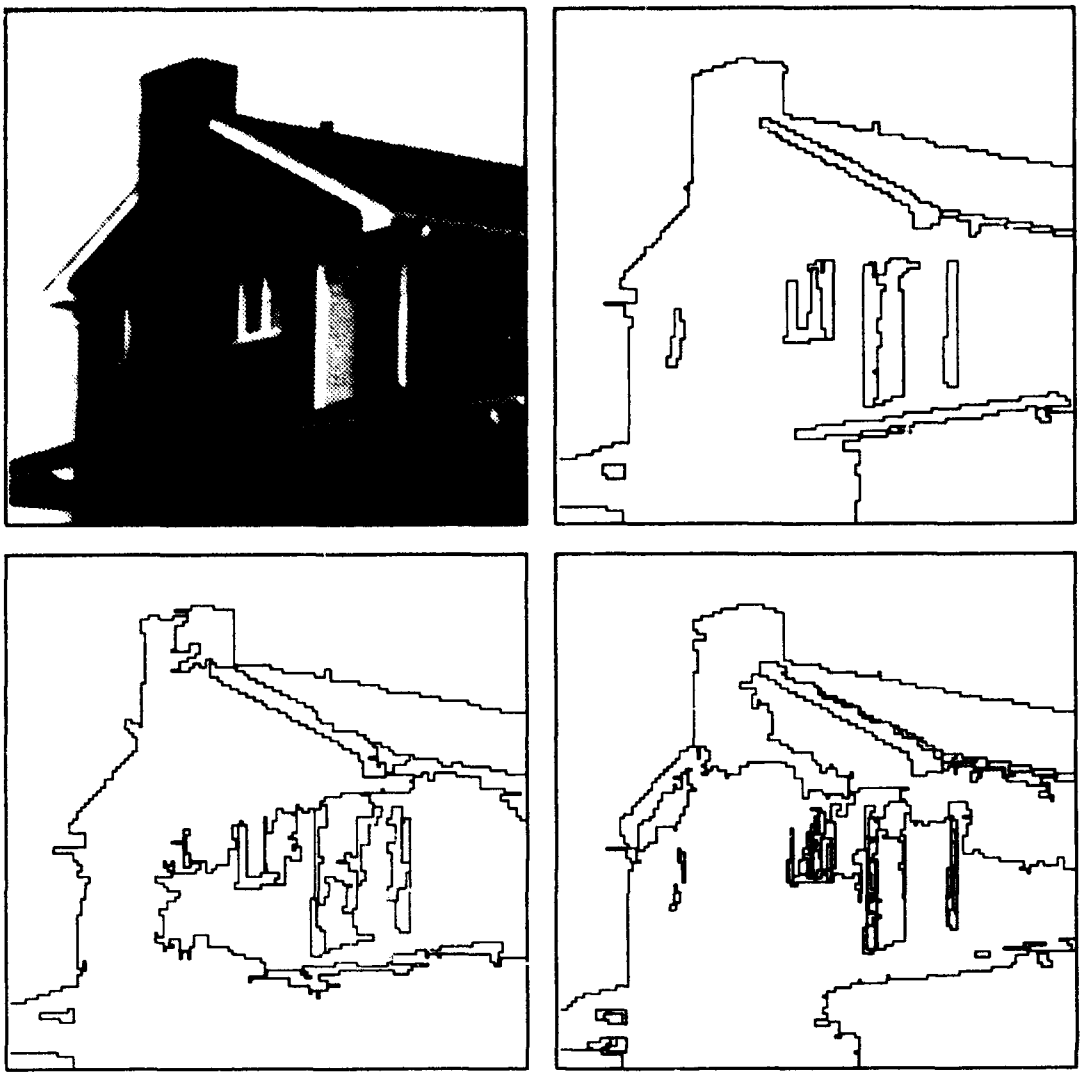


Fig. 1. The original image (top left) and segmentations obtained with the constant grey-level model (top right), the linear model (bottom left) and the linear model (bottom right) with limited slope.

a large slope. This is done by adding the magnitude of the gradient of the best fitting linear model, multiplied by some weight factor, to the error for each candidate group. This reduces clutter on larger scales, but introduces more clutter in small image details. Clutter in small details occurs because such details can correspond to regions with a high slope, while the error (with respect to both constant grey-level and linear models) of such a region can be small.

The segmentation results for a natural image are shown in Fig. 1. The image consists of 256×256 pixels with 256 possible grey-values. The threshold for the error values was set to 18 grey-level units. For the improved linear model, the weight factor for the gradient magnitude was set to 10. A nice looking segmentation is (for this image) produced by the constant grey-level model. The

linear model produces much large-scale clutter. The linear model, adapted to discourage large slopes, has problems with smaller regions.

Grey-level models and error measures should be carefully chosen: an unfortunate choice of models and measures can cause noise pixels to survive as isolated groups in an otherwise homogeneous region. This problem does not occur in practice with the measures described here.

5 Polygonal Approximation of Curves

The grouping process described in the previous sections can also be applied to the polygonal approximation of curves. In this case, the vertices of the graph represent line segments. Each vertex in the initial graph represents a line segment of unit length. The initial graph can be derived from a chain code description [5]. The edges are defined from the adjacency relations along the curve. Each vertex which is not an end point of the curve has exactly two adjacent vertices.

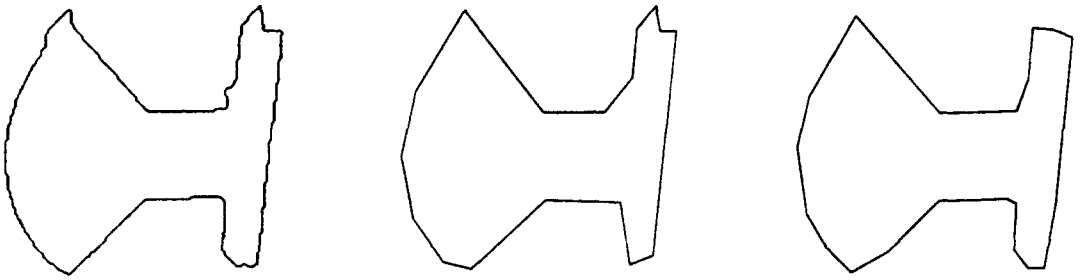


Fig. 2. From left to right: the original curve, the segmentation achieved using the measure of Wall and Danielsson [14], the segmentation achieved using the measure of Borgefors [2].

Each pair of adjacent line segments is evaluated as a possible group. Two such line segments share an end point. The union of these two line segments is the line segment that connects their unshared end points. There are several measures for determining the quality of the approximation of a curve by a line segment (e.g. [14] and [2]).

A curve segment is approximated by the line segment connecting its end points. The curve and the line segment cut out regions from the plane. The difference between the area of the regions to the right of the line segment and those to its left is called the *signed area* between the line and the curve. The error measure proposed by Wall and Danielsson [14] is the absolute value of the signed area between the line segment and the curve, divided by the length of the line segment. A nice property of this measure is that the signed area between the union of two line segments and the curve can be calculated from the signed areas between the two individual line segments and the curve. Therefore, the error of a candidate pair can be computed from the properties of the constituting line

segments, without having to consider the curve itself. The measure proposed by Borgefors [2] is based on the closest distance of points on the line segment to points of the curve. It is calculated for each point on the line segment, and averaged over the line segment. The error measure of Borgefors can be computed efficiently from the distance transformation of the original curve.

The construction of a polygonal curve approximation starts with the construction of an initial graph that contains "line segments" of the size of a single pixel. Clustering is performed until the error measure of all pairs of adjacent line segments exceeds a threshold. Figure 2 shows the result of the polygonal curve approximation procedure. Both the error measures of Wall and Danielsson and of Borgefors were used. The threshold was 2.5 pixel sizes for the Wall and Danielsson measure and 5 pixel sizes for the Borgefors measure. The size of the complete image was 128×128 pixels.

Note that the curve approximation contains a stochastic component, such that different runs of the same algorithm produce slightly different segmentation results. There are no major differences between the outcomes of different runs based on the same error measure. Differences between the two error measures can be seen in the top right and bottom right of the curve. These effects can be understood by considering the approximation of a curve consisting of two line segments AB and BC by the line segment AC. The segment AC can be regarded as the base of the triangle ΔABC with the distance from B to the line AC as its height. According to Wall and Danielsson the error of this approximation is proportional to the height of this triangle. The error according to Borgefors is proportional to the average distance of a point on AC to the curve ABC. If the height is small, this is proportional to the height of the triangle, but if the height is large and the angle ABC is sharp, it is proportional to the length of AC. Therefore, sharp corners are followed more exactly if the error measure of Wall and Danielsson is used, while obtuse angles are traced more precisely when the error measure of Borgefors is used.

6 Conclusions

A new grouping scheme for grey-level image primitives has been presented. The scheme has been described in a hierarchical graph context. This method calculates for each pair of adjacent vertices the fit of their union to the region model. This information is used to select the pairs that are actually grouped. The new technique has been applied to grey-level image segmentation and to curve segmentation.

In contrast to previous stochastic methods, the image content guides the clustering process, not through adaptation of the region adjacency graph, but through merge scores represented in the line graph. For the selection of pairs, the maximal independent set algorithm by Montanvert *et al.* [12] is extended by assigning to each vertex a random label from a different distribution.

There are several ways in which the method presented here can be improved. Problems that arise from the linear grey-level region model can probably be

alleviated by performing a number of grouping steps with the constant grey-level model until large regions have been formed, followed by a number of grouping steps in which the linear model is used. A further improvement can be expected from the use of top-down flow of information (as used for example in relinking methods; e.g. [4]) for the generation and evaluation of candidate groups. This type of information can for instance be derived from the output of an edge detector ([11]). The extension to split-and-merge type of operations and relinking methods also needs to be investigated.

References

1. Beveridge, J.R., Grifith, J., Kohler, R., Hanson, A., Riseman, M. (1989). Segmenting images using localized histograms and region merging, Int. J. of Comp. Vis. 2, pp. 311–347.
2. Borgefors, G. (1988). Hierarchical chamfer matching: a parametric edge matching algorithm, IEEE Trans. Patt. An. Mach. Int. 10, pp. 849–865.
3. Brice, C.R., Fennema, C.L. (1970). Scene analysis using regions, Artificial Int. 1, pp. 205–226.
4. Burt, P., Hong, T.H., Rosenfeld, A. (1984). Image segmentation and region property computation by cooperative hierarchical computation, IEEE Trans. on Systems, Man and Cybern. 12, pp. 611–622.
5. Freeman, H. (1961). On the encoding of arbitrary geometric configurations, IEEE Trans. Elec. Computers 10, pp. 260–268.
6. Harary, F. (1977). Graph Theory, Addison-Wesley, Reading, MA.
7. Horowitz, S.L., Pavlidis, T. (1976). Picture segmentation by a tree traversal algorithm, J. of the Ass. for Computing Machinery 23, pp. 368–388.
8. Jolion, J.M., Montanvert, A. (1992) The adaptive pyramid: a framework for 2D image analysis, Comp. Vis. Graph. Im. Proc.: Image Understanding 55, pp. 339–348.
9. Meer, P. (1989). Stochastic image pyramids, Comp. Vis. Graph. Im. Proc. 45, pp. 269–294.
10. Meyer, F., Beucher, S. (1990). Morphological segmentation, J. of Visual Comm. and Im. Repr. 1, pp. 21–46.
11. Montanvert A., Bertolino P. (1992). Irregular pyramids for parallel image segmentation, Bischof, H. and Kropatsch, W.G. (eds.), Oldenbourg Verlag 1992, pp. 13–35, 16th AGM Meeting, Vienna, Austria, May 5–9.
12. Montanvert, A., Meer, P., Rosenfeld, A. (1991). Hierarchical image analysis using irregular tesselations, IEEE Trans. Patt. An. and Mach. Int. 13, pp. 307–316.
13. Pavlidis, T. (1977). Structural Pattern Recognition, Springer-Verlag, New York.
14. Wall, K., Danielsson, P.-E. (1984). A fast sequential method for polygonal approximation of digitized curves, Comp. Vis. Graph. Im. Proc. 28, pp. 220–227.
15. Zucker, S.W. (1976). Region growing: childhood and adolescence, Comp. Vis. Graph. Im. Proc. 5, pp. 382–399.

Hierarchical Shape Representation for Image Analysis

*O Ying-Lie**

CWI Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

Abstract. Image analysis requires an appropriate description of shape. The structure of shape may be determined by a grouping of parts of the image with certain associated characteristics. A coarse-to-fine structure can be determined by an ordered sequence of hierarchical levels. Three methods of generating an ordered sequence are proposed, that is, based on grey-level images, based on shape primitives, and based on symbolic descriptions. The hierarchical representation is based on symbolic descriptions. This paper aims to generalize the hierarchical approach, and to explain the mathematical background.

Keywords: shape, order relation, grouping, primitive extraction, symbolic description, layered structure.

1 Introduction

The notion of *shape* is rather intuitive, and highly influenced by human vision of objects in real life. The human visual system is capable of recognizing a large variety of objects in different environments and circumstances. In image analysis, only specific tasks are of interest. These tasks are generally based on shapes that can be described by models that contain specific characteristics.

The structure of shape may be determined by a *grouping* of parts of the image with associated characteristics. These characteristics are determined with respect to the spatial domain and the grey-level domain. The description must be invariant with respect to irrelevant transformations and small perturbations.

A mathematical description of shape may be characterized thus:

- It is based on an underlying *topology*, supporting notions such as order relation, neighbourhood property, connectivity, adjacency, and inclusion.
- It supports *invariance* under certain affine transforms, such as translation, rotation, scaling, and deformation.
- It allows *decomposition* into various levels of details.

* Correspondence address: Zwaardemakerlaan 23, 3571 ZA Utrecht, The Netherlands.
The author wishes to acknowledge the referees for their helpful comments.

Two classes of model-based methods may be distinguished: the *geometric model* and the *expansion model*. The geometric model directly reflects the geometrical characteristics, while the expansion model is based on series expansion.

The geometric model of shapes in a two-dimensional image may be

1. *solid-based* based on connectivity of sets with nonempty interiors,
   - *region-based* based on a contiguous domain,
2. *curve-based* based on connectivity of sets with empty interiors,
   - *outline-based* based on a closed boundary curve,
   - *skeleton-based* based on a main curve with branches,
3. *point-based* based on distinct points,
   - *landmark-based* based on identifying features.

A similar classification for statistical shape analysis is presented in [5].

The invariance properties determine equivalence classes of similar shapes. These classes define the characteristics of shapes that are identified as equal. A coarse-to-fine structure can be determined by an ordered sequence of hierarchical levels. The equivalence classes increase with a coarser representation.

In Sect. 2, a structured method of generating hierarchical ordered sequences is depicted. Then, Sects. 3, 4, and 5 delineate the different types of sequences and the operations that are involved. In Sect. 6, it is explained how the hierarchical representation can be obtained. Finally, Sect. 7 concludes with some familiar examples.

2 Hierarchical Levels

The hierarchical representation is a sequence of levels, such that

- higher levels contain fewer shape characteristics than lower levels,
- successive levels are associated.

The hierarchical levels are *ordered*, that is, there exists a *hierarchical order relation* $\succeq$ that satisfies the *preorder* properties

reflexivity $L1 \succeq L1$,

transitivity $L1 \succeq L2$ and $L2 \succeq L3$ imply that $L1 \succeq L3$,

where $L1$,$L2$, and $L3$ are hierarchical levels.

The hierarchical levels of an image are given by an ordered sequence that satisfies the preorder properties, and

linearity $L_m \succeq L_n$ or $L_n \succeq L_m$,

where L_m and L_n are members of the sequence.

From an initial grey-level image, three types of ordered sequences can be generated: the *sequence of grey-level images* $\{F_n\}$, the *sequence of primitive extractions* $\{G_n\}$, and the *sequence of symbolic descriptions* $\{H_n\}$.

These sequences can be obtained by application of some of the following operations:

grey-level filter	$F_n = \phi_n(F_{n-1}),\ n = 1, \ldots, N,$
primitive extraction	$G_n = \zeta_n(F_n),\ n = 0, \ldots, N,$
primitive filter	$G_n = \psi_n(G_{n-1}),\ n = 1, \ldots, N,$
symbolic description	$H_n = \xi_n(G_n),\ n = 0, \ldots, N,$
symbolic filter	$H_n = \chi_n(H_{n-1}),\ n = 1, \ldots, N,$

where the subscript n indicates the hierarchical level, and the subscript 0 indicates the initial level.

A *filter* reduces the amount of shape characteristics by removing details. In particular, a grey-level filter decreases the variation of the grey-levels. A primitive or symbolic filter eliminates undesirable primitives or symbols.

A *primitive extraction* operation extracts shape characteristics of a different nature than the grey-level image. Examples of primitives are: extrema, differentials, line segments, and curve segments. A *symbolic description* provides a representation of the primitive extraction that is suitable for image analysis. An example of a widely used symbolic description is a graph.

The methods of generating the sequences are illustrated by the following diagrams.

Generation of the *sequence of grey-level images* from an initial image:

$$\begin{array}{lccccccc} \text{(a)} & & F_0 & \xrightarrow{\phi_1} & F_1 & \longrightarrow \ldots \longrightarrow & F_{n-1} & \xrightarrow{\phi_n} F_n \\ \text{(b)} & \zeta_0 \downarrow & & \downarrow \zeta_1 & & & \zeta_{n-1} \downarrow & \downarrow \zeta_n \\ & G_0 & & G_1 & \ldots & & G_{n-1} & G_n \\ \text{(c)} & \xi_0 \downarrow & & \downarrow \xi_1 & & & \xi_{n-1} \downarrow & \downarrow \xi_n \\ & H_0 & & H_1 & \ldots & & H_{n-1} & H_n \end{array} \tag{1}$$

Generation of the *sequence of primitive extractions* from an initial image:

$$\begin{array}{lccccccc} & & F_0 & & & & & \\ \text{(a)} & \zeta_0 \downarrow & & & & & & \\ \text{(b)} & G_0 & \xrightarrow{\psi_1} & G_1 & \longrightarrow \ldots \longrightarrow & & G_{n-1} & \xrightarrow{\psi_n} G_n \\ \text{(c)} & \xi_0 \downarrow & & \downarrow \xi_1 & & & \xi_{n-1} \downarrow & \downarrow \xi_n \\ & H_0 & & H_1 & \ldots & & H_{n-1} & H_n \end{array} \tag{2}$$

Generation of the *sequence of symbolic descriptions* from an initial image:

$$\begin{array}{lccccc} & & F_0 & & & \\ \text{(a)} & \zeta_0 \downarrow & & & & \\ & G_0 & & & & \\ \text{(b)} & \xi_0 \downarrow & & & & \\ \text{(c)} & H_0 & \xrightarrow{\chi_1} & H_1 & \longrightarrow \ldots \longrightarrow & H_{n-1} \xrightarrow{\chi_n} H_n \end{array} \tag{3}$$

The *hierarchical representation* is obtained from the sequence of symbolic descriptions by a *hierarchical association* ρ_n:

$$\text{(d)} \qquad H_0 \xleftrightarrow{\rho_1} H_1 \quad \longleftrightarrow \ldots \longleftrightarrow \quad H_{n-1} \xleftrightarrow{\rho_n} H_n \tag{4}$$

where the "$\longleftrightarrow$" sign denotes the association between primitives in different levels.

3 The Sequence of Grey-level Images

A grey-level image can be considered as a two-dimensional function. The distribution of the grey-levels determines the shape in the image.

Let $F : \mathrm{D} \to \mathrm{T}$; $F : x \mapsto t$ denote a *grey-level image*, where D is the *spatial domain* and T is the *grey-level domain*. F assigns a unique grey-level value $t \in \mathrm{T}$ to each element $x \in \mathrm{D}$, given by the elements $(x, t) \in F$. The space of grey-level images is $\mathbf{F} = \mathrm{T}^{\mathrm{D}}$. The spatial domain $\mathrm{D} \subset \mathbb{R}^2$ is connected, and the grey-level domain $\mathrm{T} \subset \mathbb{R}$. As a result of the imaging process, D is usually a square and T is bounded. In addition, T may be nonnegative, where the 0 values may indicate lack of information.

The *dual grey-level image* $\check{F}$ is obtained by mirroring the grey-levels with respect to a constant grey-level. As a result, minima become maxima, lower semi-continuous becomes upper semi-continuous and vice versa.

Definition 1. A *grey-level shape* in a grey-level image $\mathsf{Sf} \subset F$ is a grouping of subsets $\mathsf{Sf} = \mathsf{group}(\{f\})$, $f \in \mathcal{P}(F)$, such that (i) the spatial domain $\mathsf{Df} = \mathrm{dom}\,\mathsf{Sf}$ is connected; and (ii) the associated grey-level function $F(\mathsf{Df})$ is upper semi-continuous.

The *domain* Df may be based on the interior (region-based), or on the boundary (outline-based). A grey-level shape may be composed of smaller subsets which may overlap. The *grouping* of the subsets is determined by connectivity (or adjacency) of the domains. Shape *inclusion* $\mathsf{Sf1} \subseteq \mathsf{Sf2}$ is inclusion of the domains $\mathsf{Df1} \subseteq \mathsf{Df2}$. Note that this is different from set inclusion of a function [3].

"Important" shapes are called *foreground* shapes, and "unimportant" shapes are called *background* shapes. The term shape is generally used for foreground shapes. If the image is considered as a smooth geographic surface, then pronounced "hills" are foreground shapes while the remainder are background shapes. Hence, if the shapes of interest are the "valleys", then it is more convenient to use the dual image $\check{F}$.

Shape *invariance* properties with respect to transformations on the spatial domain and grey-level domain such as translation, rotation, and scaling are desirable. Specifically, the "shape" of a grey-level function is invariant with respect to grey-level translation.

The *sequence of grey-level images* is generated from the initial image by application of a sequence of grey-level filters (diagram (1), step (a)).

A *grey-level filter* is a *mapping* $\phi : \mathbf{F} \to \mathbf{F}$ that satisfies the following basic properties:

increasing	$F1 \succeq F2$ implies that $\phi(F1) \succeq \phi(F2)$,
anti-extensive	$F \succeq \phi(F)$,
neighbourhood	for each $(x1, t1)$, $(y1, s1) \in F$ with corresponding $(x2, t2)$, $(y2, s2) \in \phi(F)$, $x1, y1 \in O1 \subset \mathrm{dom}\,F$ implies that $x2, y2 \in O2 \subset \mathrm{dom}\,\phi(F)$. $O1$ and $O2$ are open neighbourhoods.

An example of the neighbourhood property is n-isomorphism [8].

The invariance properties of most filters are usually limited to translation and rotation. Scaling is only included if the filter depends on the scaling parameter.

The *sequence of grey-level filters* $\{\phi_n\}$, $F_n = \phi_n(F_{n-1})$, $n = 1, \ldots, N$, that generates the *sequence of grey-level images* $\{F_n\}$ possesses additional properties that determine the association between grey-level images:

order	ϕ_n induces an order relation such that $F_{n-1} \succeq F_n$,
scale	ϕ_n depends on a monotonic one-parameter family $s_n \geq 0$ such that $F_n \succeq F_m$ if $n < m$, $\phi_0 = 1$,
evolutionary	$\lvert d(F_n, F_m)\rvert \leq \epsilon_{n,m}$ for some $\epsilon_{n,m} \geq 0$ that depends on n, m,
composition	$\phi_n \phi_m = \phi_{\nu(n,m)}$, semigroup if $\nu(n,m)$ is linear.

The parameter s is the *scale parameter* that may incorporate scaling invariance properties. The *difference function* $d(.,.)$ indicates the distinction between two hierarchical levels.

4 The Sequence of Primitive Extractions

A shape can be specified by features that reflect the geometrical properties. Certain combinations of these features form a primitive that provides a meaningful description of shape.

Let $G \subset \{(x,u) : x \in \mathrm{D}, u \in \mathrm{U}\}$ denote a *primitive extraction*. G assigns one or more primitives $u \in \mathrm{U}$ to some $x \in \mathrm{D}$ generating the (distinct) elements $(x,u) \in G$. The space of extracted images is $\mathbf{G} \subset \mathrm{D} \times \mathrm{U}$. A primitive is typically a geometric structure that may have parameters. The space of primitives U is therefore usually richer than the grey-level domain T.
The primitive extraction does not generally have a dual representation.

The *primitive extraction* operation is a *mapping* $\zeta : \mathbf{F} \rightarrow \mathbf{G}$; $\zeta : \{(x,t)\} \mapsto (x,u)$ that derives a primitive from the grey-level image. The primitive and its parameters are derived from a subset of the grey-level function. The corresponding spatial domain, the primitive domain, will be denoted by $\mathrm{dom}\, u$. Domains of different primitives may overlap. The operation is generally nonlinear, and preserves the neighbourhood property:

neighbourhood	for each $(x1,t1)$, $(y1,t2) \in F$ with corresponding $(x2,u1)$, $(y2,u2) \in \zeta(F)$, $x1, x2 \in \mathrm{dom}\, u1$ and $y1, y2 \in \mathrm{dom}\, u2$, $x1, y1 \in O1 \subset \mathrm{dom}\, F$ implies that $x2, y2 \in O2 \subset \mathrm{dom}\, \zeta(F)$. $O1$ and $O2$ are open neighbourhoods.

Definition 2. A *primitive shape* in a grey-level image $\mathsf{Sg} \subset G$ is a grouping of primitives $\mathsf{Sg} = \mathsf{group}(\{g\})$, $g \in \mathcal{P}(G)$, such that (i) the spatial domain $\mathsf{Dg} = \mathrm{dom}\, \mathsf{Sg} \subseteq \mathsf{Df}$; and (ii) each primitive $(x1,u1) \in \mathsf{Sg}$ is associated with at least one other primitive $(x2,u2) \in \mathsf{Sg}$ in a specific manner.

The *domain* Dg may be given by the union of the primitive domains (region-based or outline-based), or by the corresponding distinct points (landmark-based). The *grouping* is determined by connectivity of the primitive domains, and by properties and parameters of the primitives. The shape *inclusion* is not as clear.

The primitive extraction operation preserves the invariance properties mentioned earlier for the grey-level shape. In fact, it is frequently used to enhance

these properties. Primitives with special invariance properties that are useful for image analysis are frequently denoted by "invariants".

Different grey-level shapes may result in the same primitive shape, yielding an *equivalence class* of *primitive-similar* shapes. It satisfies the reflexivity, symmetry, and transitivity properties.

Two grey-level shapes Sf1 and Sf2 are primitive-similar $\mathrm{Sf1} \overset{\zeta}{\simeq} \mathrm{Sf2}$ if $\zeta(\mathrm{Sf1}) = \zeta(\mathrm{Sf2})$.

Specifically, a class of deformations such that $A(\mathrm{Sf}) \overset{\zeta}{\simeq} \mathrm{Sf}$ may exist.

The *sequence of primitive extractions* may be generated (i) from the sequence of grey-level images by performing primitive extraction operations on each level (diagram (1), step (b)); or (ii) from an initial primitive extraction by application of a sequence of primitive filters (diagram (2), step (b)).

In method (i), the order relation of the grey-level images is preserved if the primitive extraction operation ζ is increasing.

The *primitive filter* is a *surjective mapping* $\psi : \mathbf{G} \to \mathbf{G}$ that is increasing, anti-extensive, and preserves the neighbourhood property and the invariance properties of the primitive extraction operation ζ. The surjective property implies that primitives may be eliminated. Primitives can be altered according to filter conditions $\{(x,u)_{n-1} \mapsto (x,u)_n \mid R_\psi\}$ as follows

$(x,u)_n = (x,u)_{n-1}$, (preservation)
$(x,u)_n = \emptyset$, (elimination)
$(x,u)_n \neq (x,u)_{n-1}$, (modification)
$\{(x,u)_{n-1}\} \mapsto (x,u)_n$, (merging)

where the filter conditions R_ψ depend on the parameters of the primitive.

The *sequence of primitive filters* $\{\psi_n\}$, $G_n = \psi_n(G_{n-1})$, $n = 1, \ldots, N$, that generates the *sequence of primitive extractions* $\{G_n\}$ has additional properties that determine the association between primitive extractions:

order ψ_n induces an order relation such that $G_{n-1} \succeq G_n$,
scale ψ_n depends on a monotonic one-parameter family $s_n \geq 0$ such that $G_n \succeq G_m$ if $n < m$, $\psi_0 = 1$,
evolutionary $|d(G_n, G_m)| \leq \epsilon_{n,m}$ for some $\epsilon_{n,m} \geq 0$ that depends on n, m.

The scale parameter s is determined by the filter conditions R_ψ. The same consideration applies to the difference function $d(.,.)$, which is not always easy to determine.

5 The Sequence of Symbolic Descriptions

The structure of shape can be described by primitives and associations between these primitives. These elements can be represented by a set of symbols.

Let $H \subset \{v, e : v \in \mathrm{V},\ e \in \mathrm{E}\}$ denote a *symbolic description*. H consists of (distinct) primitive symbols $v \in \mathrm{V}$ and interrelation symbols $e \in \mathrm{E}$ that associate pairs of primitive symbols. Both types of symbols may have parameters. The space of symbolic descriptions is $\mathbf{H} \subset (\mathrm{V} \cup \mathrm{E})$. The spaces of primitive symbols V and interrelation symbols $\mathrm{E} \subset \mathrm{V} \times \mathrm{V}$ are finite.

The *symbolic description* operation $\xi : \mathbf{G} \rightarrow \mathbf{H}$ is composed of a *bijective mapping* $\eta : \mathbf{G} \rightarrow \mathrm{V}$; $\eta : (x, u) \mapsto v$ and symmetric assignments $e(v1, v2) \in \mathrm{E}$. The first mapping assigns a primitive symbol v to each primitive (x, u). The second expression assigns an interrelation symbol e to each pair of associated primitive symbols $(v1, v2)$. Parameters of the primitives are passed to the primitive symbols. Interrelation symbols may also have parameters that are derived from parameters of the associated primitive symbols.

The *dual symbolic description* $\check{H}$ is obtained by replacing interrelation symbols with primitive symbols, and primitive symbols with interrelation symbols. The use of the dual representation may be worthy of further analysis.

Definition 3. A *symbolic shape* in a grey-level image $\mathrm{Sh} \subset H$ is a grouping of primitive symbols $\mathrm{Sh} = \mathrm{group}(\{h\})$, $h \in \mathcal{P}(H)$, such that the primitive symbols $v \in V$ are connected by interrelation symbols $e \in E$ forming certain connection paths.

The connection paths satisfy the symmetry and transitivity properties. The path is determined by the geometric model:

each symbol is connected to all other symbols by paths (region-based),
each symbol is connected by a closed path (outline-based),
symbols are connected by a main path with branches (skeleton-based),
only associated symbols may be connected (landmark-based).

Both *grouping* and *inclusion* are determined by connectivity between primitive symbols. This provides an adequate and explicite way to describe shape.

In accordance with the foregoing, an *equivalence class* of *symbolic-similar* shapes may be defined.

Two symbolic shapes Sg1 and Sg2 are symbolic-similar $\mathrm{Sg1} \overset{\xi}{\cong} \mathrm{Sg2}$ if there is a one-to-one correspondence between each $v1 \in \mathrm{Sg1}$ and $v2 \in \mathrm{Sg2}$, and each corresponding interrelation.

A similar notion in graph theory is known as graph isomorphism [2].

The *sequence of symbolic descriptions* may be generated (i) from the sequence of primitive extractions by performing symbolic description operation on each level (diagram (1), step (c)), or similarly (ii) (diagram (2), step (c)); or (iii) from an initial symbolic description by application of a sequence of symbolic filters (diagram (3), step (c)).

In methods (i) and (ii), the order relation of the primitive extractions is preserved if the symbolic description operation ξ is increasing.

The *symbolic filter* is a *surjective mapping* $\chi : \mathbf{H} \rightarrow \mathbf{H}$ that satisfies similar basic properties and conditions to the primitive filter ψ with conditions on the interrelation symbols $\{(e(v1, v2))_{n-1} \mapsto (e(v1, v2))_n \mid R_\chi\}$ such that

$e_n(v1, v2)_n \neq e_{n-1}(v1, v2)_{n-1}$ (reassignment).

Changes in the primitive symbols result in appropriate reassignment of the associated interrelation symbols. After elimination of a primitive symbol, path connectivity can be preserved by assigning additional interrelation symbols.

The *sequence of symbolic filters* $\{\chi_n\}$, $H_n = \chi_n(H_{n-1})$, $n = 1, \ldots, N$ that generates the *sequence of symbolic descriptions* $\{H_n\}$ has properties that determine the association between symbolic descriptions as specified previously for

the sequence of primitive filters $\{\psi_n\}$.
The scale parameter s is determined by the filter conditions R_χ. The same consideration applies to the difference function $d(.,.)$, which is not always as clear.

6 Hierarchical Representation

The hierarchical representation is a hierarchical structure that is obtained from the sequence of symbolic descriptions by associations of successive levels. The hierarchical association determines parent-child links between primitive symbols in these levels. The parent symbol is in a higher level than the child symbol.

The *hierarchical association* $\rho_n(H_n, H_{n-1})$, $n = 1, \ldots, N$, assigns *parent-child links* between pairs of primitive symbols $v_n \in V_n \subset H_n$ and $v_{n-1} \in V_{n-1} \subset H_{n-1}$, denoted by $\{r_n(v_n, v_{n-1})\}$.

The child-symbol can be inferred from the parent-symbol by linking conditions $\{v_n, v_{n-1} \mapsto r_n(v_n, v_{n-1}) \mid R_\rho\}$. According to these conditions, previously established links may be reassigned (relinking). In the case that the symbolic descriptions are generated by symbolic filters, the sequence of hierarchical associations $\{\rho_n\}$ directly results from the sequence of symbolic filters $\{\chi_n\}$.

The *hierarchical structure* is a *layered structure*. Links between symbols in successive levels are indicated by *directed arcs* from parent to child.

A special type of a hierarchical structure is the *tree* [2]. A tree has one root symbol that has no connection to a parent symbol, branches that connect one parent symbol with child symbols, and leaf symbols that have no connection with child symbols. The root is the ancestor, and descendants are therefore parts of the same shape. Grouping is indicated by the connections of child symbols to the same parent symbol. The structure of shapes in an image can be described by several trees, in which:

roots may be located on different levels	(vanishing),
symbols may belong to more than one tree	(overlapping),
symbols may not belong to any tree	(independence).

7 Applications

A number of useful applications are segmentation, recognition, and compression. *Segmentation* divides the grey-level image into disjunct connected subsets. It prevents the overlapping property. These segments can be regarded as "shapes" in the image. *Recognition* is carried out by comparing the shapes with an archetype. For this purpose, the description must be comprehensive, and suitable for algorithmic use. *Compression* is a method to reduce the amount of data for storage or transmission. The grey-level image must be recoverable from its representation, up to a certain equivalence class.

The number of hierarchical levels is determined by the number of meaningful shape characteristics required for the aimed purpose. A grey-level image with a constant value, a primitive extraction containing a single primitive, or a symbolic description without interrelation symbols may not be useful.

7.1 Generation of a Sequence of Grey-level Images

The sequence of grey-level images can be generated by a system of evolutionary equations. The most popular method is scale-space, satisfying the diffusion equation $\Delta F(x, s) = F_s(x, s)$, where s is the scale. The solution depend continuously on the scale and is governed by the maximum principles [16]. This property is valid for several types of parabolic equations, a few of which are presented in this volume [10, 1, 6].

The method of deriving a hierarchical representation from a *sequence of grey-level images* is delineated in diagrams (1) and (4).

(a) The *grey-level filter* is obtained by convolution with a Gaussian kernel

$$F_n = \Phi_n(F_0) = \int F_0(y)\, k(x - y, s_n)\, dy \ , \quad k(x, s_n) = c\, \exp(-\frac{\|x\|^2}{s_n}) \ ,$$

where $\|.\|$ is the l_2 norm. The filter is linear, translation and rotation invariant, and satisfies the order, scale, evolutionary, and semigroup properties. The maximum principle implies the neighbourhood property. The order relation is induced by the scale, $F_{n-1} \succeq F_n$ if $s_{n-1} < s_n$. The difference function is given by the grey-level range $d(F_n, F_m) = |T_n - T_m|$, $T_n = \max F_n - \min F_n$.

(b) The *primitive extraction* method is based on the evolutionary properties. Mostly, it only concerns one primitive type, given by the set of spatial and grey-level values $G_n = \{(x, t) \mid s = s_n\}$.
- *Region-based.* An *extremal region* or "*blob*" is defined by the function surrounding a maximum M that exceeds a threshold level $g_M = \{F(x) \mid F(x) > \tau,\ M = \max g_M(x)\}$. The threshold value τ can be determined by a delimiting saddle point [11]. These "blobs" are "hill-like" foreground shapes.
- *Outline-based.* The *zero-crossings* are surfaces that are solutions of the Laplacian-of-Gaussian equation $\Delta F = 0$. This method is suitable for data compression [4].
- *Landmark-based.* The behaviour of extrema as a function of the scale, *extremum following* [7] are curves that are solutions of $|\nabla F| = 0$.

(c) The *symbolic description* associates sets of function values with primitive symbols, $H_n = \{(v_n, x0, t0) \mid s = s_n\}$, and derives $(x0, t0)$ as principal parameters of v_n. In the region-based example each blob is a primitive symbol, and the maximum determines the principal parameters.

(d) The *hierarchical representation* gives the loci of primitive symbols as a function of the scale by connecting the principal parameters. The resulting surfaces or curves may merge or vanish.

7.2 Generation of a Sequence of Primitive Extractions

Most shape description methods are based on binary images. A method that is also suitable for grey-level images is mathematical morphology. An overview of

the theory and application on shape description is delineated in [3]. The primary operations are dilation and erosion

$$\delta(X) = X \oplus B = \bigcup_{b \in B} X_b \quad \text{and} \quad \varepsilon(X) = X \ominus B = \bigcup_{b \in B} X_{-b} \ ,$$

where X is a set, the set B is the structuring element, and the subscripts indicate the translates along vectors b, $-b$ respectively. The operations are translation invariant, and reflect how the shape of a set X relates to the shape of the structuring element B.

By using sets as structuring elements, the approach can easily be extended to grey-level functions, yielding the "flat extension" which has the same properties. The method of obtaining a hierarchical representation from a *sequence of primitive extractions* is depicted in diagrams (2) and (4).

(a) The *primitive extraction* method is based on a granulometry [3], that is, a family of openings $\{\alpha_s \mid \alpha_s(X) \supseteq \alpha_r(X) \text{ if } s \leq r\}$, $\alpha_s = \delta_{sB}\varepsilon_{sB}$, where sB is the structuring element scaled by s, and B is a convex structuring element of unit size. An opening is an increasing, anti-extensive operator that satisfies the order, scale, evolutionary, and semigroup properties. The order relation is the set inclusion relation, and the difference function is based on set difference.

(b) The *primitive filter* and the primitive extraction can be combined into a single operation $G_n = \Psi_n(F_0)$ where Ψ is an increasing, anti-extensive operator.

- *Region-based.* Shape decomposition can be obtained from granulometries with structuring elements of different shapes and sizes, given by the recursion formula

$$X^s = X^r \bigcup \bigcup_m \alpha_{sBm}(X \setminus X^r) \ , \quad s < r \ ,$$

where $Bm = B1, B2, \ldots$ are convex structuring elements of unit size. This results in

$$\psi_n(X) = X^s \mid s \geq s_n \ .$$

- *Skeleton-based.* The skeleton subset is given by [3]

$$\Sigma_{sB}(X) = \varepsilon_{sB}(X) \setminus \bigcup_r \alpha_{rB}(\varepsilon_{sB}(X)) \ .$$

The sequence of primitive extractions is given by the skeletons of the eroded sets

$$\psi_n(X) = \Sigma_B(\varepsilon_{s_n B}(X)) = \bigcup_{s \geq s_n} \Sigma_{sB}(X) \ .$$

Curves can also be generated from shape indices of the eroded set $\varepsilon_{s_n B}(X)$, yielding the erosion curve [12].

(c) The *symbolic description* can be expressed as a union of sets $H_n = \bigcup_{s \geq s_n} v$, where the primitive symbols v are determined by the structuring elements. In the region-based example, $v = sBm(x)$ are the scaled structuring elements positioned at x.

(d) The *hierarchical representation* gives the description of shape as a function of structuring elements of decreasing or increasing sizes.

7.3 Generation of a Sequence of Symbolic Descriptions

Suitable representations for symbolic descriptions are graphs. A major advantage of this method is that the theory is well established [2], a dual representation may exist, and algorithms are available. A graph consists of vertices which may be labelled, and edges that may connect pairs of vertices. A pair of vertices that are connected by an edge is called adjacent.

The method of obtaining a hierarchical representation from a *sequence of symbolic descriptions* is given in diagrams (3) and (4).

(a) The *primitive extraction* method must generate distinct primitives and clearly defined interrelations. It may be region-based, curve-based, or landmark-based.

(b) The *symbolic description* is a labelled graph $H \subset \{v, e : v \in \mathrm{V},\ e \in \mathrm{E}\}$, where v are vertices, and e are undirected edges. The vertices are the primitive symbols, and the edges are the interrelation symbols. Both vertices and edges may be labelled and have parameters passed from the primitive extraction.

(c) The *symbolic filter* is a surjective mapping

$$v_n = \chi_n(\{v_{n-1}\})$$

that merges a group of adjacent vertices into one vertex, and reassignes the corresponding edges. It consists of three steps [14]: (i) selection of a group of adjacent vertices (grouping), (ii) mapping of the group into one vertex (merging), (iii) determination of the edges (adjacency).
The grouping is determined by grouping conditions. The number of vertices that meet the conditions can be reduced by favouring specific configurations [14], or by random selection [13]. Vertices that do not belong to a group are eliminated. The merging step determines the order relation $V_{n-1} \succ V_n$. The scale parameter and the difference function are determined by these conditions. The adjacency step generally preserves path connectivity.

- *Region-based.* Region-based methods are generally used for segmentation. The regions are derived from adjacency and similarity of the grey-levels, given by edges $e(d(t1, t2))$ of adjacent vertices $((v1(t1), v2(t2))$, where the difference function $d(t1, t2) = |t1 - t2|$, $t1$ and $t2$ are the grey-levels. Vertices are selected if $d(t1, t2) \leq \tau$. The threshold value τ determines the scale $s_n = \tau n$. The initial level is obtained by dividing the grey-level image into small regions, usually a square grid.

- *Curve-based.* Curve-based methods can be based on several assumptions: connectivity of curve-segments, outline-based or skeleton-based, and geometric forms, allowing particular kinds of connectivity such as intersection and end-point connection.
  Grouping can also be based on spatial sampling. In the $2 \times 2 \setminus 2$ sampling [9], 2×2 squares of the grid are split by their diagonals, and the 4 inner triangles are merged into a new square, rotated by 45 degrees. This induces a spatial scale $s_n = \frac{1}{2}\sqrt{2}\, n$.

(d) The *hierarchical representation* is a layered graph, in which the parent-child links are indicated by directed edges. These links result from the merging operation, yielding a tree structure. A layered graph also allows the decomposition of a shape into parts, for instance by using a binary tree [17].

References

1. van den Boomgaard, R., Smeulders, A.W.M. (1993). Towards a morphological scale-space theory, this volume, pp. 631–640.
2. Harary, F. (1972). Graph Theory, Addison Wesley, Reading, MA.
3. Heijmans, H.J.A.M. (1993). Mathematical morphology as a tool for shape description, this volume,pp. 147–176.
4. Hummel, R., Moniot, R. (1989). Reconstructions from zero-crossings in scale-space, IEEE Trans. ASSP 37, pp. 2111–2130.
5. Kent, J.T., Mardia, K.V. (1993). Statistical shape methodology in image analysis, this volume, pp. 443–452.
6. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1993). Exploring the shape manifold: the role of conservation laws, this volume, pp. 601–620.
7. Koenderink, J.J. (1984). The structure of images, Biological Cybernetics 50, pp. 363–370.
8. Kovalevsky, V.A. (1993). A new concept for digital geometry, this volume, pp. 37–51.
9. Kropatsch, W.G., Willersinn, D. (1993). Irregular curve pyramids, this volume, pp. 525–538.
10. Lindeberg, T.P. (1993). Scale-space for N-dimensional discrete signals, this volume, pp. 571–590.
11. Lindeberg, T.P. (1993). Scale-space behaviour and invariance properties of differential singularities, this volume, pp. 591–600.
12. Mattioli, J., Schmitt, M. (1993). On information contained in the erosion curve, this volume, pp. 177–196.
13. Montanvert, A., Meer, P., Bertolino, P. (1993). Hierarchical shape analysis in grey-level images, this volume, pp. 511–524.
14. Nacken, P., Toet, A. (1993). Candidate grouping for bottom up segmentation, this volume, pp. 549–558.
15. Porteous, I.R. (1981). Topological Geometry, Cambridge Univ. Press., Cambridge, UK.
16. Protter, M.H., Weinberger, H.F. (1967). Maximum Principles in Differential Equations, Prentice-Hall.
17. Segen, J. (1993). Inference of stochastic graph models for 2-D and 3-D shape, this volume, pp. 493–510.

Scale-Space for N-Dimensional Discrete Signals *

Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computing Science
Royal Institute of Technology, S-100 44 Stockholm, Sweden
Email: tony@bion.kth.se

Abstract. This article shows how a (linear) scale-space representation can be defined for discrete signals of arbitrary dimension. The treatment is based upon the assumptions that (i) the scale-space representation should be defined by convolving the original signal with a one-parameter family of symmetric smoothing kernels possessing a semi-group property, and (ii) local extrema must not be enhanced when the scale parameter is increased continuously.

It is shown that given these requirements the scale-space representation must satisfy the differential equation $\partial_t L = \mathcal{A}_{ScSp} L$ for some linear and shift invariant operator $\mathcal{A}_{ScSp}$ satisfying locality, positivity, zero sum, and symmetry conditions. Examples in one, two, and three dimensions illustrate that this corresponds to natural semi-discretizations of the continuous (second-order) diffusion equation using different discrete approximations of the Laplacean operator. In a special case the multidimensional representation is given by convolution with the one-dimensional discrete analogue of the Gaussian kernel along each dimension.

Keywords: scale, scale-space, diffusion, Gaussian smoothing, multiscale representation, wavelets, image structure, causality.

1 Introduction

Image structures are intrinsically of a multiscale nature. Objects in the world and, hence, image features only exist as meaningful entities over certain ranges of scale. The idea behind a *scale-space* is to explicitly cope with this inherent property of measured data, by embedding a given signal into a family of gradually smoothed and simplified signals, in which the fine scale information is successively suppressed. Each member of the scale-space family should be associated with a specific value of a so-called *scale parameter*, somehow describing the current level of scale. A natural requirement of such an embedding is that features at coarser scales should correspond to (abstractions of) features at finer scales — they should not be just accidental phenomena created by the smoothing method.

* The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

This property has been formalized in different ways by different authors. When Witkin [21] introduced the term "scale-space" he observed a decreasing number of zero-crossings when subjecting a signal to Gaussian smoothing. Koenderink [9] showed that natural constraints; causality, homogeneity, and isotropy, necessarily imply that the scale-space of a two-dimensional signal must satisfy the diffusion equation. Other formulations were given by Yuille and Poggio [22], regarding the zero-crossings of the Laplacean, Babaud *et al.* [2], and Lindeberg [12] who combined a decreasing number of local extrema with a semi-group structure on the smoothing transformation. Recently, Florack *et al.* [6] showed that the uniqueness of the Gaussian kernel for scale-space representation can be derived under weaker assumptions by imposing scale invariance on a semi-group of convolution kernels.

From the similarity of these results it can by now be regarded as well established that within the class of linear transformations the natural way to construct a scale-space $L : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a *continuous* signal $f : \mathbb{R}^N \to \mathbb{R}$ is by convolution with the Gaussian kernel

$$L(\cdot;\, t) = g(\cdot;\, t) * f(\cdot) \ , \tag{1}$$

where

$$g(x;\, t) = \frac{1}{\sqrt{2\pi t}^N} e^{-(x_1^2 + \ldots + x_N^2)/2t} \ , \tag{2}$$

or equivalently, by solving the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L \tag{3}$$

with initial condition $L(\cdot;\, 0) = f$. In contrast to many other multiscale representations like pyramids (see, e.g., Burt [3]) or orthogonal wavelets (see, e.g., Mallat [17]), structures in the scale-space representation can be easily related across scales, since it is described by a differential equation (see, e.g., [14, 16]).

When applying scale-space theory in practice it should, however, be noted that real-life signals from standard detectors are *discrete*. The subject of this paper is to develop how scale-space theory can be discretized while still maintaining the scale-space properties exactly.

2 Scale-Space Theory for 1-D Discrete Signals

For one-dimensional signals it is possible to develop a complete discrete theory based on the assumption that the number of local extrema in a signal must not increase with scale. Below, are briefly summarized some of the main results from earlier work on this [12, 13]. The hasty reader may proceed directly to Sec. 3, where higher-dimensional signals are treated.

Definition 1 Discrete scale-space kernel (1-D). A kernel $K : \mathbb{Z} \to \mathbb{R}$ is said to be a scale-space kernel if for any signal $f_{in} : \mathbb{Z} \to \mathbb{R}$ the number of local extrema in $f_{out} = K * f_{in}$ does not exceed the number of local extrema in f_{in}.

Using classical results (mainly by Edrei and Schoenberg; see Karlin [8]) it is possible to completely classify those kernels that satisfy this definition.

Theorem 2 Classification of discrete scale-space kernels (1-D). *A kernel $K : \mathbb{Z} \to \mathbb{R}$ is a scale-space kernel if and only if its generating function $\varphi_K(z) = \sum_{n=-\infty}^{\infty} K(n)z^n$ is of the form*

$$\varphi_K(z) = c\, z^k\, e^{(q_{-1}z^{-1}+q_1 z)} \prod_{i=1}^{\infty} \frac{(1+\alpha_i z)(1+\delta_i z^{-1})}{(1-\beta_i z)(1-\gamma_i z^{-1})} ,$$

$$c > 0; \quad k; \in Z; \quad q_{-1}, q_1, \alpha_i, \beta_i, \gamma_i, \delta_i \geq 0 \quad \beta_i, \gamma_i < 1; \quad \sum_{i=1}^{\infty} (\alpha_i + \beta_i + \gamma_i + \delta_i) < \infty.$$

The interpretation of this result is that discrete scale-space kernels obey the following decomposition property:

Corollary 3 Primitive discrete smoothing transformations (1-D). *For discrete signals $\mathbb{Z} \to \mathbb{R}$ there are five primitive types of linear and shift-invariant smoothing transformations, of which the last two are trivial;*

- *two-point weighted average or generalized binomial smoothing*

$$f_{out}(x) = f_{in}(x) + \alpha_i f_{in}(x-1) \qquad (\alpha \geq 0),$$
$$f_{out}(x) = f_{in}(x) + \delta_i f_{in}(x+1) \qquad (\delta_i \geq 0),$$

- *moving average or first order recursive filtering*

$$f_{out}(x) = f_{in}(x) + \beta_i f_{out}(x-1) \qquad (0 \leq \beta_i < 1),$$
$$f_{out}(x) = f_{in}(x) + \gamma_i f_{out}(x+1) \qquad (0 \leq \gamma_i < 1),$$

- *infinitesimal smoothing or diffusion smoothing (see Theorem 4 for an example),*
- *rescaling, and*
- *translation.*

It follows that a discrete kernel is a scale-space kernel if and only if it can be decomposed into the above primitive transformations. Moreover, the only non-trivial smoothing kernels of *finite support* arise from binomial smoothing.

If Definition 1 is combined with a requirement that the family of smoothing transformations must possess a semi-group property and have a *continuous* scale parameter, then the result is that there is in principle only one way to construct a scale-space for discrete signals.

Theorem 4 Scale-space for discrete signals; Necessity and sufficiency. *Given any signal $f : \mathbb{Z} \to \mathbb{R}$, let $L : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ be a one-parameter family of functions defined by $L(x;\, 0) = f(x)$ $(x \in \mathbb{Z})$ and*

$$L(x;\, t) = \sum_{n=-\infty}^{\infty} T(n;\, t) f(x-n) \ , \tag{4}$$

($x \in \mathbb{Z}$, $t > 0$), where $T : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ is a one-parameter family of symmetric functions satisfying the semi-group property $T(\cdot;\ s) * T(\cdot;\ t) = T(\cdot;\ s+t)$ and the normalization criterion $\sum_{n=-\infty}^{\infty} T(n;\ t) = 1$. For any signal f and any $t_2 > t_1$ it is required that the number of local extrema (zero-crossings) in $L(x;\ t_2)$ must not exceed the number of local extrema (zero-crossings) in $L(x;\ t_1)$. Then, necessarily (and sufficiently),

$$T(n;\ t) = e^{-\alpha t} I_n(\alpha t) \tag{5}$$

for some non-negative real α, where I_n are the modified Bessel functions of integer order. This kernel T is called the discrete analogue of the Gaussian kernel.

Similar arguments in the continuous case uniquely lead to the Gaussian kernel.

The term "diffusion smoothing" can be understood by noting that the scale-space family L satisfies a semi-discretized version of the diffusion equation:

Theorem 5 Diffusion formulation of the discrete scale-space. *The representation $L : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ given by (4) with $T : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ according to (5) and $\alpha = 1$ satisfies the system of ordinary differential equations*

$$\partial_t L(x;\ t) = \frac{1}{2}(L(x+1;\ t) - 2L(x;\ t) + L(x-1;\ t)) = \frac{1}{2}(\nabla_2^2 L)(x;\ t) \tag{6}$$

with initial condition $L(x;\ 0) = f(x)$ for any discrete signal $f : \mathbb{Z} \to \mathbb{R}$ in l_1.

Despite the completeness of these results, they cannot be extended directly to higher dimensions, since in two (and higher) dimensions there are no non-trivial kernels guaranteed to never increase the number of local extrema in a signal. One example of this, originally due to Lifshitz and Pizer [11], can be found in [12] (see also Yuille [23]). Anyway, an important point about this study, is that it gives a deep understanding of what one-dimensional linear transformations can be regarded as smoothing transformations. It also shows that the only reasonable way to convert the one-dimensional scale-space theory from continuous signals to discrete signals is by discretizing the diffusion equation.

3 Selecting Scale-Space Axioms in Higher Dimensions

Koenderink [9] derives the scale-space for two-dimensional continuous images from three assumptions; causality, homogeneity, and isotropy. The main idea is that it should be possible to trace every grey-level at a coarse scale to a corresponding grey-level at a finer scale. In other words, no new level curves should be created when the scale parameter increases. Using differential geometry he shows that these requirements uniquely lead to the diffusion equation.

It is of course impossible to apply these ideas directly in the discrete case, since there are no direct correspondences to level curves or differential geometry for discrete signals. Neither can the scaling argument by Florack *et al.* [6] be carried out in a discrete situation. An alternative way of expressing the first property, however, is by requiring that *if for some scale level t_0 a point x_0 is a local maximum for the scale-space representation at that level (regarded as a*

function of the space coordinates only) then its value must not increase when the scale parameter increases. Analogously, if a point is a local minimum then its value must not decrease when the scale parameter increases.

It is clear that this formulation is equivalent to the formulation in terms of level curves for continuous images, since if the grey-level value at a local maximum (minimum) would increase (decrease) then a new level curve would be created. Conversely, if a new level curve is created then some local maximum (minimum) must have increased (decreased). An intuitive description of this requirement is that it prevents local extrema from being enhanced and from "popping up out of nowhere". In fact, this is closely related to the maximum principle for parabolic differential equations (see, e.g., Widder [20]).

In next section it will be shown that this condition combined with a continuous scale parameter means a strong restriction on the smoothing method also in the discrete case, and again it will lead to a discretized version of the diffusion equation. In a special case, the scale-space representation will be reduced to the family of functions generated by separated convolution with the discrete analogue of the Gaussian kernel, $T(n;\, t)$.

3.1 Basic Definitions

Given a point $x \in \mathbb{Z}^N$ denote its neighbourhood of connected points by $N(x) = \{\xi \in Z^N : (\| x - \xi \|_\infty \leq 1) \wedge (\xi \neq x)\}$ (corresponding to what is known as eight-connectivity in the two-dimensional case). The corresponding set including the central point x is written $N_+(x)$. Define (weak) extremum points as follows:

Definition 6 Discrete local maximum. A point $x \in \mathbb{Z}^N$ is said to be a (weak) local maximum of a function $g : \mathbb{Z}^N \to \mathbb{R}$ if $g(x) \geq g(\xi)$ for all $\xi \in N(x)$.

Definition 7 Discrete local minimum. A point $x \in \mathbb{Z}^N$ is said to be a (weak) local minimum of a function $g : \mathbb{Z}^N \to \mathbb{R}$ if $g(x) \leq g(\xi)$ for all $\xi \in N(x)$.

The following operators are natural discrete correspondences to the Laplacean operator ∇^2 in one (∇_3^2), two ($\nabla_5^2, \nabla_{\times 2}^2$) and three ($\nabla_7^2, \nabla_{+3}^2, \nabla_{\times 3}^2$) dimensions respectively (below the notation $f_{-1,0,1}$ stands for $f(x-1, y, z+1)$ etc.):

$$(\nabla_3^2 f)_0 = f_{-1} - 2f_0 + f_1,$$

$$(\nabla_5^2 f)_0 = f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4f_{0,0},$$

$$(\nabla_{\times 2}^2 f)_{0,0} = 1/2(f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4f_{0,0}),$$

$$(\nabla_7^2 f)_{0,0,0} = f_{-1,0,0} + f_{+1,0,0} + f_{0,-1,0} + f_{0,+1,0} + f_{0,0,-1} + f_{0,0,+1} - 6f_{0,0,0},$$

$$\begin{aligned}(\nabla_{+3}^2 f)_{0,0,0} = 1/4 \quad & (f_{-1,-1,0} + f_{-1,+1,0} + f_{+1,-1,0} + f_{+1,+1,0} + \\ & f_{-1,0,-1} + f_{-1,0,+1} + f_{+1,0,-1} + f_{+1,0,+1} + \\ & f_{0,-1,-1} + f_{0,-1,+1} + f_{0,+1,-1} + f_{0,+1,+1} - 12f_{0,0,0}),\end{aligned}$$

$$\begin{aligned}(\nabla_{\times 3}^2 f)_{0,0,0} = 1/4 \quad & (f_{-1,-1,-1} + f_{-1,-1,+1} + f_{-1,+1,-1} + f_{-1,+1,+1} + \\ & f_{+1,-1,-1} + f_{+1,-1,+1} + f_{+1,+1,-1} + f_{+1,+1,+1} - 8f_{0,0,0}).\end{aligned}$$

4 Axiomatic Discrete Scale-Space Formulation

Given that the task is to state an axiomatic formulation of the first stages of visual processing, *the visual front end*, a list of desired properties may be long; *linearity*, *translational invariance*, *rotational symmetry*, *mirror symmetry*, *semi-group*, *causality*, *positivity*, *unimodality*, *continuity*, *differentiability*, *normalization* to one, nice *scaling* behaviour, *locality*, rapidly *decreasing* for large x and t, existence of an *infinitesimal generator* (explained below), and *invariance* with respect to certain *grey-level transformations*, etc. Such a list will, however, contain redundancies, as does this one. Here, a (minimal) subset of these properties will be taken as axioms. In fact, it can be shown that all the other above-mentioned properties follow from the selected subset (see also [15, 16]).

The scale-space representation for higher-dimensional signals is constructed analogously to the one-dimensional case. To start with, postulate that the scale-space should be generated by convolution with a one-parameter family of kernels, i.e., $L(x;\ 0) = f(x)$ and

$$L(x;\ t) = \sum_{\xi \in \mathbb{Z}^N} T(\xi;\ t) f(x - \xi) \quad (t > 0) \ . \tag{7}$$

This form of the smoothing formula corresponds to natural requirements about *linear shift-invariant smoothing* and the existence of a *continuous scale parameter*. It is natural to require that all coordinate directions should be handled identically. Therefore all kernels should be *symmetric*. Impose also a *semi-group* condition on the family T. This means that all scale levels will be treated similarly, that is, the smoothing operation does not depend on the scale value, and the transformation from a lower scale level to a higher scale level is always given by convolution with a kernel from the family:

$$\begin{aligned} L(\cdot;\ t_2) = \{\text{definition}\} = T(\cdot;\ t_2) * f = \{\text{semi-group}\} = \\ = (T(\cdot;\ t_2 - t_1) * T(\cdot;\ t_1)) * f = \{\text{associativity}\} = \\ = T(\cdot;\ t_2 - t_1) * (T(\cdot;\ t_1) * f) = \{\text{definition}\} = T(\cdot;\ t_2 - t_1) * L(\cdot;\ t_1) \ . \end{aligned} \tag{8}$$

As smoothing criterion the *non-enhancement* requirement for local extrema is taken. It is convenient to express it as a condition of the derivative of the scale-space family with respect to the scale parameter. In order to ensure a proper statement of this condition, where differentiability is guaranteed, it is necessary to state a series of preliminary definitions leading to the desired formulation.

4.1 Definitions

Let us summarize this (minimal) set of basic properties, which a family should satisfy in order to be a candidate family for generating a (linear) scale-space.

Definition 8 Pre-scale-space family of kernels. A one-parameter family of kernels $T : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to be a pre-scale-space family of kernels if it satisfies

- $T(\cdot;\ 0) = \delta(\cdot)$,
- the semi-group property $T(\cdot;\ s) * T(\cdot;\ t) = T(\cdot;\ s+t)$,
- the symmetry properties $T(-x_1, x_2, ..., x_N;\ t) = T(x_1, x_2, ..., x_N;\ t)$ and $T(P_k^N(x_1, x_2, ..., x_N);\ t) = T(x_1, x_2, ..., x_N;\ t)$ for all $x = (x_1, x_2, ..., x_N) \in \mathbb{Z}^n$, all $t \in \mathbb{R}_+$, and all possible permutations P_k^N of N elements, and
- the continuity requirement $\| T(\cdot;\ t) - \delta(\cdot) \|_1 \to 0$ when $t \downarrow 0$.

Definition 9 Pre-scale-space representation. Let $f : \mathbb{Z}^N \to \mathbb{R}$ be a discrete signal and let $T : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ be a pre-scale-space family of kernels. Then, the one-parameter family of signals $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ given by (7) is said to be the pre-scale-space representation of f generated by T.

Provided that the input signal f is sufficiently regular, these conditions on the family of kernels T guarantee that the representation L is differentiable and satisfies a system of linear differential equations.

Lemma 10 A pre-scale-space representation is differentiable. *Let $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ be the pre-scale-space representation of a signal $f : \mathbb{Z}^N \to \mathbb{R}$ in l_1. Then L satisfies the differential equation*

$$\partial_t L = \mathcal{A} L \tag{9}$$

for some linear and shift-invariant operator $\mathcal{A}$.

Proof. If f is sufficiently regular, e.g., if $f \in l_1$, define a family of operators $\{\mathcal{T}_t, t > 0\}$, here from from l_1 to l_1, by $\mathcal{T}_t f = T(\cdot;\ t) * f$. Due to the conditions imposed on the kernels it will satisfy the relation

$$\lim_{t \to t_0} \| (\mathcal{T}_t - \mathcal{T}_{t_0}) f \|_1 = \lim_{t \to t_0} \| (\mathcal{T}_{t-t_0} - \mathcal{I})(\mathcal{T}_{t_0} f) \|_1 = 0 \ , \tag{10}$$

where $\mathcal{I}$ is the identity operator. Such a family is called a strongly-continuous semigroup of operators (see Hille and Phillips [7] p. 58-59). A semi-group is often characterized by its *infinitesimal generator* $\mathcal{A}$ defined by

$$\mathcal{A} f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h} \ . \tag{11}$$

The set of elements f for which $\mathcal{A}$ exists is denoted $\mathcal{D}(\mathcal{A})$. This set is not empty and never reduces to the zero element. Actually, it is even dense in l_1 (see Hille and Phillips [7] p. 307). If this operator exists then

$$\lim_{h \downarrow 0} \frac{L(\cdot, \cdot;\ t+h) - L(\cdot, \cdot;\ t)}{h} = \lim_{h \downarrow 0} \frac{\mathcal{T}_{t+h} f - \mathcal{T}_t f}{h} = \tag{12}$$

$$\lim_{h \downarrow 0} \frac{\mathcal{T}_h(\mathcal{T}_t f) - (\mathcal{T}_t f)}{h} = \mathcal{A}(\mathcal{T}_t f) = \mathcal{A} L(\cdot;\ t).$$

According to a theorem by Hille and Phillips ([7] p. 308) strong continuity implies $\partial_t(\mathcal{T}_t f) = \mathcal{A}\mathcal{T}_t f = \mathcal{T}_t \mathcal{A} f$ for all $f \in \mathcal{D}(\mathcal{A})$. Hence, the scale-space family L must obey the differential equation $\partial_t L = \mathcal{A} L$ for some linear operator $\mathcal{A}$. Since L is generated from f by a convolution operation it follows that $\mathcal{A}$ must be shift-invariant. □

This property makes it possible to formulate the previously indicated scale-space property in terms of derivatives of the scale-space representation with respect to the scale parameter. As in the maximum principle, the grey-level value in every local maximum point must not increase, while the grey-level value in every local minimum point must not decrease.

Definition 11 Pre-scale-space property: Non-enhancement of extrema. A differentiable one-parameter family of signals $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to possess pre-scale-space properties, or equivalently not to enhance local extrema, if for every value of the scale parameter $t_0 \in \mathbb{R}_+$ it holds that if $x_0 \in \mathbb{Z}^N$ is a local extremum point for the mapping $x \mapsto L(x;\ t_0)$ then the derivative of L with respect to t in this point satisfies

$$\partial_t L(x_0;\ t_0) \leq 0 \quad \text{if } x_0 \text{ is a local maximum point,} \tag{13}$$

$$\partial_t L(x_0;\ t_0) \geq 0 \quad \text{if } x_0 \text{ is a local minimum point.} \tag{14}$$

Now it can be stated that a pre-scale-space family of kernels is a scale-space family of kernels if it satisfies this property for *any* input signal.

Definition 12 Scale-space family of kernels. A one-parameter family of pre-scale-space kernels $T : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to be a scale-space family of kernels if for any signal $f : \mathbb{Z}^N \to \mathbb{R} \in l_1$ the pre-scale-space representation of f generated by T possesses pre-scale-space properties, i.e., if *for any* signal local extrema are never enhanced.

Definition 13 Scale-space representation. A pre-scale-space representation $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a signal $f : \mathbb{Z}^N \to \mathbb{R}$ generated by a family of kernels $T : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$, which are scale-space kernels, is said to be a scale-space representation of f.

In the next section it will be shown how these requirements strongly restrict the possible class of kernels and scale-space representations. For example, they will lead to a number of restrictions on the operator $\mathcal{A}$ in Lemma 10:

Definition 14 Infinitesimal scale-space generator. A shift-invariant linear operator $\mathcal{A}$ from l_1 to l_1

$$(\mathcal{A}L)(x;\ t) = \sum_{\xi \in \mathbb{Z}^N} a_\xi L(x - \xi;\ t) \ , \tag{15}$$

is said to be an infinitesimal scale-space generator, denoted $\mathcal{A}_{ScSp}$, if the coefficients $a_\xi \in R$ satisfy

- the locality condition $a_\xi = 0$ if $\xi \notin N_+(0)$,
- the positivity constraint $a_\xi \geq 0$ if $\xi \neq 0$,
- the zero sum condition $\sum_{\xi \in \mathbb{Z}^N} a_\xi = 0$, as well as
- the symmetry requirements $a_{(-\xi_1,\xi_2,\ldots,\xi_N)} = a_{(\xi_1,\xi_2,\ldots,\xi_N)}$ and $a_{P_k^N(\xi_1,\xi_2,\ldots,\xi_N)} = a_{(\xi_1,\xi_2,\ldots,\xi_N)}$ for all $\xi = (\xi_1, \xi_2, \ldots, \xi_N) \in \mathbb{Z}^N$ and all possible permutations P_k^N of N elements.

4.2 Necessity

It will first be shown that these conditions necessarily imply that the family L satisfies a semi-discretized version of the diffusion equation.

Theorem 15 Scale-space for discrete signals: Necessity. *A scale-space representation $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a signal $f : \mathbb{Z}^N \to \mathbb{R}$ satisfies the differential equation*

$$\partial_t L = \mathcal{A}_{ScSp} L \tag{16}$$

with initial condition $L(\cdot;\, 0) = f(\cdot)$ for some infinitesimal scale-space generator $\mathcal{A}_{ScSp}$. In one, two and three dimensions respectively (16) reduces to

$$\partial_t L = \alpha_1 \nabla_3^2 L \ , \tag{17}$$

$$\partial_t L = \alpha_1 \nabla_5^2 L + \alpha_2 \nabla_{\times 3}^2 L \ , \tag{18}$$

$$\partial_t L = \alpha_1 \nabla_7^2 L + \alpha_2 \nabla_{+3}^2 L + \alpha_3 \nabla_{\times 3}^2 L \ , \tag{19}$$

for some constants $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_3 \geq 0$.

Proof. The proof consists of two parts. The first part has already been presented in Lemma 10, where it was shown that the requirements on the kernels imply that the family L obeys a linear differential equation. Because of the shift invariance $\mathcal{A}L$ can be written in the form (15). In the second part counterexamples will be constructed from various simple test functions in order to delimit the class of possible operators.

The extremum point conditions (13), (14) combined with Definitions 12-13 mean that $\mathcal{A}$ must be *local*, i.e., that $a_\xi = 0$ if $\xi \notin N_+(0)$. This is easily understood by studying the following counterexample: First, assume that $a_{\xi_0} > 0$ for some $\xi_0 \notin N_+(0)$ and define a function $f_1 : \mathbb{Z}^N \to \mathbb{R}$ by

$$f_1(x) = \begin{cases} \varepsilon > 0 \text{ if } x = 0, \\ \quad 0 \text{ if } x \in N(0), \\ \quad 1 \text{ if } x = \xi_0, \text{ and} \\ \quad 0 \text{ otherwise.} \end{cases} \tag{20}$$

Obviously, 0 is a local maximum point for f_1. From (9) and (15) one obtains $\partial_t L(0;\, 0) = \epsilon a_0 + a_{\xi_0}$. It is clear that this value can be positive provided that ε is chosen small enough. Hence, L cannot satisfy (13). Similarly, it can also be shown that $a_{\xi_0} < 0$ leads to a violation of the non-enhancement property (14) (let $\varepsilon < 0$). Consequently, a_ξ must be zero if $\xi \notin N_+(0)$.

Moreover, the *symmetry* conditions imply that permuted and reflected coefficients must be equal, i.e., $a_{(-\xi_1,\xi_2,\ldots,\xi_N)} = a_{(\xi_1,\xi_2,\ldots,\xi_N)}$ and $a_{P_k^N(\xi_1,\xi_2,\ldots,\xi_N)} = a_{(\xi_1,\xi_2,\ldots,\xi_N)}$ for all $\xi = (\xi_1, \xi_2, \ldots, \xi_N) \in \mathbb{Z}^N$ and all possible permutations P_k^N of N elements. For example, the two-dimensional version of (15) reads

$$\partial_t L = \begin{pmatrix} a & b & a \\ b & c & b \\ a & b & a \end{pmatrix} L \tag{21}$$

for some a, b and c. Then, consider the function given by

$$f_2(x, y) = \begin{cases} 1 \text{ if } x \in N_+(0), \text{ and} \\ 0 \text{ otherwise.} \end{cases} \tag{22}$$

With the given (weak) definitions of local extremum points it is clear that 0 is both a local maximum point and a local minimum point. Hence $\partial_t L(0;\ 0)$ must be zero, and the *coefficients sum to zero* $\sum_{\xi \in \mathbb{Z}^N} a_\xi = 0$, which in two dimensions reduces to $4a + 4b + c = 0$ in (21). Obviously, (15) can be written

$$\partial_t L = (\mathcal{A}L)(x;\ t) = \sum_{\xi \in N(0)} a_\xi (L(x - \xi;\ t) - L(x;\ t)) \ , \tag{23}$$

and the two-dimensional special case (21) reduces to

$$\partial_t L = \alpha_1 \begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix} L + \alpha_2 \begin{pmatrix} 1/2 & & 1/2 \\ & -2 & \\ 1/2 & & 1/2 \end{pmatrix} L = \alpha_1 \nabla_5^2 L + \alpha_2 \nabla_{x^2}^2 L \ . \tag{24}$$

Finally, by considering the test function

$$f_3(x, y) = \begin{cases} \epsilon > 0 \text{ if } x = 0, \\ -1 \text{ if } x = \tilde{\xi}, \text{ and} \\ 0 \text{ otherwise.} \end{cases} \tag{25}$$

for some $\tilde{\xi}$ in $N(0)$ one easily realizes that a_ξ must be *non-negative* if $\xi \in N(0)$. It follows that $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ in (24), which proves (18). (17) and (19) follow from similar straightforward considerations. The initial condition $L(\cdot;\ 0) = f$ is a direct consequence of the definition of pre-scale-space kernel. □

4.3 Sufficiency

The reverse statement of Theorem 15 is also true.

Theorem 16 Scale-space for discrete signals: Sufficiency. *Let $f : \mathbb{Z}^N \to \mathbb{R}$ be a discrete signal in l_1, let $\mathcal{A}_{ScSp}$ be an infinitesimal scale-space generator, and let $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ be the representation generated by the solution to the differential equation*

$$\partial_t L = \mathcal{A}_{ScSp} L$$

with initial condition $L(\cdot;\ 0) = f(\cdot)$. Then, L is a scale-space representation of f.

Proof. It follows almost trivially that L possesses pre-scale-space properties, i.e., that L does not enhance local extrema, if the differential equation is rewritten in the form

$$\partial_t L = (\mathcal{A}L)(x;\ t) = \sum_{\xi \in N(0)} a_\xi (L(x - \xi;\ t) - L(x;\ t)) \ . \tag{26}$$

If at some scale level t a point x is a local maximum point then all differences $L(x-\xi;\ t) - L(x;\ t)$ are non-positive, which means that $\partial_t L(x;\ t) \leq 0$ provided that $a_\xi \geq 0$. Similarly, if a point is a local minimum point then the differences are all non-negative and $\partial_t L(x;\ t) \geq 0$.

What remains to be verified is that L actually satisfies the requirements for being a pre-scale-space representation. Since L is generated by a linear differential equation, it follows that L can be written as the convolution of f with some kernel T, i.e., $L(\cdot;\ t) = T(\cdot;\ t) * f$. The requirements of pre-scale-space kernels can be shown to hold by letting the input signal f be the discrete delta function. The semi-group property of the kernels follows from the fact that the coefficients ξ are constant, and the solution at a time $s+t$ hence can be computed from the solution at an earlier time s by letting the time increase by t. The symmetry properties of the kernel are obvious from the symmetry of the differential equation. The continuity at the origin follows directly from the differentiability. □

These results show that *a one-parameter family of discrete signals is a scale-space representation if and only if it satisfies the differential equation (16) for some infinitesimal scale-space generator.*

5 Parameter Determination

For simplicity, from now on mainly two-dimensional signals will be considered. If (18) is rewritten in the form

$$\partial_t L = C\left((1-\gamma)\nabla_5^2 L + \gamma \nabla_{\times 2}^2 L\right) = C\nabla_\gamma^2 L\ , \tag{27}$$

the interpretation of the parameter C is just a trivial rescaling of the scale parameter. Thus, without loss of generality C may be set to $\frac{1}{2}$ in order to get the same scaling constant as in the one-dimensional case. What is left to investigate is how the remaining degree of freedom in the parameter $\gamma \in [0, 1]$ affects the scale-space representation.

If $\gamma = 1$ then a undesirable situation appears. Since the cross-operator only links diagonal points, the system of ordinary differential equations given by (27) can then be split into two *uncoupled* systems, one operating on the points with even coordinate sum $x + y$ and the other operating on the points with odd coordinate sum. It is clear that this is really an unwanted behaviour, since then even after a substantial amount of "blurring", for certain types of input signals the "smoothed" grey-level landscape may still have a rather saw-toothed shape.

5.1 Derivation of the Fourier Transform

Further arguments showing that γ must not be too large can be obtained by studying the Fourier transform of the corresponding scale-space family of kernels.

Proposition 17 Fourier transform of the 2D discrete scale-space. *Let* $L : \mathbb{Z}^2 \times \mathbb{R}_+ \to \mathbb{R}$ *be the scale-space representation of a discrete signal* $f : \mathbb{Z}^2 \to \mathbb{R}$

generated by (27) with initial condition $L(\cdot;\ 0) = f(\cdot)$. ***Assume that*** $f \in l_1$. ***Then the generating function of the kernel describing the transformation from the original signal to the representation at a certain scale*** t ***is given by***

$$\varphi_T(z, w) = \sum_{(m,n) \in \mathbb{Z}^2} T(m, n;\ t) z^m w^n =$$

$$e^{-(2-\gamma)+\frac{(1-\gamma)}{2}(z^{-1}+z+w^{-1}+w)+\frac{\gamma}{4}(z^{-1}w^{-1}+z^{-1}w+zw^{-1}+zw)} \quad . \tag{28}$$

Its Fourier transform is

$$\varphi_T(z, w) = \psi_T(e^{-iu}, e^{-iv}) = e^{-(2-\gamma)t\ +\ (1-\gamma)(\cos u+\cos v)t\ +\ (\gamma \cos u \cos u)t} \quad . \tag{29}$$

Proof. Discretizing (27) further in scale using Euler's explicit method with scale step Δt, gives an iteration formula of the form

$$\begin{aligned} L_{i,j}^{k+1} = (1 - (2-\gamma)\Delta t)\ L_{i,j}^{k} + \\ \frac{(1-\gamma)\Delta t}{2}\ (L_{i-1,j}^{k} + L_{i+1,j}^{k} + L_{i,j-1}^{k} + L_{i,j+1}^{k}) + \\ \frac{\gamma \Delta t}{4}\ (L_{i-1,j-1}^{k} + L_{i-1,j+1}^{k} + L_{i+1,j-1}^{k} + L_{i+1,j+1}^{k}), \end{aligned} \tag{30}$$

where the subscripts i and j denote the spatial coordinates x and y respectively, and the superscript k denotes the iteration index. The generating function describing one iteration with this transformation is

$$\begin{aligned} \varphi_{step}(z, w) = (1 - (2-\gamma)\Delta t) + \frac{(1-\gamma)\Delta t}{2}(z^{-1} + z + w^{-1} + w) + \\ \frac{\gamma \Delta t}{4}(z^{-1}w^{-1} + z^{-1}w + zw^{-1} + zw) \quad . \end{aligned} \tag{31}$$

Assume that the scale-space representation at a scale level t is computed using n iterations with a scale step $\Delta t = \frac{t}{n}$. Then, the generating function describing the composed transformation can be written $\varphi_{composed,n}(z, w) = (\varphi_{step}(z, w))^n$. After substitution of Δt for $\frac{t}{n}$ and using $\lim_{n \to \infty}(1 + \frac{\alpha_n}{n})^n = e^{\alpha}$ if $\lim_{n \to \infty} \alpha_n = \alpha$, it follows that $\varphi_{composed,n}(z)$ tends to $\varphi_T(z, w)$ according to (28) when $n \to \infty$, provided that the discretization (30) converges to the actual solution of (27). □

5.2 Unimodality in the Fourier Domain

It is easy to verify that the Fourier transform is unimodal if and only if $\gamma \leq \frac{1}{2}$.

Proposition 18 Unimodality of the Fourier transform (2D).
The Fourier transform (29) of the kernel describing the transformation from the original signal to the smoothed representation at a coarser level of scale is unimodal if and only if $\gamma \leq \frac{1}{2}$.

Proof. Differentiation of (29) gives $\partial_u \psi = -\psi(u, v) \sin u\,(1 - \gamma(1 + \cos v))\,t$ and $\partial_v \psi = -\psi(u, v) \sin v\,(1 - \gamma(1 + \cos u))\,t$. The Fourier transform decreases with $|u|$ and $|v|$ for all u and v in $[-\pi, \pi]$ if and only if the factors $(1 - \gamma(1 + \cos v))$ and $(1 - \gamma(1 + \cos u))$ are non-negative for all u and v, i.e., if and only if $\gamma \leq \frac{1}{2}$. Then, any directional derivative away from the origin is negative. □

5.3 Separability

The transformation kernel is separable if and only if its Fourier transform is separable, that is, if and only if $\psi_T(u, v)$ can be written on the form $U_T(u)V_T(v)$ for some functions U_T and V_T. From (29) it is realized that this separation is possible if and only if $\gamma = 0$. Hence,

Proposition 19 Separability of the 2D discrete scale-space. *The convolution kernel associated with the scale-space representation defined by* $L(x, y;\ 0) = f(x, y)$ *and*

$$\partial_t L = \frac{1}{2}\left((1-\gamma)\nabla_5^2 L + \gamma \nabla_{\times 2}^2\right) \tag{32}$$

is separable if and only if $\gamma = 0$. *Then* L *is given by*

$$L(x, y;\ t) = \sum_{m=-\infty}^{\infty} T(m;\ t) \sum_{n=-\infty}^{\infty} T(n;\ t) f(x-m, y-n) \quad (t > 0), \tag{33}$$

where $T(n;\ t) = e^{-t} I_n(t)$ *and* I_n *are the modified Bessel functions of integer order.*

Proof. The Fourier transform $\psi_T(u, v)$ can be written in the form $U_T(u)V_T(v)$ for some functions U_T and V_T if and only if the term with $\cos u \cos v$ can be eliminated from the argument of the exponential function, i.e., if and only if γ is zero. In that case the Fourier transform reduces to

$$\psi_T(u, v) = e^{(-2+\cos u+\cos v)t} = e^{(-1+\cos u)t} e^{(-1+\cos v)t} \tag{34}$$

which corresponds to separated smoothing with the one-dimensional discrete analogue of the Gaussian kernel along each coordinate direction.

It can also be verified directly that (33) satisfies (32). Consider the possible scale-space representation of an N-dimensional signal generated by separable convolution with the one-dimensional discrete analogue of the Gaussian kernel; i.e., given $f : \mathbb{Z}^N \to \mathbb{R}$ define $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ by

$$L(x;\ t) = \sum_{\xi \in \mathbb{Z}^N} T_N(\xi;\ t) f(x-\xi) \quad (t > 0), \tag{35}$$

where $T_N : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$T_N(\xi;\ t) = \prod_{i=1}^{N} T_1(\xi_i;\ t)\ , \tag{36}$$

$\xi = (\xi_1, ..., \xi_N)$, and $T_1 : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ is the discrete analogue of the Gaussian kernel, $T_1(n;\ t) = e^{-t} I_n(t)$. It will be shown that this representation satisfies a semi-discretized version of the two-dimensional diffusion equation

$$\partial_t L = \frac{1}{2}\nabla_{2N+1}^2 L\ , \tag{37}$$

where

$$(\nabla^2_{2N+1} L)(x;\ t) = \sum_{i=1}^{N} L(x + e_i;\ t) - 2L(x;\ t) + L(x - e_i;\ t)\ , \qquad (38)$$

and e_i denotes the unit vector in the ith coordinate direction. Consider

$$(\partial_t T_N)(\xi;\ t) = \sum_{i=1}^{N} (\partial_t T_1)(\xi_i;\ t) \prod_{j \neq i} T_1(\xi_j;\ t)\ . \qquad (39)$$

Since T_1 satisfies (6), this expression can be written

$$(\partial_t T_N)(\xi;\ t) = \sum_{i=1}^{N} \frac{1}{2}(T_1(\xi_i - 1;\ t) - 2T_1(\xi_i;\ t) + T_1(\xi_i + 1;\ t)) \prod_{j \neq i} T_1(\xi_j;\ t),$$

which is obviously equivalent to

$$\partial_t T_N = \frac{1}{2} \nabla^2_{2N+1} T_N\ . \qquad (40)$$

The same relation holds for L provided that the differentiation and infinite summation operators commute. □

In other words, in the separable case the resulting higher-dimensional discrete scale-space corresponds to repeated application of the one-dimensional scale-space concept along each coordinate direction.

5.4 Discrete Iterations

The discretization of (27) in (30) using Euler's explicit method with scale step Δt corresponds to iterating with a kernel with the computational molecule

$$\begin{pmatrix} \frac{\gamma \Delta t}{4} & \frac{(1-\gamma)\Delta t}{2} & \frac{\gamma \Delta t}{4} \\ \frac{(1-\gamma)\Delta t}{2} & 1 - (2-\gamma)\Delta t & \frac{(1-\gamma)\Delta t}{2} \\ \frac{\gamma \Delta t}{4} & \frac{(1-\gamma)\Delta t}{2} & \frac{\gamma \Delta t}{4} \end{pmatrix}. \qquad (41)$$

Clearly, this kernel is unimodal if and only if $\gamma \leq \frac{2}{3}$. It is separable if and only if $\gamma = \Delta t$ (see below). In that case, the corresponding one-dimensional kernel is a discrete scale-space kernel in the sense of Definition 1 if and only if $\Delta t \leq \frac{1}{2}$ (see Proposition 10 in [12]). This gives a further indication that γ should not exceed the value $\frac{1}{2}$.

Proposition 20 Separability of the iteration kernel. *The iteration kernel (41), corresponding to discrete forward iteration with Euler's explicit method, is separable if and only if $\gamma = \Delta t$. In that case, the corresponding one-dimensional kernel is a discrete scale-space kernel if and only if $0 \leq \gamma \leq 1/2$.*

Proof. Since the kernel is symmetric and the coefficients sum to one, the kernel is separable if and only if it can be written as a kernel $(a,\ 1-2a,\ a)$ convolved with itself, i.e., if and only if there exists an $a \geq 0$ such that $a^2 = \gamma\Delta t/4$, $a(1-a) = (1-\gamma)\Delta t/2$, and $(1-a)^2 = 1-(2-\gamma)\Delta t$. The first equation has one non-negative root $a = \sqrt{\gamma\Delta t}/2$. Insertion into the second equation gives two conditions for Δt; either $\Delta t = 0$ or $\Delta t = \gamma$. One verifies that these roots satisfy the third equation. The kernel $(a,\ 1-2a,\ a)$ is a discrete scale-space kernel if and only if $a \leq \frac{1}{2}$ (see Equations (30) and (31) in [12]; compare also with Theorem 2). □

The boundary case $\gamma = \Delta t = \frac{1}{2}$ gives the iteration kernel in Fig. 1(a) corresponding to separated convolution with the one-dimensional binomial kernel $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ frequently used in pyramid generation (see, e.g., Crowley [4]).

$$\begin{pmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{pmatrix} \quad \begin{pmatrix} 1/8 & 2/8 & 1/8 \\ 2/8 & -12/8 & 2/8 \\ 1/8 & 2/8 & 1/8 \end{pmatrix} \quad \begin{pmatrix} 1/36 & 1/9 & 1/36 \\ 1/9 & 4/9 & 1/9 \\ 1/36 & 1/9 & 1/36 \end{pmatrix} \quad \begin{pmatrix} 1/6 & 4/6 & 1/6 \\ 4/6 & -20/3 & 4/6 \\ 1/6 & 4/6 & 1/6 \end{pmatrix}$$

(a) (b) (c) (d)

Fig. 1. Computational molecules corresponding to (from left to right); (a) discrete iteration with $\Delta t = \gamma = \frac{1}{2}$, (b) the Laplacean operator when $\gamma = \frac{1}{2}$, (c) discrete iteration with $\Delta t = \gamma = \frac{1}{3}$, and (d) the Laplacean operator when $\gamma = \frac{1}{3}$.

5.5 Spatial Isotropy

Another aspect that might affect the selection of γ is spatial isotropy. It is not clear that rotational invariance is a primary quality to be aimed at in the discrete case, since then one is locked to a fixed square grid. It is also far from obvious what should be meant by spatial isotropy in a discrete situation. Possibly, it is better to talk about the lack of spatial isotropy, spatial anisotropy, or rotational asymmetry. However, since the Fourier transform is a continuous function of u and v, one can regard its variation as a function of the polar angle, given a fixed value of the radius, as one measure of this property. By expressing $\psi_T(u,v)$ in polar coordinates $u = \omega\cos\phi$, $v = \omega\sin\phi$ and examining the resulting expression,

$$\psi_T(\omega\cos\phi, \omega\sin\phi) = e^{h(\omega\cos\phi, \omega\sin\phi)t}\ , \tag{42}$$

where

$$h(\omega\cos\phi, \omega\sin\phi) = -(2-\gamma) + (1-\gamma)(\cos(\omega\cos\phi) + \cos(\omega\sin\phi)) + \gamma\cos(\omega\cos\phi)\cos(\omega\sin\phi)\ , \tag{43}$$

one realizes that the value of γ that gives the smallest angular variation for a fixed value of ω, depends on ω. Hence, with this formulation, the "rotational invariance" is scale dependent. At coarse scales one obtains:

Proposition 21 Rotational invariance in the Fourier domain (2D). *The value of γ that gives the least rotational asymmetry for large scale phenomena in the solution to the differential equation (27) is $\gamma = \frac{1}{3}$.*

Proof. The Taylor expansion of h for small values of ω is (see [13, Appendix A.2.3])

$$h(\omega \cos\phi, \omega \sin\phi) = -\frac{1}{2}\omega^2 + \frac{1}{24}(1 + (6\gamma - 2)\cos^2\phi \sin^2\phi)\omega^4 + O(\omega^6) \ , \quad (44)$$

where the $O(\omega^6)$ term depends on both ϕ and γ. Observe that if $\gamma = \frac{1}{3}$ then the ϕ-dependence decreases with ω as ω^6 instead of as ω^4. □

This means that $\gamma = \frac{1}{3}$ *asymptotically*, i.e., with increasing spatial scale, gives the most isotropic smoothing effect on coarse-scale events. The reason why spatial isotropy is desired at coarse scales rather than at fine scales is because the grid effects become smaller for coarse-scale phenomena, which in turn makes it more meaningful to talk about rotational invariance. This selection of γ corresponds to approximating the Laplacean operator with the "the nine-point operator" (see Fig. 1(d) and Dahlquist [5]). Note that when the separability is violated by using a non-zero value of γ, the discrete scale-space representation can anyway be computed efficiently in the Fourier domain using (29).

6 Summary and Discussion

The proper way to apply the scale-space theory to discrete signals is apparently by discretizing the diffusion equation. Starting from a requirement that local extrema must not be enhanced when the scale parameter is increased continuously, it has been shown that within the class of linear transformations a necessary and sufficient condition for a one-parameter family of representations $L : \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ to be a scale-space family of a discrete signal $f : \mathbb{Z}^N \to \mathbb{R}$ is that it satisfies the differential equation

$$\partial_t L = \mathcal{A}_{ScSp} L \ , \quad (45)$$

with initial condition $L(\cdot;\ 0) = f(\cdot)$ for some infinitesimal scale-space generator $\mathcal{A}_{ScSp}$. In one, two and three dimensions respectively it can equivalently be stated that a family is a scale-space family if and only if for some linear reparameterization of the scale parameter t and for some $\gamma_i \in [0, 1]$ it satisfies

$$\partial_t L = \frac{1}{2}\nabla_3^2 L \ , \quad (46)$$

$$\partial_t L = \frac{1}{2}\left((1 - \gamma_1)\nabla_5^2 L + \gamma_1 \nabla_{x^2}^2 L\right) \ , \quad (47)$$

$$\partial_t L = \frac{1}{2}\left((1 - \gamma_1 - \gamma_2)\nabla_7^2 L + \gamma_1 \nabla_{+^3}^2 L + \gamma_2 \nabla_{x^3}^2 L\right) \ . \quad (48)$$

The essence of (45)–(48) is that these equations correspond to discretizations of *first-order* differential operators in *scale*, and *second-order* differential operators in *space*.

The effect of using different values of γ_1 in the two-dimensional case has been analysed in detail. Nevertheless, the question about definite selection is left open. Unimodality considerations indicate that γ must not exceed $\frac{1}{2}$, while $\gamma = \frac{1}{3}$ gives the least degree of rotational asymmetry in the Fourier domain.

The family of scale-space kernels is separable if and only if $\gamma = 0$. In this case the scale-space family is given by convolution with the one-dimensional discrete analogue of the Gaussian kernel along each dimension. For this parameter setting the closed-form expressions for several derived entities simplify (see, e.g., [12, 15]). Observe also that $\gamma = 0$ arises a necessary consequence if the neighbourhood concept (defined in Sec. 3.1) is redefined as $N(x) = \{\xi \in Z^N : (\| x_i - \xi \|_1 \leq 1) \wedge (\xi \neq x)\}$ (corresponding to what is known as four-connectivity in the two-dimensional case), since then necessarily $\alpha_i = 0$ $(i > 1)$ in (18) and (19). Similar results hold in higher dimensions. A possible disadvantage with choosing $\gamma = 0$ is that it emphasizes the role of the coordinate axes as being special directions.

Finally, it should be remarked that if a linear and shift-invariant operator $\mathcal{L}$, commuting with the smoothing operator $T*$, is applied to the scale-space representation L of a signal f, then $\mathcal{L}L$ will be a scale-space representation of $\mathcal{L}f$. One consequence of this is that multiscale discrete derivative approximations defined by linear filtering of the smoothed signal preserve the scale-space properties. This property, which provides a natural way to discretize the multiscale N-jet representation proposed by Koenderink and van Doorn [10], is developed in [15].

7 Possible Extensions

The treatment so far has been restricted to signals defined on infinite and uniformly sampled square grids using uniform smoothing of all grid points. Below the ways in which these notions can be generalized are outlined.

7.1 Anisotropic Smoothing

Perona and Malik [19] propose anisotropic smoothing as a way to reduce the shape distortions arising in edge detection by smoothing across object boundaries (see also Nordström [18]). The suggested methodology is to modify the diffusion coefficients in order to favour intraregion smoothing over interregion smoothing.

Using the maximum principle they show that the resulting anisotropic scale-space representation possesses a suppression property for local extrema similar to that used in Koenderink's [9] continuous scale-space formulation and this discrete treatment. From the proofs of Theorems 15-16 it is obvious that the discrete scale-space concept can easily be extended to such anisotropic diffusion by letting the coefficients in the operator $\mathcal{A}_{ScSp}$ depend upon the input signal. By this, the locality, positivity, and zero sum conditions will be preserved, while the symmetry requirements must be relaxed. Introducing such an anisotropic diffusion equation, however, violates the convolution form of smoothing as well as the semi-group property. Therefore, when proving the necessity of the representation a certain form of the smoothing formula may have to be assumed, for

example, of the form (9) with the filter coefficients depending upon the input signal. Note that, if the translational invariance and the symmetry with respect to coordinate interchanges are relaxed in (45), then this equation corresponds the (spatial) discretization of the (second-order) diffusion equation with *variable conductance*, $c(x;\ t)$,

$$(\partial_t L)(x;\ t) = \nabla(c(x;\ t)\, \nabla L(x;\ t)) \ . \tag{49}$$

Throughout this work uniform smoothing has been used at the cost of possible smoothing across object boundaries. The motivation behind this choice has been the main interest in using scale-space for *detecting* image structures. Therefore, in the absence of any prior information, it is natural that the first processing steps should be as *uncommitted* as possible. The approach taken has been to first detect candidate regions of interest, and then, once candidates have been detected as regions, improve their localization. Possibly, variable conductance could be useful in the second step of this process. Another natural application is to avoid the negative effects of smoothing thin or elongated structures.

There are, however, some problems that need to be further analyzed. Modifying the diffusion coefficients requires some kind of a priori information concerning which structures in the image are to be smoothed and which are not. In the method by Perona and Malik there is a tuning function to be determined, giving the diffusion coefficient as function of the gradient magnitude. When the scale parameter t tends to infinity, the solution to the anisotropic diffusion equation tends to a function with various sharp edges. Hence, choosing a tuning function somehow implies an implicit assumption about a "final segmentation" of the image. It is not clear that such a concept exists or can be rigorously defined.

7.2 Finite Data

A practical problem always arising in linear filtering is what to do with pixels near the image boundary for which a part of the filter mask stretches outside the available image.

The most conservative outlook is, of course, to regard the output as undefined as soon as a computation requires image data outside the available domain. This is, however, hardly desirable for scale-space smoothing, since the (untruncated) convolution masks have infinite support, while the peripheral coefficients decrease towards zero very rapidly. A variety of ad hoc methods have been proposed to deal with this; extension methods, subtraction of steady-state components, solving the diffusion equation on a limited domain with (say, adiabatic) boundary conditions, etc. However, no such technique can overcome the problem with missing data. In some simple situations ad hoc extensions may do. But this requires some kind of a priori information about the contents of the image.

Inevitably, the peripheral image values of a smoothed finite image will be less reliable than the central ones. Instead, if accurate values really are required near the image boundary, then the vision system should try to acquire additional data such that the convolution operation becomes well-defined up to the

prescribed accuracy. This is easily achieved within the active vision paradigm simply by moving the camera so that more values become available in a sufficiently large neighbourhood of the object of interest. The task of analyzing an object manifesting itself at a certain scale requires input data in a region around the object. The width of this frame depends both on the current level of scale and the prescribed accuracy of the analysis.

Of course, a genuinely finite approach is also possible. In this presentation this subject has not been developed, since the associated problems are somehow artificial and difficult to handle in a consistent manner, although the non-enhancement property can be easily formulated for finite data and although in the one-dimensional case the concepts of sign-regularity and semi-groups of totally positive matrices [8] in principle provide possible tools for dealing with this issue. One way to avoid both the infiniteness and the boundary problems is by using a spherical camera. Then, the ordinary planar camera geometry appears as an approximate description for foveal vision, that is, small solid angles in the central field of vision.

7.3 Other Types of Grids

The assumption of a square grid is not a necessary restriction. The same type of treatment can be carried out on, for example, a hexagonal grid with the semi-group property preserved, and also on a grid corresponding to non-uniform spatial sampling provided that the diffusion coefficients are modified accordingly. In the latter case some a priori form of the smoothing formula may have to be adopted when proving the necessity of the representation. An interesting case to consider might actually be the non-uniformly sampled spherical camera.

7.4 Further Work

Finally, it should be pointed out that there is one main issue that has not been considered here, namely *scale-dependent spatial sampling*. This issue is certainly of importance in order to improve the computational efficiency both when computing the representation and for algorithms working on the data. The scale-space concept outlined here uses the same spatial resolution at all levels of scale. The pyramid representations (see, e.g., Burt [3]) on the other hand imply a fixed relation between scale and resolution beyond which refinements are not possible.

Since the smoothed images at coarser scales become progressively more redundant, it seems plausible that some kind of subsampling can be done at the coarser scales without too much loss of information. It would be interesting to carry out an analysis about how much information is lost by such an operation, and to which extent a subsampling operator can be introduced in this representation, while still maintaining the theoretical properties associated with having a continuous scale parameter, and without introducing any severe discontinuities along the scale direction that would be a potential source to numerical difficulties for algorithms working on the output from the representation.

References

1. Abramowitz, M., Stegun, I.A. (1964). Handbook of Mathematical Functions, Applied Mathematics Series 55, National Bureau of Standards.
2. Babaud, J., Witkin, A.P., Baudin, M., Duda, R.O. (1986). Uniqueness of the Gaussian kernel for scale-space filtering, IEEE Trans. Patt. Anal. Mach. Intell. 8 (1), pp. 26–33.
3. Burt, P.J., Adelson, E.H. (1983). The Laplacian pyramid as a compact image code, IEEE Trans. Comm. 9 (4), pp. 532–540.
4. Crowley, J.L., Stern, R.M. (1984). Fast computation of the difference of low pass transform, IEEE Trans. Patt. Anal. Mach. Intell. 6, pp. 212–222.
5. Dahlquist, G., Björk, Å., Anderson, N. (1974). Numerical Methods, Prentice-Hall.
6. Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A. (1992). Scale-space: its natural operators and differential invariants, Image and Vision Computing 10 (6), pp. 376–388.
7. Hille, E., Phillips, R.S. (1957). Functional Analysis and Semi-Groups, Am. Math. Soc. Coll. Publ., Vol. XXXI.
8. Karlin, S. (1968). Total Positivity, Vol.I, Stanford Univ. Press.
9. Koenderink, J.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
10. Koenderink, J.J., van Doorn, A.J. (1990). Receptive field families, Biol. Cybern. 63, pp. 291–297.
11. Lifshitz, L.M., Pizer, S.M. (1987). A multiresolution hierarchical approach to image segmentation based on intensity extrema, Technical report, Depts. Comp. Sci. and Radiology, Univ. North Carolina, Chapel Hill, N.C., USA
12. Lindeberg, T.P. (1990). Scale-space for discrete signals, IEEE Trans. Patt. Anal. Mach. Intell. 12 (3), pp. 234–254.
13. Lindeberg, T.P. (1991). Discrete Scale-Space Theory and the Scale-Space Primal Sketch, Ph.D. Thesis, ISRN KTH/NA/P–91/8–SE, Dept. Num. Anal. Comp. Sci., Royal Inst. Tech., S-100 44 Stockholm, Sweden. A revised and extended version to appear in The Kluwer International Series in Engineering and Computer Science.
14. Lindeberg, T.P. (1992). Scale-space behaviour of local extrema and blobs, J. Math. Imaging Vision 1, pp. 65–99.
15. Lindeberg, T.P. (1993). Discrete derivative approximations with scale-space properties, J. Math. Imaging and Vision, to appear.
16. Lindeberg T.P. (1993). Scale-space behaviour and invariance properties of differential singularities, this volume, pp. 591–600.
17. Mallat, S.G. (1989). A theory of multiresolution signal processing: The wavelet representation, IEEE Trans. Patt. Anal. Mach. Intell. 11 (6), pp. 674–693.
18. Nordström, N. (1990). Biased anisotropic diffusion — A unified regularization and diffusion approach to edge detection. In: Faugeras, O. (ed.), Proc 1st Eur. Conf. Comp. Vision, Antibes, France, Apr. 23-37, pp. 18–27, Springer-Verlag.
19. Perona, P., Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Patt. Anal. Mach. Intell. 12 (7), pp. 629–639.
20. Widder, D.V. (1975). The Heat Equation, Academic Press, New York.
21. Witkin, A.P. (1983). Scale-space filtering, Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany, Aug. 8-12, pp. 1019–1022.
22. Yuille, A., Poggio, T. (1986). Scaling theorems for zero-crossings, IEEE Trans. Patt. Anal. Mach. Intell. 9 (1), pp. 15–25.
23. Yuille, A.L. (1988). The creation of structure in dynamic shape, Proc 2nd Int. Conf. Comp. Vision, Tampa, Florida, Dec. 5-8, pp. 685–689.

Scale-Space Behaviour and Invariance Properties of Differential Singularities *

Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)
Department of Numerical Analysis and Computing Science
Royal Institute of Technology, S-100 44 Stockholm, Sweden
Email: tony@bion.kth.se

Abstract. This article describes how a certain way of expressing low-level feature detectors, in terms of singularities of differential expressions defined at multiple scales in scale-space, simplifies the analysis of the effect of smoothing. It is shown how such features can be related across scales, and generally valid expressions for drift velocities are derived with examples concerning edges, junctions, Laplacean zero-crossings, and blobs. A number of invariance properties are pointed out, and a particular representation defined from such singularities, the scale-space primal sketch, is treated in more detail.

Keywords: scale-space, drift velocity, feature detection, primal sketch, singularity, invariance.

1 Introduction

A common way of implementing low-level feature detectors in computer vision and image processing is by applying non-linear operations to smoothed input data. Examples of this are edge detection, junction detection, and blob detection. The pre-smoothing step can be motivated either heuristically by the need for noise suppression in real-world signals, or by the fact that image structures only exist as meaningful entities over certain ranges of scale. The latter argument is one of the main motivations for the development of the multi-scale representation known as *scale-space representation*, in which a given signal is subjected to smoothing by Gaussian kernels of successively increasing width.

This aim of this article is to show why a certain way of formulating such low-level feature detectors, in terms of singularities of differential expressions defined from the scale-space representation, is attractive from theoretical viewpoint. An overview of the scheme proposed is shown in Fig. 1. Any given signal is subjected to the following operations: (i) smoothing to a number of scales, (ii) derivative computations at each scale, (iii) combination of the derivatives at each scale into (non-linear) differential geometric entities, and (iv) detection of zero-crossings in

* The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

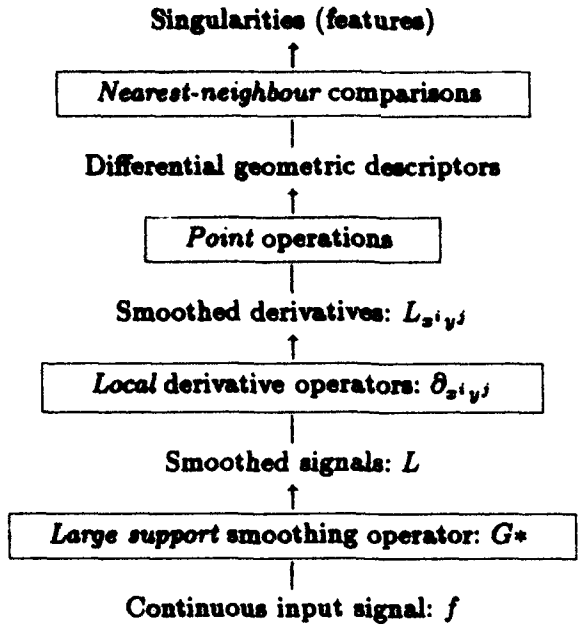


Fig. 1. Schematic overview of the different types of computations required for detecting features in terms of differential singularities at multiple scales.

these. It will be shown how the effect of the smoothing operation in this scheme can be analyzed by (i) showing how features defined in this way can be related across scales; a subject which can be referred to as the "deep structure of scale-space", and (ii) by deriving drift velocities for a large class of feature detectors. A number of invariance properties with respect to natural transformations of the spatial coordinates and the grey-level domain will also be listed.

Before starting it should be pointed out that this scheme is not presented as solely new. Some of the results presented in the paper are (at least partly) known, or have been touched upon before; see, e.g., Koenderink and van Doorn [8], who proposed the multi-scale N-jet representation, Florack *et al.* [5], who showed how a minimal set of differential invariants can be derived at any scale, or Lindeberg [11] who analysed the behaviour of scales of local extrema and related entities. The purpose with this presentation is rather to emphasize the role of the *singularities* in the scheme in Fig. 1, and to illustrate how they are attractive for the theoretical analysis of different feature detectors. For simplicity, the treatment is developed for two-dimensional signals. The approach, however, is valid in arbitrary dimensions.

The scale-space concept dealt with is the traditional diffusion-based scale-space for continuous signals developed by Witkin [16], Koenderink [7], and Babaud *et al.* [1], which is given by the solution to

$$\partial_t L = \frac{1}{2}\nabla^2 L = \frac{1}{2}(L_{xx} + L_{yy}) \tag{1}$$

with initial condition $L(\cdot;\, 0) = f(\cdot)$. At any scale in this representation and at any point $P_0 = (x_0, y_0) \in \mathbb{R}^2$, denote by $\partial_{\tilde{v}}$ the directional derivative in the gradient direction of L, and by $\partial_{\tilde{u}}$ the derivative in the perpendicular direction. In terms of derivatives along the Cartesian coordinate directions it holds that

$$\partial_{\tilde{u}} = \sin\phi\partial_x - \cos\phi\partial_y \;\; ; \quad \partial_{\tilde{v}} = \cos\phi\partial_x + \sin\phi\partial_y \;\; , \tag{2}$$

where $(\cos\phi, \sin\phi)$ is the normalized gradient direction of L at P_0.

2 Feature Detection from Singularities in Scale-Space

2.1 Examples of Differential Formulations of Feature Detectors

A natural way to define *edges* from a continuous grey-level image $L : \mathbb{R}^2 \to \mathbb{R}$ is as the union of the points for which the gradient magnitude assumes a maximum in the gradient direction. This method is usually referred to as "non-maximum suppression", (see e.g. Canny [4]). Assuming that the second and third order directional derivatives of L in the v-direction are not simultaneously zero, a necessary and sufficient condition for P_0 to be a gradient maximum in the gradient direction may be stated as:

$$\begin{cases} L_{\tilde{v}\tilde{v}} = 0 \;\; , \\ L_{\tilde{v}\tilde{v}\tilde{v}} < 0 \;\; . \end{cases} \tag{3}$$

Since only the sign information is important, this condition can be restated as

$$\begin{cases} \tilde{L}_{\tilde{v}\tilde{v}} = \;\; L_{\tilde{v}}^2 L_{\tilde{v}\tilde{v}} = \;\; L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy} = 0 \;\; , \\ \tilde{L}_{\tilde{v}\tilde{v}\tilde{v}} = L_{\tilde{v}}^3 L_{\tilde{v}\tilde{v}\tilde{v}} = L_x^3 L_{xxx} + 3L_x^2 L_y L_{xxy} + 3L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} < 0 \;\; . \end{cases} \tag{4}$$

An entity commonly used for *junction detection* is the curvature of level curves in intensity data (see e.g. Kitchen [6] or Koenderink and Richards [9]). In terms of directional derivatives it can be expressed as

$$\kappa = \frac{L_{\tilde{u}\tilde{u}}}{L_{\tilde{v}}} \;\; . \tag{5}$$

In order to to give a stronger response near edges, the level curve curvature is usually multiplied by the gradient magnitude (see, e.g., Brunnström *et al.* [3])

$$|\tilde{\kappa}| = |L_{\tilde{v}}^2 L_{\tilde{u}\tilde{u}}| = |L_y^2 L_{xx} - 2L_x L_y L_{xy} + L_x^2 L_{yy}| \;\; . \tag{6}$$

Assuming that the first- and second-order differentials of $\tilde{\kappa}$ are not simultaneously degenerate, a necessary and sufficient condition for a point P_0 to be a maximum in this rescaled level curve curvature is that:

$$\begin{cases} \partial_{\tilde{u}}(\tilde{\kappa}) = 0 \;\; , \quad \partial_{\tilde{v}}(\tilde{\kappa}) = 0 \;\; , \\ \mathcal{H}(\tilde{\kappa}) = \tilde{\kappa}_{\mathcal{H}} = \tilde{\kappa}_{\tilde{u}\tilde{u}}\tilde{\kappa}_{\tilde{v}\tilde{v}} - \tilde{\kappa}_{\tilde{u}\tilde{v}}^2 > 0 \;\; , \\ \mathrm{sign}(\tilde{\kappa})\tilde{\kappa}_{\tilde{u}\tilde{u}} < 0 \;\; . \end{cases} \tag{7}$$

Zero-crossings of the Laplacean

$$\nabla^2 L = L_{\tilde{u}\tilde{u}} + L_{\tilde{v}\tilde{v}} = L_{xx} + L_{yy} = 0 \tag{8}$$

have been used for *stereo matching* (see, e.g., Marr [15]) and *blob detection* (see, e.g., Blostein and Ahuja [2]). Blob detection methods can also be formulated in terms of local extrema (see, e.g., Lindeberg and Eklundh [10]).

2.2 Invariance Properties of Differential Singularities

One of the main reasons why the formulation in terms of singularities is important is because these singularities do not depend on the actual numerical values of the differential geometric entities, but only on their *relative* relations. In this way, they will be less affected by scale-space smoothing, which is well-known to decrease the amplitude of the variations in a signal and its derivatives.

In fact, the differential entities used above are *invariant* to a number of primitive transformations of both the original and the smoothed grey-level signal; *translations*, *rotations*, and (uniform) *rescalings* in space as well as *affine intensity transformations*. (This set is similar but not equal to the set of transformations used by Florack *et al.* [5] to derive necessity results about differential invariants from intensity data; the main difference is that [5] considers invariance with respect to arbitrary *monotone intensity transformations*, while the differential singularities used here are invariant to uniform rescalings of the coordinate axes, i.e., *size changes*.)

To give a precise formulation of this, let $L_{\bar{u}^m \bar{v}^n} = L_{\bar{u}^\alpha}$ denote a mixed directional derivative of order $|\alpha| = m + n$, where $\alpha = (m, n)$, and let $\mathcal{D}$ be a (possibly non-linear) homogeneous differential expression of the form

$$\mathcal{D}L = \sum_{i=1}^{I} c_i \prod_{j=1}^{J} L_{\bar{u}^{\alpha_{ij}}} \quad , \tag{9}$$

where $|\alpha_{ij}| > 0$ for all $i = [1..I]$ and $j = [1..J]$, and $\sum_{j=1}^{J} |\alpha_{ij}| = N$ for all $i \in [1..I]$. Moreover, let $\mathcal{S}_\mathcal{D} L$ denote the *singularity set* of this operator, i.e., $\mathcal{S}_\mathcal{D} L = \{(x;\ t) : \mathcal{D}L(x;\ t) = 0\}$, and let $\mathcal{G}$ be the Gaussian smoothing operator, i.e., $L = \mathcal{G}f$. Under these transformations of the spatial domain (represented by $x \in \mathbb{R}^2$) and the intensity domain (represented by either the unsmoothed f or the smoothed L) the singularity sets transform as follows:

Transformation	Definition	Invariance
translation	$(\mathcal{T}L)(x;\ t) = L(x + \Delta x;\ t)$	$\mathcal{S}_\mathcal{D}\mathcal{G}\mathcal{T}f = \mathcal{S}_\mathcal{D}\mathcal{T}\mathcal{G}f = \mathcal{T}\mathcal{S}_\mathcal{D}\mathcal{G}f$
rotation	$(\mathcal{R}L)(x;\ t) = L(Rx;\ t)$	$\mathcal{S}_\mathcal{D}\mathcal{G}\mathcal{R}f = \mathcal{S}_\mathcal{D}\mathcal{R}\mathcal{G}f = \mathcal{R}\mathcal{S}_\mathcal{D}\mathcal{G}f$
uniform scaling	$(\mathcal{U}L)(x;\ t) = L(sx;\ t)$	$\mathcal{S}_\mathcal{D}\mathcal{G}\mathcal{U}f = \mathcal{S}_\mathcal{D}\mathcal{U}\mathcal{G}f = \mathcal{U}\mathcal{S}_\mathcal{D}\mathcal{G}f$
affine intensity	$(\mathcal{A}L)(x;\ t) = aL(x;\ t) + b$	$\mathcal{S}_\mathcal{D}\mathcal{G}\mathcal{A}f = \mathcal{S}_\mathcal{D}\mathcal{A}\mathcal{G}f = \mathcal{S}_\mathcal{D}\mathcal{G}f$

Above, R is a rotation matrix, Δx is a vector ($\in \mathbb{R}^2$), while a, b and s are scalar constants. The definitions of the transformed singularity sets are as follows; $\mathcal{T}\mathcal{S}_\mathcal{D}L = \{(x;\ t) : \mathcal{D}L(x + \Delta x;\ t) = 0\}$, $\mathcal{R}\mathcal{S}_\mathcal{D}L = \{(x;\ t) : \mathcal{D}L(Rx;\ t) = 0\}$, and $\mathcal{U}\mathcal{S}_\mathcal{D}L = \{(x;\ t) : \mathcal{D}L(sx;\ s^2 t) = 0\}$. The commutative properties of $\mathcal{G}$ with $\mathcal{T}$, $\mathcal{R}$, and $\mathcal{A}$ are trivial consequences of the translational invariance, rotational invariance, and linearity of Gaussian smoothing. Under uniform rescalings $f'(x, y) = f(sx, sy)$, however, it holds that $L' = \mathcal{G}f'$ is related to L by $L'(x, y;\ t_0) = L(sx, sy;\ s^2 t_0)$, which means that $\mathcal{U}$ applied to a singularity set also affects the scale levels. The commutative properties of $\mathcal{S}_\mathcal{D}$ with $\mathcal{T}$, $\mathcal{U}$, and $\mathcal{A}$ follow from corresponding invariance properties of linear derivative operators combined with the homogeneity of (9), while the commutativity with respect to $\mathcal{R}$ follows from the rotational invariance of the directional derivatives $L_{\bar{u}^{\alpha_{ij}}}$.

3 Relating Differential Singularities at Different Scales

Consider a feature, which at any level of scale can be defined by

$$h(x, y;\ t) = 0 \tag{10}$$

for some function $h : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}^N$, where N is either 1 or 2. Using the implicit function theorem it is easy to analyze the dependence of (x, y) on t in the solution to (10). The results to be derived give estimates of the drift velocity of different features due to scale-space smoothing, and provides a theoretical basis for relating corresponding features at adjacent scales. It does hence enable well-defined linking and/or identification of features across scales.

3.1 Zero-Dimensional Entities (Points)

Assume first that N is equal to 2, that is, that $h(x, y;\ t) = (h_1(x, y;\ t), h_2(x, y;\ t))$ for some functions $h_1, h_2 : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$. The derivative of the mapping h at a point $P_0 = (x_0, y_0;\ t_0)$ is

$$h'|_{P_0} = \begin{pmatrix} \partial_x h_1 & \partial_y h_1 & \partial_t h_1 \\ \partial_x h_2 & \partial_y h_2 & \partial_t h_2 \end{pmatrix}\Big|_{P_0} = \begin{pmatrix} \frac{\partial(h_1, h_2)}{\partial(x, y)} & \frac{\partial(h_1, h_2)}{\partial(t)} \end{pmatrix}\Big|_{P_0} . \tag{11}$$

If $\partial(h_1, h_2)/\partial(x, y)$ is a non-singular matrix at P_0, then the solution (x, y) to $h(x, y;\ t_0) = 0$ will be an isolated point. Moreover, the implicit function theorem guarantees that there exists some local neighbourhood around P_0 where (x, y) can be expressed as a function of t. The derivative of that mapping $t \mapsto (x, y)$ is:

$$\begin{pmatrix} \partial_t x \\ \partial_t y \end{pmatrix}\Big|_{P_0} = - \begin{pmatrix} \partial_x h_1 & \partial_y h_1 \\ \partial_x h_2 & \partial_y h_2 \end{pmatrix}\Big|_{P_0}^{-1} \begin{pmatrix} \partial_t h_1 \\ \partial_t h_2 \end{pmatrix}\Big|_{P_0} \tag{12}$$

If h is a function of the spatial derivatives of L only, which is the case, for example, for the feature extractors treated in Sect. 2.1, then the fact that spatial derivatives of L satisfy the diffusion equation $\partial_t = (\partial_{xx} + \partial_{yy})/2$, can be used for replacing derivatives with respect to t by derivatives with respect to x and y. Hence, closed form expression can be obtained containing only partial derivatives of L with respect to x and y. For example, the *junction candidates* given by (7) satisfy $(\tilde{\kappa}_{\tilde{u}}, \tilde{\kappa}_{\tilde{v}}) = (0, 0)$. In terms of directional derivatives, (12) can then be written

$$\begin{pmatrix} \partial_t u \\ \partial_t v \end{pmatrix}\Big|_{P_0} = - \begin{pmatrix} \tilde{\kappa}_{\tilde{u}\tilde{u}} & \tilde{\kappa}_{\tilde{u}\tilde{v}} \\ \tilde{\kappa}_{\tilde{u}\tilde{v}} & \tilde{\kappa}_{\tilde{v}\tilde{v}} \end{pmatrix}\Big|_{P_0}^{-1} \begin{pmatrix} \partial_t \tilde{\kappa}_{\tilde{u}} \\ \partial_t \tilde{\kappa}_{\tilde{v}} \end{pmatrix}\Big|_{P_0} . \tag{13}$$

By differentiating the expressions for $\tilde{\kappa}_{\tilde{u}}$ and $\tilde{\kappa}_{\tilde{v}}$ with respect to t, by using the fact that the spatial derivatives satisfy the diffusion equation, and by expressing

the result in directional derivatives in the u- and v-directions the following expressions can be obtained (the calculations have been done using Mathematica):

$$
\begin{aligned}
\tilde{\kappa}_{uu} &= L_v^2 L_{uuuu} + 2L_{uu}(L_{uu}L_{vv} - L_{uv}^2) + 2L_v(L_{uv}L_{uuu} - L_{uu}L_{uuv}),\\
\tilde{\kappa}_{uv} &= L_v^2 L_{uuuv} + 2L_{uv}(L_{uu}L_{vv} - L_{uv}^2) + 2L_v(L_{vv}L_{uuu} - L_{uv}L_{uuv}),\\
\tilde{\kappa}_{vv} &= L_v^2 L_{uuvv} + 2L_{vv}(L_{uu}L_{vv} - L_{uv}^2) + 2L_v(L_{uu}L_{vvv} + 2L_{vv}L_{uuv} - 3L_{uv}L_{uvv}),\\
\partial_t\tilde{\kappa}_u &= L_v^2(L_{uuuuu} + L_{uuuvv})/2\\
&\quad +(L_{uu}L_{vv} - L_{uv}^2)(L_{uuu} + L_{uvv}) + L_v(L_{uuu}L_{vvv} - L_{uuv}L_{uvv}),\\
\partial_t\tilde{\kappa}_v &= L_v^2(L_{uuuuv} + L_{uuvvv})/2 + (L_{uu}L_{vv} - L_{uv}^2)(L_{uuv} + L_{vvv})\\
&\quad +L_v(L_{vv}(L_{uuuu} + L_{uuvv}) + L_{uu}(L_{vvvv} + L_{uuvv}) - 2L_{uv}(L_{uuuv} + L_{uvvv}))\\
&\quad +L_v(L_{uuv}(L_{uuv} + L_{vvv}) - L_{uvv}(L_{uuu} + L_{uvv})).
\end{aligned}
$$

(These expressions simplify somewhat if we make use of $L_{uuu}|_{P_0} = 0$, which follows from $\tilde{\kappa}_u = 0$.) Note that as long as the Hessian matrix of $\tilde{\kappa}$ is non-degenerate, the sign of the $\tilde{\kappa}_{\mathcal{H}}$ and $\tilde{\kappa}_{uu}$ will be constant. This means that the type of extremum will be the same. For *local extrema* of the grey-level landscape, given by $(L_x, L_y) = (0, 0)$, the expression for the drift velocity reduces to

$$r_t = -\frac{1}{2}(\mathcal{H}L)^{-1}\nabla^2(\nabla L) \ , \tag{14}$$

where $\mathcal{H}L$ denotes the Hessian matrix and ∇L the gradient vector. Observe that regularity presents no problem, since L satisfies the diffusion equation, and for $t > 0$ any solution to the diffusion equation is infinitely differentiable.

3.2 One-Dimensional Entities (Curves)

If N is equal to 1, then there will no longer be any unique correspondence between points at adjacent scales. An ambiguity occurs, very similar to what is called the aperture problem in motion analysis. Nevertheless, we can determine the drift velocity in the normal direction of the curve. Given a function $h : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ consider the solution to $h(x, y;\ t) = 0$. Assume that $P_0 = (x_0, y_0;\ t_0)$ is a solution to this equation and that the gradient of the mapping $(x, y) \mapsto h(x, y;\ t_0)$ is non-zero. Then, in some neighbourhood around (x_0, y_0) the solution (x, y) to $h(x, y;\ t_0) = 0$ defines a curve. Its normal at (x_0, y_0) is given by $(\cos\phi, \sin\phi) = (h_x, h_y)/(h_x^2 + h_y^2)^{1/2}$ at P_0. Consider the function $\tilde{h} : R \times \mathbb{R}_+ \to \mathbb{R}$ $\tilde{h}(s;\ t) = h(x_0 + s\cos\phi, y_0 + s\sin\phi;\ t)$, which has the derivative

$$\tilde{h}_s(0;\ t_0) = h_x(x_0, y_0;\ t_0)\cos\phi + h_y(x_0, y_0;\ t_0)\sin\phi = \sqrt{h_x^2 + h_y^2}\Big|_{P_0} \ . \tag{15}$$

Since this derivative is non-zero, we can apply the implicit function theorem. It follows that there exists some neighbourhood around P_0 where $\tilde{h}(s;\ t) = 0$ defines s as a function of t. The derivative of this mapping is

$$\partial_t s|_{P_0} = -\tilde{h}_s\Big|_{P_0}^{-1}\tilde{h}_t\Big|_{P_0} = -\left.\frac{h_t}{\sqrt{h_x^2 + h_y^2}}\right|_{P_0} \ . \tag{16}$$

As an example of this consider an *edge* given by non-maximum suppression

$$h = \alpha = L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy} = 0 \ . \tag{17}$$

By differentiating (17), by using the fact that the derivatives of L satisfy the diffusion equation, and by expressing the result in terms of the directional derivatives we get

$$\begin{aligned} \alpha &= L_{\tilde{v}}^2 L_{\tilde{v}\tilde{v}} = 0, \\ \alpha_{\tilde{u}} &= L_{\tilde{v}}^2 L_{\tilde{u}\tilde{v}\tilde{v}} + 2L_{\tilde{v}} L_{\tilde{u}\tilde{v}} L_{\tilde{u}\tilde{u}}, \\ \alpha_{\tilde{v}} &= L_{\tilde{v}}^2 L_{\tilde{v}\tilde{v}\tilde{v}} + 2L_{\tilde{v}} L_{\tilde{u}\tilde{v}}^2, \\ \alpha_t &= L_{\tilde{v}}^2 (L_{\tilde{u}\tilde{u}\tilde{v}\tilde{v}} + L_{\tilde{v}\tilde{v}\tilde{v}\tilde{v}})/2 + L_{\tilde{v}} L_{\tilde{u}\tilde{v}} (L_{\tilde{u}\tilde{u}\tilde{u}} + L_{\tilde{u}\tilde{v}\tilde{v}}) \ . \end{aligned} \tag{18}$$

To summarize, the drift velocity in the normal direction of a *(curved) edge* in scale-space is (with $\alpha_{\tilde{u}}$ and $\alpha_{\tilde{v}}$ according to (18))

$$(\partial_t u, \partial_t v) = -\frac{L_{\tilde{v}}(L_{\tilde{u}\tilde{u}\tilde{v}} + L_{\tilde{v}\tilde{v}\tilde{v}}) + 2L_{\tilde{u}\tilde{v}}(L_{\tilde{u}\tilde{u}\tilde{u}} + L_{\tilde{u}\tilde{v}\tilde{v}})}{2((L_{\tilde{v}} L_{\tilde{u}\tilde{v}\tilde{v}} + 2L_{\tilde{u}\tilde{v}} L_{\tilde{u}\tilde{u}})^2 + (L_{\tilde{v}} L_{\tilde{v}\tilde{v}\tilde{v}} + 2L_{\tilde{u}\tilde{v}}^2)^2)} (\frac{\alpha_{\tilde{u}}}{L_{\tilde{v}}}, \frac{\alpha_{\tilde{v}}}{L_{\tilde{v}}}) \ . \tag{19}$$

Unfortunately, this expression cannot be further simplified unless additional constraints are posed on L. For a *straight edge*, however, where all partial derivatives with respect to u are zero, it reduces to

$$(\partial_t u, \partial_t v) = -\frac{1}{2} \frac{L_{\tilde{v}\tilde{v}\tilde{v}\tilde{v}}}{L_{\tilde{v}\tilde{v}\tilde{v}}} (0, 1) \tag{20}$$

(see also [11, 14]). For a curve given by the *zero-crossings of the Laplacean* we have

$$(\partial_t u, \partial_t v) = -\frac{\nabla^2 (\nabla^2 L)}{2((\nabla^2 L_{\tilde{u}})^2 + (\nabla^2 L_{\tilde{v}})^2)} (\nabla^2 L_{\tilde{u}}, \nabla^2 L_{\tilde{v}}) \ , \tag{21}$$

which also simplifies to (20) if all directional derivatives in the u-direction are set to zero. Similarly, for a *parabolic curve*, given by $\det(\mathcal{H}L) = L_{xx} L_{yy} - L_{xy}^2 = 0$, the drift velocity in the normal direction is

$$\begin{aligned} (\partial_t x, \partial_t y) = \ - \ & \frac{L_{yy} L_{xxxx} - 2L_{xy} L_{xxxy} + (L_{xx} + L_{yy}) L_{xxyy} - 2L_{xy} L_{xyyy} + L_{xx} L_{yyyy}}{2((L_{xx} L_{xyy} - 2L_{xy} L_{xxy} + L_{yy} L_{xxx})^2 + (L_{xx} L_{yyy} - 2L_{xy} L_{xyy} + L_{yy} L_{xxy})^2)} \\ & (L_{xx} L_{xyy} - 2L_{xy} L_{xxy} + L_{yy} L_{xxx}, L_{xx} L_{yyy} - 2L_{xy} L_{xyy} + L_{yy} L_{xxy}). \end{aligned}$$

This expression simplifies somewhat in a pq-coordinate system, with the p- and q-axes aligned to the principal axes of curvature so that the mixed second-order directional derivative $L_{\tilde{p}\tilde{q}}$ is zero.

4 Invariance Properties of the Scale-Space Primal Sketch

A particular type of representation constructed in this way is the scale-space primal sketch [10, 14]. It is a tree-like multi-scale representation aimed at *making explicit* blobs in scale-space as well as the relations between blobs at different scales. It is constructed by first defining one type of blobs, called *grey-level blobs*, at all levels of scale. The definition of this concept should be obvious from Fig. 2. Every local extremum is associated with a grey-level blob, whose extent is determined by the level curve through a specific saddle point, called *delimiting saddle point*. Formally, grey-level blobs can be defined by a water-shed analogy:

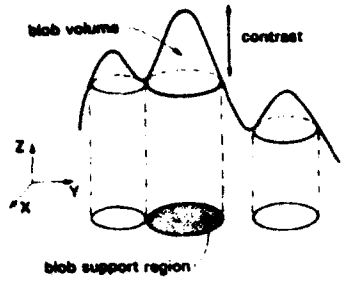


Fig. 2. Grey-level blob definition for bright blobs of a two-dimensional signal. In two dimensions a grey-level blob is generically given by a local extremum and the level curve through a specific saddle point, denoted *delimiting saddle point*.

Given any differentiable signal $f : \mathbb{R} \to \mathbb{R}$ consider any pair of maxima, A and B. They are connected by an infinite set of paths $P_{A,B}$. On each path, $p_{A,B}$, the grey-level function assumes a minimum. To reach another maximum from A, one must at least descend to the grey-level

$$z_{base}(A) = \sup_{B \in M} \sup_{p_{A,B} \in P_{A,B}} \inf_{(\xi,\eta) \in p_{A,B}} f(\xi,\eta) \ , \tag{22}$$

where M is the set of all local maxima; $z_{base}(A)$ is the grey-level value of the delimiting saddle point associated with the local maximum A. The support region $D_{supp}(A)$ of the blob is the region

$$D_{supp}(A) = \{\mathbf{r} \in \mathbb{R}^2 : \sup_{p_{A,r} \in P_{A,r}} \inf_{(\xi,\eta) \in p_{A,r}} f(\xi,\eta) \geq z_{base}(A)\} \ . \tag{23}$$

Finally, the grey-level blob associated with A is the (three-dimensional) set

$$G_{blob}(A) = \{(x,y,z) \in \mathbb{R}^2 \times \mathbb{R} : ((x,y) \in D_{supp}(A)) \wedge (z_{base}(A) \leq z \leq f(x,y))\} \ . \tag{24}$$

In general, to every grey-level blob existing at some level of scale there will correspond a similar blob both at a finer scale and a coarser scale. This notion can be made precise by applying the implicit function theorem to the extremum and saddle point associated with each blob. Grey-level blobs along an *extremum path* across scales are linked as long as neither the extremum point nor its delimiting saddle point is involved in any bifurcation, that is, as long as the Hessian remains non-degenerate. The resulting (four-dimensional) objects, which have extent in both space (x,y), grey-level z, and scale t, are called *scale-space blobs* (see [10, 11, 14] for a detailed description). From the bifurcations between these objects, which can be of four types; annihilation, merge, split, and creation, a tree-like data structure can be constructed with the scale-space blobs as primitives and the bifurcations as arcs between them.

Earlier work has shown that this representation can be used for extracting blob-like image structures from grey-level images without any prior information

about the contents of the image. A significance measure is postulated as the *(four-dimensional) volume the scale-space blobs occupy in scale-space*,

$$S_{vol,norm}(r) = \int_{t=t'_{min}}^{t'_{max}} V_{trans}(G_{volume}(r(t));\ t)\ d(\tau_{eff}(t)) \ , \qquad (25)$$

however normalized in order to enable uniform treatment of structures at different scales. $V_{trans}(G_{volume}(r(t));\ t)$ denotes a transformed volume of a grey-level blob along an extremum path r delimited by two scale values t'_{min} and t'_{max}, while $\tau_{eff} : \mathbb{R} \to \mathbb{R}$ is a transformation mapping the ordinary scale parameter into a transformed scale parameter called effective scale. For continuous signals τ_{eff} is given by $\tau_{eff}(t) = C_1 + C_2 \log t$ for some constants, C_1 and C_2 (see [12]).

Since the scale-space primal sketch is defined solely in terms of local extrema, level curves through saddle points, and bifurcations between critical points, it inherits the invariance properties listed in Sect. 2.2. This means that the topological relations in the tree-like data structure are preserved under *translations*, *rotations*, and (uniform) *rescalings* in space as well as *affine intensity transformations*. The relative ranking on significance obeys the following properties:

Invariance with respect to translations and rotations is trivial, since the scale-space representation and volumes are invariant to such operations. Concerning affine intensity transformations, it is obvious that the grey-level blob volumes are insensitive to the grey-level offset. Invariance with respect to linear stretching is achieved by dividing the measured grey-level volumes by the variation level of the input image in the transformation function V_{trans}. Because of the invariance of the scale-space primal sketch with respect to coordinate rescaling, the only way an extremum path is affected by this operation is by moving it so that the scale values t'_{min} and t'_{max} are multiplied by a constant factor. Clearly the logarithmic measure τ_{eff} is invariant to this, since it corresponds to a translation of the integration domain, which affects all scale-space blobs in the same way. The intention with the transformation function V_{trans} is that the integrand should also be well-behaved under this operation.

The scale-space primal sketch satisfies the following properties, which are essential for a low-level image representation: (i) it is based on the underlying *topology*, since it is defined from families of level curves; (ii) it is *hierarchical* in the sense that the primitives are related through a tree-like data structure, and there is a natural ranking of events in order of significance; (iii) it is *local* in the sense that the primitives of the representation have finite support and only influence their nearest neighbours (in fact, it can be used for delimiting regions in space (and intervals in scale) for further processing); (iv) it is *continuous* in the sense that the topology of the overall representation is preserved as long as the relations in the underlying images remain the same; (v) it is *invariant* to transformations such as rotations and translations in space; (vi) it is *compatible with rescalings* of both the spatial coordinates and the grey-level intensity. For discrete signals, the rotations must be multiples of 90 degrees, the translations must be pixel-wise, and the spatial rescaling factor must be an integer in order to preserve the invariance.

5 Summary and Discussion

It has been shown how the formulation of feature detectors in terms of differential singularities makes it theoretically simple to analyze their behaviour over scales. Even though further work may be needed before implementing the drift velocity estimates derived for the different feature detectors, these expressions completely describe the theoretical evolution properties of such non-linear combinations defined in scale-space. The discretization of the scheme in Fig. 1 is treated in [13, 14], where experimental results are also presented.

References

1. Babaud, J., Witkin, A.P., Baudin, M., Duda, R.Q. (1986). Uniqueness of the Gaussian kernel for scale-space filtering, IEEE Trans. Patt. Anal. Mach. Intell. 8 (1), pp. 26–33.
2. Blostein D., Ahuja N. (1987). Representation and three-dimensional interpretation of image texture: An integrated approach. In: Proc. 1st Int. Conf. Comp. Vision, London, England, Jun. 8-11, pp. 444–449.
3. Brunnström K., Lindeberg T.P, Eklundh J.-O. (1992). Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch. In: Proc. 2nd Eur. Conf. Comp. Vision, Santa Margherita Ligure, Italy, pp. 701–709.
4. Canny J. (1986). A computational approach to edge detection, IEEE Trans. Patt. Anal. Machine Intell. 8 (6), pp. 679–698.
5. Florack L.M.J., ter Haar Romeny B.M., Koenderink J.J., Viergever M.A. (1991). General intensity transformations and second order invariants, Proc. 7th Scand. Conf. Image Analysis, Aalborg, Denmark, Aug 13-16, pp. 338–345.
6. Kitchen, L., Rosenfeld, R., (1982). Gray-level corner detection, Patt. Recogn. Lett. 1 (2), pp. 95–102.
7. Koenderink J.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
8. Koenderink J.J., van Doorn A.J. (1987). Representation of local geometry in the visual system, Biol. Cybern. 55, pp. 367–375.
9. Koenderink J.J., Richards W. (1988). Two-dimensional curvature operators, J. Opt. Soc. Am. 5 (7), pp. 1136–1141.
10. Lindeberg T.P., Eklundh J.-O. (1992). The scale-space primal sketch: Construction and experiments, Image and Vision Comp., 10, pp. 3–18.
11. Lindeberg T.P. (1992). Scale-space behaviour of local extrema and blobs, J. Math. Imaging Vision 1, pp. 65–99.
12. Lindeberg T.P. (1993). Effective scale: A natural unit for measuring scale-space lifetime, IEEE Trans. Pattern Anal. Machine Intell., in press.
13. Lindeberg, T.P. (1993). Discrete derivative approximations with scale-space properties, J. Math. Imaging and Vision, in press.
14. Lindeberg, T.P. (1993). Scale-Space Theory in Early Vision, Kluwer Academic Publishers, Boston, to appear.
15. Marr D. (1982). Vision, Freeman, San Francisco.
16. Witkin, A.P. (1983). Scale-space filtering, In: Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany, Aug. 8-12, pp. 1019–1022.

Exploring the Shape Manifold: the Role of Conservation Laws

Benjamin B. Kimia[1], *Allen R. Tannenbaum*[2], *and Steven W. Zucker*[3] *

[1] Laboratory for Man/Machine Systems, Brown University, Providence RI, 02912, USA, kimia@lems.brown.edu
[2] Department of Electrical Engineering, University of Technion, Israel, and University of Minnesota, USA.
[3] McGill Research Center for Intelligent Machines, Montreal, Canada

Abstract. A general theory of shape is developed from basic principles. These principles are organized around two intuitions: first, if a boundary changes only slightly, then its shape should change only slightly. This leads to a proposal of an operational theory of shape based on incremental contour deformations. The second intuition is that not all contours are shapes, but rather only those that can enclose "physical" material. A novel theory of contour deformation is derived from these principles, based on abstract conservation principles and the Hamilton-Jacobi theory. The result is a characterization of the computational elements of shape: deformations, parts, bends, and seeds, which show where to place the components of a shape. The theory unifies many of the diverse aspects of shapes, and leads to a space of shapes (the reaction/diffusion space), which places shapes within a neighbourhood of "similar" ones. Such similarity relationships underlie descriptions suitable for recognition.

Keywords: shape description, object recognition, shape evolution, shape deformation, parts, reaction-diffusion, entropy, conservation, viscosity solution, topology of shapes.

1 Introduction

While there is a sense in which the meaning of shape is effortlessly and intuitively understood, a formal definition of it has been elusive: there is currently no generally accepted definition of shape in either computational vision or psychology. This gap in understanding is important, because shape may be considered as the bottleneck between early visual processes operating on edges, texture, colour, shading, etc., and higher level processes acting on representations of objects. A theory of shape sufficiently powerful to provide a language for describing shapes is thus needed. It follows that such a theory must be robust to variations within scenes, for example, due to small changes in viewpoint, to the changing

* This research was supported by grants from AFSOR, ARO, MRC, NSERC, and NSF. We thank David Mumford for an early reference to the work of Grayson, and Allan Dobbins and Lee Iverson for discussions and technical help.

appearance of objects due to local motion and emergent occlusions, as well as to variations within objects, for example, due to flexibility, growth, and inflation. To meet these needs, a formal framework for our theory is derived from a mathematical model of deformations. But the results are not just mathematical: in an attempt to capture the intuitions underlying shape, a series of natural principles are postulated to which any such theory should be subject. This complements the examples presented throughout the paper, which exhibit the numerical stability of this new technique.

2 The Need for a Novel Geometry of Shape

Objects come in all forms. As they deform and grow incrementally, their shape does not change drastically. For example, one's perception of a tree is not drastically altered each day as it grows, nor as a flock of birds rests on it. That the primary perception is one of an object with modifications is so intuitive that to mention it might appear redundant. Yet, it is essential that any general computational representation of shape must behave similarly, so that, for example, an industrial tool that is slightly bent or chipped will be described as "a tool that was bent or chipped". Analogously, when objects deform with motion, growth, erosion, etc., the percept is only slightly modified. There seems to be great stability with regard to such changes [20, 19, 18].

Unfortunately, standard geometries do not satisfactorily address these aspects of shape for the purposes of object recognition. This points to the need for a language that makes the morphogenesis of shape explicit. A second point concerns the treatment of singularities. Attneave showed that the most salient portions of a shape are corners and high curvature points [3]. However, singularities do occur in nature, and they play a different role than their smoothed versions [21]. They must have an explicit place in a theory of shape.

2.1 Preview of Results

A preview of results is presented to close the introductory portion of the paper and to provide a concrete focus for the ensuing theoretical discussion. First, the notion of deformation and how it leads to robust descriptions of parts is illustrated. Figure 1 contains four images of pears, presented by Richards *et al.* in [23], and which were intended as gross modifications of an object category (pear). The original shapes are across the top, and each column contains samples from a continuous sequence in which the bounding contour has evolved according to the deformation rules (to be developed in Sect. 3). The samples were chosen to illustrate how the deformation process eliminates the noise (first row) to reveal the fundamental part structure for the pear (second row). This structure is a pair of lobes, with the most significant one on the bottom. The relevant shocks in this case signal the part structure, and correspond to the orientation discontinuities that develop on the evolving contour in ordered pairs. Note how the lobe structure and the dominant lobe (bottom row) are comparable for each

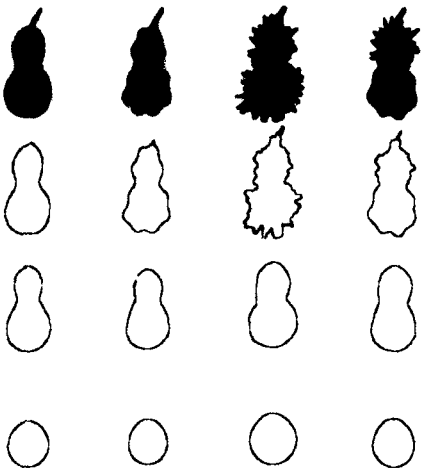


Fig. 1. An illustration of how our deformation approach to shape leads to natural descriptions despite large quantities of noise and texture. The original shapes are across the top, in black. Each column contains samples from a continuous sequence in which the bounding contour has evolved according to our deformation rules. The samples were chosen to illustrate how the deformation process eliminates the noise (first row) to reveal the fundamental part structure for the pear (second row). This structure is a pair of lobes, with the most significant one on the bottom. Note how the lobe structure and the dominant lobe (bottom row) are comparable for each of these different pair images, even though the noise and texture were so prominent.

of these different pair images. The continuous space of shapes which supports such descriptions is called the *entropy scale-space*.

A second example shown in Fig. 2 illustrates the notion of hierarchy in more detail. An image of a doll was chosen to show how the different "parts" emerge according to one's natural intuitions about significance. Observe that the "feet" partition from the "legs"(via second-order shocks, to be discussed in Sect. 5) between frames 3 and 4, and the "hands" from the "arms" between frames 2 and 3. Following these second-order shocks, first-order shocks develop as the "arms" are "absorbed" into the chest. Running this process in the other direction would illustrate how the arms "protrude" from the chest. Note, in addition, how hands and feet are less significant than limbs, which are in turn less significant than the torso. This example also illustrates that several different types of shocks arise within the system, with first-order shocks signalling deformations, second-order shocks signalling part connections, third-order shocks signalling bends, and fourth-order shocks signalling part centres. Note that occlusion will not affect decomposition into parts, a desirable feature for recognition.

3 Shape from Deformations

To begin the development of our framework for shape, note that, since any recognition strategy requires a notion of similarity between shapes, or of a "neighbour-

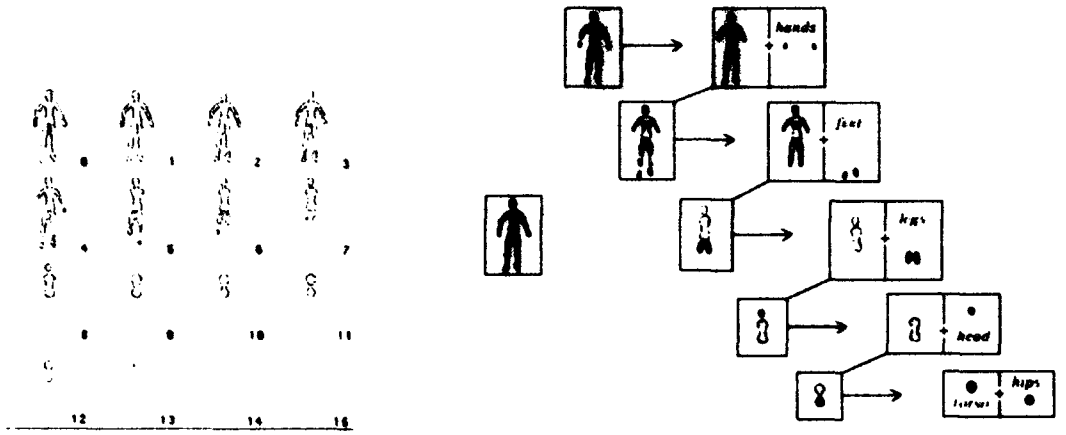


Fig. 2. (left) The evolution of shocks leads to parts, protrusions, and bends. This figure shows the development of an image of a doll (National Research Council of Canada Laser Range Image Library CNRC9077 Cat No 422; 128X128). The contour shown in each box corresponds to a boundary evolution at the time shown in the lower right hand corder of the box. (right) The hierarchical decomposition of a doll into parts. Selected frames were organised into a hierarchy according to the principle that the significance of a part is directly proportional to its survival duration. These notions are described precisely in [17].

hood" around each shape, the shape of an object should be intimately interconnected to "nearby" shapes. To illustrate this, consider the shapes in Fig. 3; these are readily seen as similar, and as variations within the category of "peanuts". Such variations among shape are captured via deformations, and deformations are characterised in a differential geometry. Our approach is to apply arbitrary deformations to shapes and, through incremental change, to observe the emerging organization of the space of shapes. In the first subsection, it is shown that the space of local arbitrary deformations is qualitatively spanned by two simple deformations: constant deformations and curvature deformations. In the next subsection, a distinction is made between the evolution of *contours* and the evolution of *shapes*. The entire development is guided by several principles that, in our view, are fundamental and self-evident. More formally, they are proposed to ensure that evolving curves remain "valid" shapes, that is, possible projections of three-dimensional objects onto a two-dimensional image. Then, it is shown that the contrasting and complementary properties of shape are captured in the interaction between constant and curvature deformations.

3.1 Shape from Deformations of Contours

Much of early vision is organized around inferring boundaries [27], from which it may be observed that

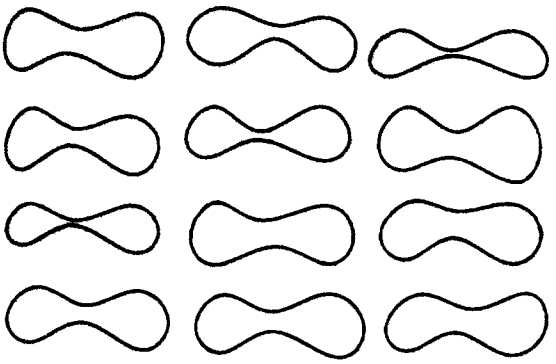


Fig. 3. These shapes seem to belong to the same group of objects. This concept of a neighbourhood of "nearby" shapes is the key to recognition.

Principle 1. *Slight changes in the boundary of an object cause only slight changes to its shape.*

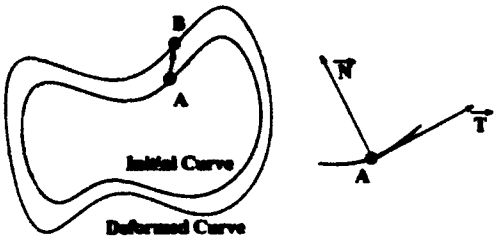


Fig. 4. The points on the initial curve A move to B to generate a new curve. The direction and magnitude of this motion is arbitrary in order to capture general deformations. However, with mild restrictions appropriate to shape, one can classify this deformation as a sum of *constant deformation* and *curvature deformation* along the normal.

Consider a shape represented by the curve $C_0(s) = (x_0(s), y_0(s))$ undergoing a deformation, where s is the parameter along the curve (not necessarily the arclength), x_0 and y_0 are the Cartesian coordinates and the subscript $_0$ denotes the initial curve prior to deformation. Now, let each point of this curve move by some arbitrary amount in some arbitrary direction; see Fig. 4. This evolution is then described as

$$\begin{cases} \dfrac{\partial C}{\partial t} = \alpha(s,t)\mathbf{T} + \beta(s,t)\mathbf{N} \ , \\ C(s,0) = C_0(s) \ , \end{cases}$$

where $\mathbf{T}$ is the tangent, $\mathbf{N}$ is the outward normal, s is again the parametrization, t is the time duration (magnitude) of the deformation, and α, β are arbitrary

functions. This, by a reassignment (i.e. reparameterization) of points, can be reduced [13, 14] to

$$\begin{cases} \dfrac{\partial \mathcal{C}}{\partial t} = \beta(s,t)\mathbf{N} \ , \\ \mathcal{C}(s,0) = \mathcal{C}_0(s) \ , \end{cases}$$

where β is again arbitrary, but not necessarily the same as that of the previous equation. Now, concentrate on intrinsic deformations that depend only on the *local geometry* of the curve at that point, namely those dependent on the curvature [8],

$$\begin{cases} \dfrac{\partial \mathcal{C}}{\partial t} = \beta(\kappa(s,t))\mathbf{N} \ , \\ \mathcal{C}(s,0) = \mathcal{C}_0(s) \ , \end{cases} \tag{1}$$

where κ is the curvature.

Since the deformations are intended to bring out the relationships among shapes, it is reasonable to require that the process relating shape $\mathbf{S}_1$ to shape $\mathbf{S}_2$ is independent of *when* it is applied to $\mathbf{S}_1$. For example, the way an ellipse relates to a circle in the space of deformations should not be dependent on the time of the deformation, but rather on the amount and nature of the deformation itself. Hence our second principle concerns time-invariance and it is proposed that:

Principle 2. *The class of contour deformations necessary to articulate shape consists of those deformations that do not depend on the time the deformation is applied.*

Then from (1)

$$\begin{cases} \dfrac{\partial \mathcal{C}}{\partial t} = \beta(\kappa(s))\mathbf{N} \ , \\ \mathcal{C}(s,0) = \mathcal{C}_0(s) \ . \end{cases} \tag{2}$$

To obtain the generic character of (2), assume $\beta(.)$ is analytic, that is, that it admits a Taylor series expansion, and consider the following "first-order approximation" of $\beta(\kappa)$:

$$\begin{cases} \dfrac{\partial \mathcal{C}}{\partial t} = (\beta_0 - \beta_1\kappa)\mathbf{N} \ , \\ \mathcal{C}(s,0) = \mathcal{C}_0(s) \ . \end{cases} \tag{3}$$

The remaining terms in a Taylor expansion of an analytic β involving higher orders of κ qualitatively resemble κ for the purposes of shape [14].

Equation (3) contains two terms. The first term describes a deformation that is a constant motion along the normal, or *constant deformation*. The second term describes a deformation that is proportional to the curvature along the normal, or *curvature deformation*. To summarize the discussion of deformations, it has been shown that:

Result 3. Arbitrary local deformations of a curve in an arbitrary direction are qualitatively captured by a linear combination of two *basis* deformations, namely, the *constant deformation* and the *curvature deformation*, of the curve along its normal.

Such deformations will be fundamental to our framework, and will provide the basis for forming a topology of shape.

3.2 Shape Deformation vs. Contour Deformation

The next principles relate to the observation that not all contours are valid shapes. Recall that, informally, shape derives from the projection of three-dimensional objects, or volumes of material, onto two dimensions. The basic constraint is thus that *for contours to represent shapes they must be (the projection of) boundaries which could enclose "material"*. This notion also seems to hold psychophysically [9]. To restrict curve evolution to shape evolution, consider the situation in which two remote points of the boundary touch each other. This would occur in the process of pinching a ball of clay, for example. At the point when the two extremal points come together, the object falls apart into two pieces.

Principle 4. *If, during the process of deformation, distinct points of the boundary touch, then the evolved shape, or its background, splits into two subshapes.*

It follows, of course, that once a shape has split it cannot be joined together again by continuing the process of deformation:

Principle 5. *During the process of deformation the boundary of the shape must not cross over itself.*

This principle had an earlier expression in the "grass-fire" transformation of Blum [4], who observed that grass could not burn twice.

It is obvious that open curves cannot contain material. Therefore,

Principle 6. *The boundary of a shape must remain closed during the process of deformation.*

Next consider the singularities of the contour, such as corners and cusps. Since objects often have sharp edges, bends, etc., these project to corners and cusps in the contour. In fact, these are among the salient points of a shape and deserve an explicit representation. However, there cannot be infinitely many such singularities, or for that matter extrema in curvature, because physical objects are composed of materials with a finite grain size and are observed by devices with finite resolution limits. This implies a finite total undulation in the two-dimensional shape, and such total variation may be measured by total absolute curvature as defined by

$$\bar{\kappa}(t) = \int_0^{2\pi} |\kappa(s,t)| g(s,t)\, ds \ ,$$

where $g(s,t)$ is the length metric along the curve:

$$g(s,t) = \left|\frac{\partial \mathcal{C}}{\partial t}\right| = [x_s^2 + y_s^2]^{1/2}$$

Note that this definition allows for the representation of curves with tangent discontinuities, for example, a square, for the infinite curvature can be countered by infinitesimal speed [14]. Therefore,

Principle 7. *During the process of deformation the boundary of the shape must remain of finite total absolute curvature.*

Notice that closed curves evolving by (3) must remain closed (as long as the classical solution exists). Moreover, from the maximum principle for parabolic equations, one can show (see for example, [1]), that

Theorem 8. *Simple closed curves evolving by (3) remain simple and closed (as long as the classical solution exists).*

The final principle relates the deformation process to the change in similarity.

Principle 9. *The deformation of shape is required to preserve similarity.*

4 Conservation and Shape: The Role of Shocks and Entropy

It is most intriguing that an arbitrary combination of constant and curvature deformations satisfies a *conservation law with viscosity.* The relevance of conservation laws to shape is subtle. First, there is an intuitive connection in which our deformations leave certain aspects of the shape conserved, for example, contour orientation. The second connection is technical: the conservation laws allow our deformation models to continue beyond the formation of singularities. This is particularly important for shapes with discontinuities, such as a square which has four corners, where the normal is undefined.

4.1 Conservation of Orientation

How does an infinitesimal piece of a curve change its orientation when viewed externally? To motivate our approach, recall that, when matter flows through a small section of pipe, matter is conserved in the sense that the amount of matter flowing into this piece of pipe is precisely equal to that which flows out plus that which stays in. Similarly then, consider a small piece of the external x-axis coordinate, the interval $(x, x + \Delta x)$. This infinitesimal interval can be regarded as a small section of a pipe through which "orientation flows". It is shown that orientation is not annihilated or created by this flow for curves evolving according to (3) with $\beta_1 = 0$, that is, orientation is conserved. When $\beta_1 \neq 0$, viscosity is introduced into the system:

Theorem 10. *The orientation of a curve deformed by constant deformation satisfies*

$$\frac{\partial \theta}{\partial t} + \mathcal{H}(\theta)_x = 0 \ , \tag{4}$$

where θ is the orientation of the curve in some Cartesian coordinate frame, namely, the angle the curve tangent makes with the x axis, $\mathcal{H}(\theta) = -\cos(\theta)$ is the flux of orientation flow, $-\pi/2 < \theta \leq \pi/2$; clearly a hyperbolic conservation law for orientation θ.

Intuitively, each coordinate frame's horizontal axis can be viewed as a pipe through which "orientation" flows. The conservation law asserts that in this process orientation does not annihilate or regenerate. Rather, it flows from one section to another, governed by a flux $\mathcal{H}(\theta) = \cos(\theta)$. Adding curvature motion, on the other hand, adds viscosity to the system:

Theorem 11. *The orientation of a curve deformed by a combination of* constant motion *and* curvature motion *satisfies*

$$\theta_t + \beta_0[\mathcal{H}(\theta)]_x = \beta_1 \cos^2(\theta)\theta_{xx} \ , \tag{5}$$

where $\mathcal{H}(\theta) = -\cos(\theta)$, namely a viscous hyperbolic conservation law for orientation θ.

This kind of viscosity or "diffusion" changes the character of the deformation. Whereas with no viscosity, deformations conserved the local orientation identity of each piece, with viscosity the local orientation of each piece is "blended" with its neighbouring points. This blending, in its pure form, is equivalent to a form of Gaussian smoothing of the boundary coordinates. Informally then, one view of the constant deformation and curvature deformation trade-off is that of area versus length or region versus boundary.

4.2 The Formation of Shocks

Other than the intuitive appeal of these conservation laws, their more significant role is that the original process of deformation (2), a local differential model, can now be extended to handle singularities using an "integral" form of the conservation law and the notion of "weak" solutions of partial differential equations.

The key lies in Principle 9 and the introduction of a notion of *entropy* for shape. Since the conservation law of orientation is valid beyond the point when singularities form, the principle of conservation can be postulated as the fundamental principle underlying the deformation of shape, and it can be used to guide shape evolution beyond the formation of "corners". Informally, the entropy condition, which forces characteristics to lean into a shock, translates into the condition of removing dashed curves for shape. These dashed lines are portions of the curve which cross over each other. A physical analogy may help: in gas dynamics, when particles of unit velocity and stationary ones collide at a shock, they "reach an agreement", namely, the formation of a shock moving with a

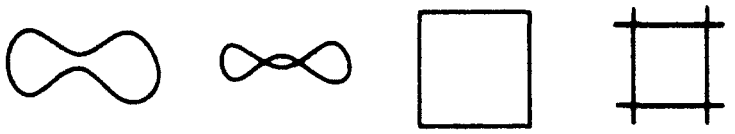


Fig. 5. Curve evolution is not shape evolution. Note that in the process of deformation of the curve, local portions of the boundary may cross over each other, like the corner of the square. Similarly, remote portions of the boundary may cross over each other. Since shapes are curves that are filled with material as required by Principles 5 and 6, the local curve deformation does not always lead to shape deformation. To resolve this dilemma, the interior must be represented explicitly.

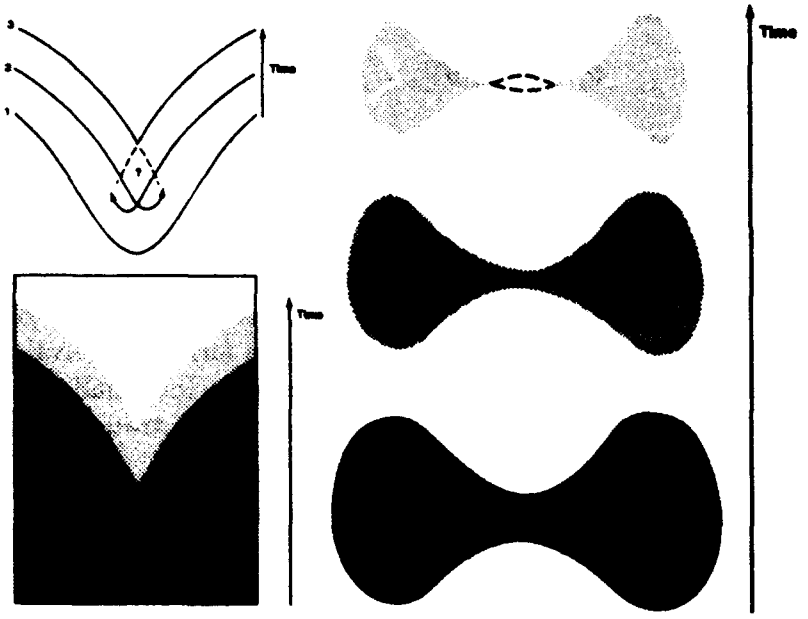


Fig. 6. Nonlinear processes can transform initially smooth functions to functions with singularities. (left) shows a curve with a negative curvature extremum which, when evolved by constant motion along the normal, leads to a singularity. This evolution can be based entirely on boundary information until the singularity arises. However, at this point the entropy condition is required to further control evolution, so that the curve does not cross over itself and the swallowtail configuration can be properly handled. The entropy condition is region-based, and controls how interior information interacts with the boundary. It plays another key role in controlling topological evolution, by globally managing the splitting of a single boundary into two closed boundaries (right). In both cases the entropy condition dictates that the solution does not include the "dashed" portions of the contour – these annihilate into the shock.

"compromise" velocity which depends on the form of the flux. Certain particles must annihilate into a shock. For shape, the dashed lines would be present if each portion were to evolve independently by deforming along the normal. However, given the conservation law (4), during the collision a similar "agreement" must be reached between the orientation of "particles", namely, the crossed-over portions — the dashed lines of Fig. 6 — must annihilate into a shock. Since the dashed segments are points which the boundary has crossed already, the following definition of *entropy* for shape is appropriate:

Definition 12 Entropy condition. In the process of inward deformation, once a point is dislodged from a shape, it remains disjoint from it forever. Similarly, in the process of outward deformation, once a point becomes part of a shape, it remains part of it forever.

This condition is reminiscent of the "grass fire" algorithms in vision, where once an area is burnt, it cannot be burned again; see also Sethian's analogous entropy condition for flame propagation [24].

Whereas the conservation law may be postulated in the above fashion to be a beginning for our approach, it can also be shown that it is necessarily the way to continue deformations past the point of singularity formation, based on previously postulated principles. The key connection here is one between the entropy-satisfying solution of hyperbolic conservation laws and their *viscosity solutions*. The idea is to introduce an infinitesimal amount of viscosity to the system and reduce it to zero. It is reasonable to require that the limit of such a solution be the solution of a system in the absence of viscosity [12, 10, 6].

4.3 Embedding Curve Evolution in a Higher-dimensional Space

The conservation law formulation resolves the first of the two problems depicted in Fig. 5, that is, the local collision of the boundary and the consequent formation of singularities. However, as the peanut shape of Fig. 5 evolves in time, remote portions of the boundary collide and pass over each other. This collision does not manifest itself in either the local curve deformation model of (2) or the more general conservation model of (5). The missing ingredient is a notion of "interior". Observe that a comprehensive understanding of shape involves the notion of *both* its boundary *and* its interior. To allow for this extra "dimension" of information, consider an evolution in a higher-dimensional space, for example, the evolution of a two-dimensional surface in a three-dimensional space constrained to embed the original problem.

In [16] it is proved that:

Result 13. Consider a surface $z = \phi(x, y, t)$ evolving according to the Hamilton-Jacobi equation

$$\phi_t(x, y, t) + \beta_0 \mathcal{H}_0(\phi_x, \phi_y, t) = \beta_1 \mathcal{H}_1(\phi_{xx}, \phi_{yy}, \phi_x, \phi_y, t) \ , \qquad (6)$$

with the initial condition $z = \phi(x, y, 0) = \phi_0(x, y)$. If the explicit representation of $\phi_0(x, y) = 0$ is denoted by $\mathcal{C}_0(s) = (x_0(s), y_0(s))$, then the (x, y, t) satisfying $\phi(x, y, t) = 0$ also satisfy evolution (3).

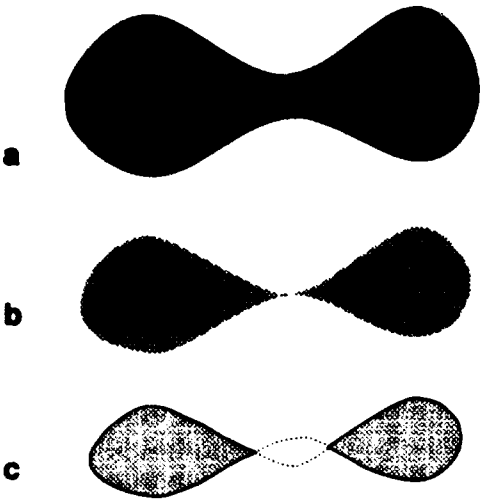


Fig. 7. This figure depicts the case when two points of a shape (a) that are distant along its boundary come together during an arbitrary deformation (b). How should the deformation proceed beyond this point? A pointwise deformation along the normal would produce the dashed-lines, which clearly violate Principle 5 since they do not correspond to an actual object.

Thus, by invoking the entropy condition as presented in Definition 12, when portions of the shape's boundary collide and pass over each other, the shape is segmented into two disjoint subshapes, each evolving separately, according to Principle 5. In this way, topologically connected shapes composed of two components, like that in Fig. 7(a) and those with two disjoint components, like that in Fig. 7(c), become neighbours in a deformation process and therefore are similar. To summarize:

Result 14. Solutions of (6) satisfying the entropy condition in Definition 12 are the proper "physical" solutions as they are also obtained by the viscosity method.

5 The Reaction-Diffusion Space and Formation of Shocks

In the previous section shocks were shown to form in the course of evolution of shapes. In this section, these shocks are classified and are shown to lead to our proposal for the computational elements of shape: parts, protrusions, and bends. These shocks occur for various combinations of constant deformation and curvature deformation, or reaction and diffusion. The space generated by these combinations and by time is thus referred to as the *reaction-diffusion space*. It is in the context of this space that shocks will be related to shape.

5.1 The Reaction-Diffusion Space

Definition 15. The representation of a shape $\mathbf{S}$ in all possible time and all possible ratios β_0/β_1 ($\beta_1 < 0$) is called the *reaction-diffusion space* for that shape.

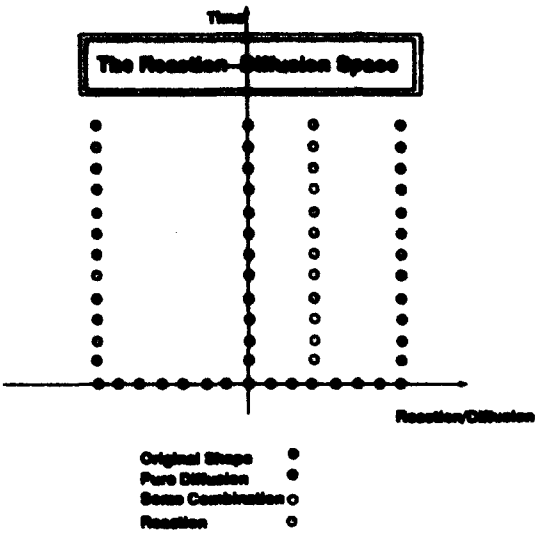


Fig. 8. The reaction-diffusion space represents a range of deformations: diffusion corresponds to deformation by curvature motion alone, while reaction corresponds to deformation by constant motion (both in and out) along tne normal. While diffusion represents the "y-axis", reaction represents the extreme verticals. Intermediate vertical line represent combinations of the two deformations. Time is the amount of deformation in all cases.

In other words, the reaction-diffusion space for a shape $\mathbf{S}$ is the set of all shapes $\mathbf{S}'$ generated by

$$\mathcal{RD}_{(\alpha,t)}\,\mathbf{S} \longmapsto \mathbf{S}' \ , \qquad (\alpha \in (-\infty,\infty), t \in [0,\infty)) \ ,$$

where $\mathcal{RD}$ is the deformation with α as the ratio of constant motion to curvature motion magnitudes and t is time; see Fig. 8. This space therefore spans all combinations of reaction and diffusion, and time. Note that $\beta_1 > 0$ represents the heat equation running backward in time and $\beta_1 = 0$ represents the case of no diffusion; a nonrealistic situation. Alternatively, one may view the reaction-diffusion deformation as an operator acting on the space of all shapes $\mathbb{S}$, generating a reaction-diffusion space, a collection of new shapes, for each shape $\mathbf{S} \in \mathbb{S}$:

$$\begin{cases} \mathcal{RD} : \mathbb{S} \longrightarrow \mathbb{S} \ , \\ \mathcal{RD}_{(\alpha,t)}(\mathbf{S}) = \mathbf{S}', \qquad \mathbf{S} \in \mathbb{S} \ . \end{cases}$$

In the reaction-diffusion space, the shock formation process is governed differently as the ratio of reaction to diffusion is altered. Also, the time of formation of a shock is related to its significance. This two-dimensional space affords a much richer analysis than the representation of a single curve in isolation. Within the reaction-diffusion space, the following shocks may arise [14].

5.2 First-order Shocks

Consider the shape in Fig. 9 which is formed by pushing a portion of a circle outwards. It would not be uncommon to describe this shape as a "circle with a protrusion". Now, consider a constant motion type of deformation on this shape. Adhering closely to the terminology of classical conservation laws, let us preserve the term *shock* and define:

Definition 16. When in the process of deformation and orientation flow, curvature builds up to create an orientation discontinuity, then a *first-order shock* has formed.

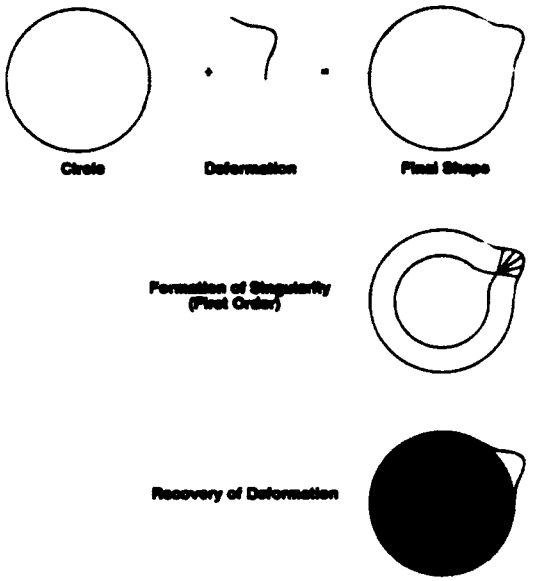


Fig. 9. The shape on the right is perceived as a circle with a deformation. While a number of other interpretations are possible, this interpretation seems to be favoured naturally.

Thus, a *first-order shock* is a discontinuity in orientation of the boundary of a shape, which often arises from curvature extrema of the boundary:

Theorem 17. *In the process of evolution by constant motion, each local curvature extremum leads to a first-order shock, provided that only this local portion of the curve evolves.*

First-order shocks are associated with protrusions (indentations) in the absence of other shocks. They arise because curvature accumulates most rapidly at extrema. Note that several smaller protrusions may merge to form one at a larger "scale" as in Fig. 10.

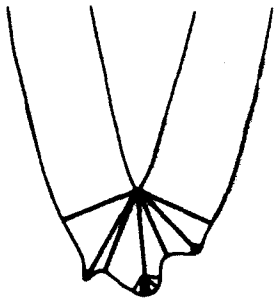


Fig. 10. This figure illustrates the hierarchy of first-order shocks, depicting protrusion in multiple scales.

5.3 Second-order Shocks

A second kind of shock forms, not due to curvature build-up as in the first type of shock, but due to a collision of boundaries. Consider the shape in Fig. 7. As the shape (a) evolves in time due to a *constant deformation*, portions of the boundary collide and give rise to two cusps (b). These cusps are discontinuities, not in tangent, but in curvature. These are referred to as *second-order shocks*. Note the change of connectivity at this instant. Beyond this instant, portions of the boundaries cross each other (the dashed lines). The role of entropy in this case is to remove portions of the boundary that have reached a previously visited point (c). Formally,

Definition 18. When in the process of deformation two distinct non-neighbouring boundary points join and not all the other neighbouring boundary points have collapsed together, a *second-order shock* is formed.

Thus, a second-order shock is a dicontinuity of curvature, but not of orientation. The second-order shocks define *parts* of a shape. This notion of *parts* is different than that proposed in [11], where parts are defined by negative minima of curvature. Our parts are more intuitive; see, for example, Fig. 11. These ideas have been extended in [25, 26] to include neck-based and limb-based parts.

5.4 Third-order Shocks

A third type of shock point is generated when distinct boundary points are brought together as in second-order shocks, but, unlike the second-order shock, the neighbouring boundary points on each side have also joined with other distant boundary points. Formally,

Definition 19. When in the process of deformation two distinct non-neighbouring boundary points join, so that neighbouring boundaries of each point also collapse together, a *third-order shock* is formed.

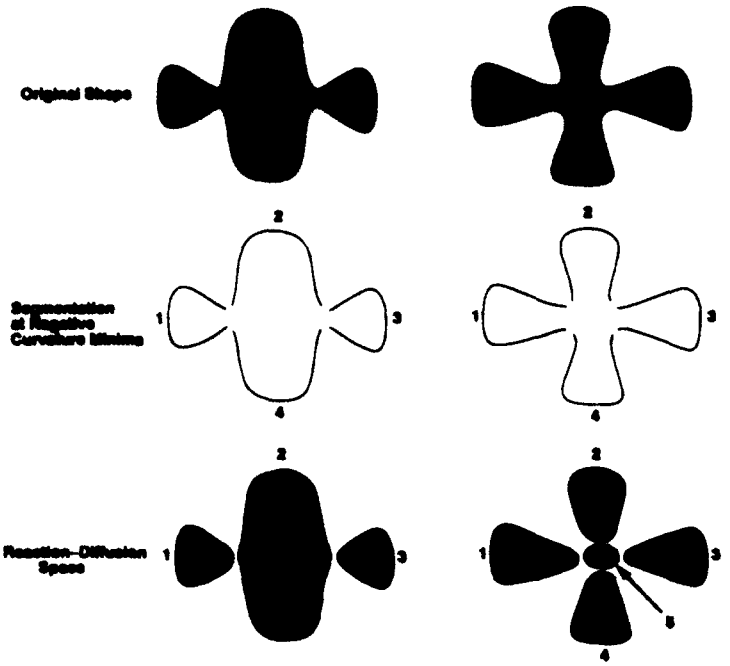


Fig. 11. Partitioning of a two-dimensional shape requires not only boundary, but also region information. The top row shows two shapes and the unnatural part structure implied by [11]. Our theory leads to much more natural descriptions (bottom row).

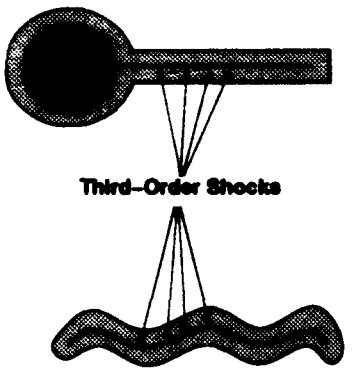


Fig. 12. The snake shape forms third-order shocks when distant points of the boundary come together not in isolation, but rather in conjunction with neighbours. Third-order shocks indicate the "bending" of an object. The interpretation of the snake therefore is as a "bent stick". Note that the partitioning of the "snake" at the negative curvature minima would give rise to five (unnatural) parts, as indicated by a pure boundary-based approach.

Thus, a curve of third-order shocks represents the median of two parallel curve segments of the boundary. As defined above, third-order shocks cannot possibly change the topological connectivity of the shape. Rather, they indicate a symmetric axis, as in the case of an ellipse. However, this axis is not composed of first-order shocks where portions of the boundary collapse into a single point. Rather, this axis is the result of a region collapsing into points, Fig. 12. Therefore, the locus of these points indicates a *bending* of an extended region, rather than a protrusion of the boundary.

5.5 Fourth-order Shocks

In the process of inward evolution of a shape, regions shrink and form shocks. In time, remaining regions finally shrink to a point and disappear due to the entropy condition. All parts of a shape must eventually annihilate to a point, since the shape may be entirely embedded inside some circle of radius R which will, in R/β_0 units of time, disappear.

Definition 20. When in the process of deformation a closed boundary collapses to a single point, a *fourth-order shock* is formed.

Thus, fourth-order shocks are the *seeds* for parts of shape.

5.6 Examples

In this section, the reaction-diffusion space and the formation therein of shocks is illustrated. To recall, the *reaction-diffusion space* is the collection of all deformations of shape for all combinations of reaction β_0 and diffusion β_1, and time t. Since of these three variables only two are independent, time and the ratio of reaction to diffusion (β_0/β_1) are selected as the two variables which span the space. This choice is motivated by the fact that there is always some amount of diffusion present, so that *the vertical lines at the extremes of the x-axis are associated with pure reaction and the y-axis is associated with pure diffusion.* Furthermore, the case of pure diffusion is a natural seam between inward and outward reactions; see Figs. 8 and 13.

To interpret our reaction-diffusion convention, then, each vertical line at x is a deformation of the original shape with the ratio $\beta_0/\beta_1 = x$ (the absolute values of β_0 and β_1 are not relevant in that they are absorbed in time t). Note that for $\beta_1 = 0$, or pure reaction, some diffusion manifests itself in the numerical implementation, so that shapes along this line diffuse minimally. The vertical dimension of the line represents time, or the amount of deformation. This vertical axis is best depicted on a logarithmic scale and the numerals indicate the time frame of the computation from which the image was taken. However, note that there is no inherent "vertical topology" in this space and that the space is intended to generate a topology in both dimensions; this is a visual choice for representing this space.

There is an interesting connection between the reaction-diffusion and the symmetric axis transform (SAT): the shocks along the pure reaction axis form

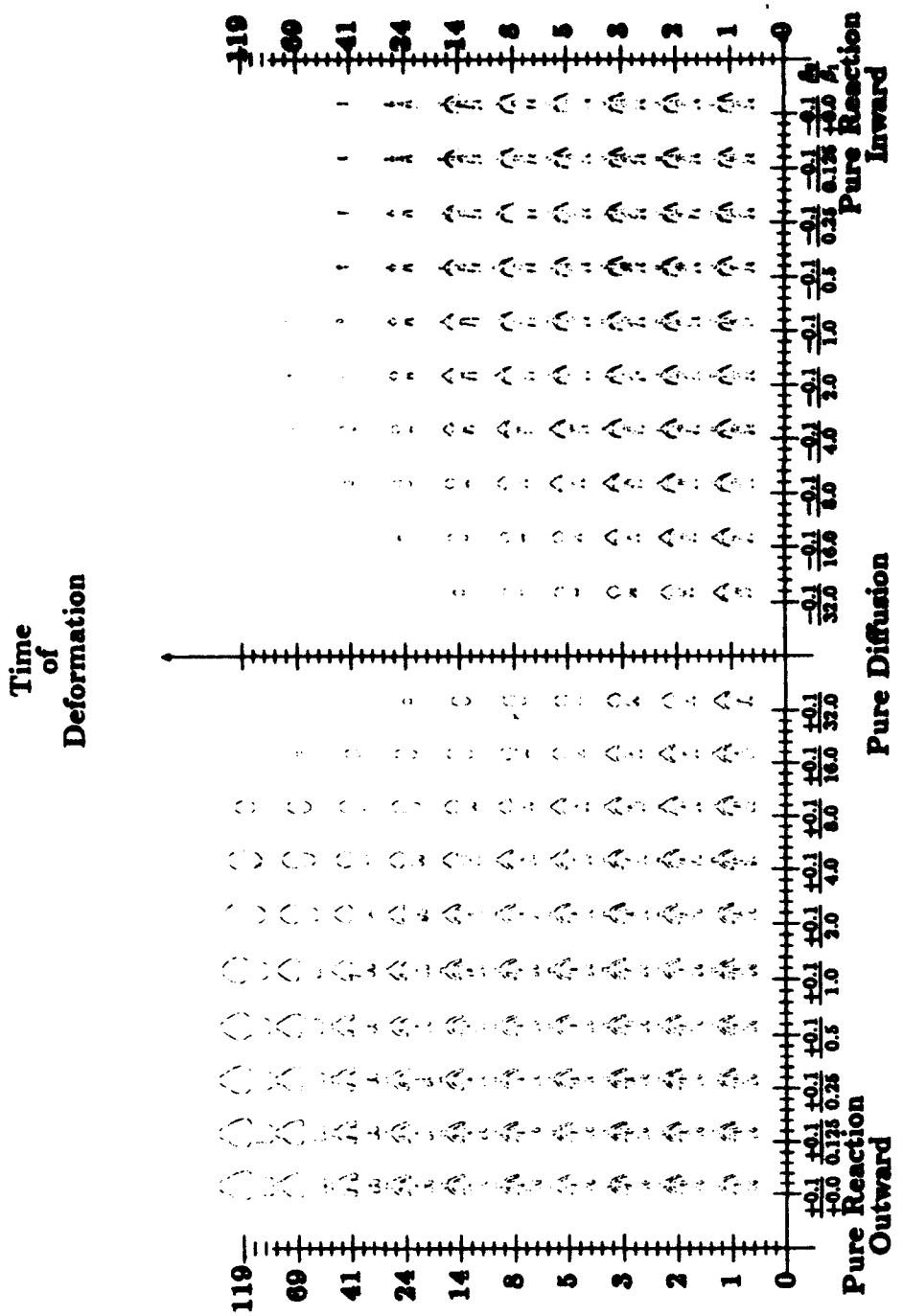


Fig. 13. This figure is an example of a reaction and diffusion involving both inward and outward deformation for the DOLL image. The DOLL image was taken from a range image collection of the National Research Council of Canada's Laser Range Image Library CNRC9077 Cat No 422. The image was thresholded and stored as a 128x128 image. The numbers on the x-axis are indicative of the two values β_0, β_1 in relation to time. Note the formation of shocks with both inward and outward reaction. Also, observe that outward reaction may be thought of as an inward reaction when the role of figure and ground is reversed.

the loci of the symmetric axis transform. Since SAT is susceptible to noise, various smoothing algorithms have been proposed [7, 22]. In our framework, however, diffusion is naturally part of the deformations. As such, not only the resulting description is a "colouration", or classification, of the symmetric axis into meaningful portions via the classification of shocks, but a sense of significance also emerges. The other interesting connection is that the reaction-diffusion space under the pure reaction axes embeds the mathematical morphology operations of erosion and dilation with a ball structuring element. Furthermore, the "shape from deformation" framework also embeds mathematical morphology operations with all convex structuring elements [2]; also see [5].

Shape representation is perhaps most important to object recognition. Any object-matching method employs a similarity metric, whether it is explicit or implicit in the algorithm. The formation of shocks in the reaction-diffusion space and their classification yields a complete representation of the shape. These shocks as discrete events represent the shape not only statically, but also dynamically (in the spirit of Koenderink's dynamic shape [19]) in relation to its "nearby" shapes [15]. Since these deformations simplify shapes in time, the longer it takes two shapes to become similar under these deformations, the more dissimilar they are. Therefore, the degree of similarity of the shock-based representation of shapes in the reaction-diffusion space is indicative of their degree of similarity for object recognition.

References

1. Angenent, S.B. (1988). The zero set of a solution of a parabolic equation, J. für die Reine und Angewandte Mathematik 390, pp. 79–96.
2. Arehart, A., Vincent, L., Kimia, B.B. (1993). Mathematical morphology: The Hamilton-Jacobi connection, Proc. 4th Int. Conf. on Computer Vision, Germany, Berlin, May 11–13.
3. Attneave, F. (1954). Some informational aspects of visual perception, Psych. Review 61, pp. 183–193.
4. Blum, H. (1973). Biological shape and visual science, J. Theor. Biol. 38, pp. 205–287.
5. Boomgaard van den, R. (1993). Towards a morphological scale-space theory, this volume, pp. 631–640.
6. Crandall, M.G., Lions, P.L. (1983). Viscosity solutions of Hamilton-Jacobi equations, Trans. Am. Math. Soc. 277, pp. 1–42.
7. Dill, A.R., Levine, M.D., Noble, P.B. (1987). Multiple resolution skeletons, IEEE Trans. on Pattern Analysis and Machine Intelligence 9(4), pp. 495–504.
8. do Carmo, M.P. (1976). Differential Geometry of Curves and Surfaces. Prentice-Hall, New Jersey.
9. Elder, J., Zucker, S.W. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes, Vision Research 33 (7), pp. 981–991.
10. Gelfand, I. (1963). Some problems in the theory of quasilinear equations, Am. Math. Soc. Translation Ser. 2, pp. 291–381.
11. Hoffman, D.D. Richards, W.A. (1985). Parts of recognition, Cognition 18, pp. 65–96.

12. Hopf, E. (1950). The partial differential equation $u_t + uu_x = \epsilon u_{xx}$, Comm. Pure Appl. Math. 3, pp. 201–230.
13. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1990). Toward a computational theory of shape: An Overview. In: Faugeras, O. (ed.), Lecture Notes in Computer Science, 427 pp. 402–407, Berlin, Springer Verlag.
14. Kimia, B.B. (1990). Conservation Laws and a Theory of Shape, Ph.D. dissertation, McGill Centre for Intelligent Machines, McGill University, Montreal, Canada.
15. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1992). The Shape Triangle: Parts, Protrusions, and Bends, Technical Report TR-92-15, McGill University Research Center for Intelligent Machines.
16. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1993). Shapes, shocks, and deformations, I: The components of shape and the reaction-diffusion space, Int. J. of Comp. Vision, submitted.
17. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1991). Entropy scale-space. In: Arcelli, C. (ed.), Visual Form: Analysis and Recognition, pp. 333–344, New York, Plenum Press.
18. Koenderink, J.J. (1990). Solid Shape, MIT Press, Cambridge, Massachusetts.
19. Koenderink, J.J. van Doorn, A.J. (1986). Dynamic shape, Biol. Cybern. 53, pp. 383–396.
20. Leyton, M. (1992). Symmetry, Causality, Mind, MIT press.
21. Link, N.K., Zucker, S.W. (1987). Sensitivity to corners in flow patterns, Spatial Vision 12(3), pp. 233–244.
22. Pizer, S.M., Oliver, W.R., Bloomberg, S.H. (1987). Hierarchical shape description via the multiresolution symmetric axis transform, IEEE Trans. on Pattern Analysis and Machine Intelligence 9(4), pp. 505–511.
23. Richards, W., Dawson, B., Whittington, D. (1986). Encoding contour shape by curvature extrema, J. Opt. Soc. Am. A 3(9), pp. 1483–1489.
24. Sethian, J.A. (1985). Curvature and the evolution of fronts, Comm. Math. Physics 101, pp. 487–499.
25. Siddiqi, K., Kimia, B.B. (1993). Parts of visual form: computational aspects, Proc. Conf. on Computer Vision and Pattern Recognition, New York.
26. Tresness, K.J., Siddiqi, K., Kimia, B.B. (1992). Parts of Visual Form: Ecological and Psychophysical Aspects, Technical Report LEMS 104, LEMS, Brown University.
27. Zucker, S.W., Dobbins, A., Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation, Neural Computation 1, pp. 68–81.

Performance in Noise of a Diffusion-based Shape Descriptor *

Murray H. Loew and Sheng-Yuan Hwang

Department of Electrical Engineering and Computer Science, George Washington University, Washington DC 20052, USA

Abstract. A diffusion-like process, analogous to the thermodynamic diffusion of heat or of gas molecules, is used to describe the shape of two- or three-dimensional objects. It is effective at identifying extrema of curvature that might be used to segment the boundary, and also at characterizing the types of line segments that lie between the extrema. Both of those operations are essential for qualitative descriptions of images, as would be required by an approach based on geometrical icons (geons). The region need not be convex. The descriptor is invariant to several common transformations, including rotation. It can be implemented easily on parallel machines, does not pose problems with the definition of slope, and appears to be capable of dealing with the matching of partially-occluded objects. The descriptor's performance is essentially independent of user-supplied parameters.

It is shown that noise does not affect the accuracy of identification of the extrema — a simple stopping rule for the process ensures that the structural parts of the boundary are preserved while the noise is suppressed. The procedure is compared and contrasted to scale-based boundary-description methods. Connections are drawn between this work and that of others who use diffusion in scale-space and edge-detection methods.

Keywords: curvature extrema, diffusion, scale-space.

1 Introduction

Previous work [11, 8] has presented a method for describing the shape of a region, based on a simulated discrete-time diffusion process. The region was required to be simply-connected, but need not be convex. The method worked as well in three dimensions as in two; here only the two-dimensional case is considered.

The earlier work showed that the descriptor was invariant under translations, rotations by multiples of 90 degrees, and scale changes, and claimed that the

* We appreciate the constructive comments of the reviewers. This work was supported in part by the Office of Naval Research under Grant N00014-91-J-1539.

method was relatively insensitive to noise; the claim was not, however, supported by proof or by extensive experimentation.

Others have examined diffusion because of the relationship that the diffusion equation [4] has to multiscale representations of grey-scale images [9, 6, 7, 5]. Those methods aim to create families of images, in which the original image is successively blurred, and intensity features become less and less distinct. A feature at a coarse level of resolution is required to possess a "cause" at a finer level of resolution, though the converse need not be true. The features describe regions within the image, which would then lead to robust edge detection or segmentation. Lindeberg [7] provides a fundamental basis and establishes the necessary conditions for performing scale-space operations in a discrete domain. He derives the values of parameters that ensure desirable characteristics such as isotropy and separability in the discrete space. In all cases the aim is to produce a sequence of successively smoothed images that may give insights into the structure of the images and be useful in segmentation.

In contrast, this work seeks to use diffusion only to detect extrema of boundary curvature. It possesses many of the desirable characteristics of other diffusion approaches, and is essentially parameter-free. Because the method (described below) is effective at detecting extrema of curvature, it would appear to be useful in identifying the regions that are important in human object recognition [1, 2]. An issue identified in our earlier work was the choice of stopping rule — when should the process be stopped, and the results examined? The work presented here links the two issues of stopping rule and sensitivity to noise.

2 Methods and Data

2.1 The Diffusion Method

The diffusion-type procedure simulates the release at an initial time of a given number of particles from each pixel along the boundary of a region to be. (It is assumed that the boundary pixels have been defined and that non-boundary pixels within the region are empty initially.) At each instant of discrete time thereafter, new values of pixel contents are computed based on an assumed diffusion constant and the isotropic assumption (i.e., that the diffusion law applies equally in all directions for all parts of the region under study). The process consists of an initial transient and a subsequent steady-state condition. In steady-state, all pixels contain the same number of particles. During the transient condition, however, the number of particles in each boundary pixel depends on the shape of the boundary. The concentration is greater in concavities than in convexities, with straight or nearly-straight regions having intermediate concentrations. It is necessary therefore to stop the diffusion process during the transient condition to detect those characteristics of the boundary. When the simulated diffusion process is stopped, the sequence of numbers of particles in the boundary pixels can be used to generate a shape-related code. Specifically, the positive- and negative-going peaks in a plot of pixel-content vs. boundary-location correspond to regions of high concavity and convexity, respectively.

Let $N_{i,j}(t)$ be the number of particles contained in the pixel at coordinates (i,j) at time t. Then the fundamental algorithm to be used is

$$N_{i,j}(t+dt) = N_{i,j}(t) - 4KN_{i,j}(t) + K(N_{i-1,j}(t) + N_{i+1,j}(t) + N_{i,j-1}(t) + N_{i,j+1}(t)) \quad (1)$$

This is the two-dimensional computational evaluation of

$$I_t = \Delta I = K(I_{xx} + I_{yy}), \quad (2)$$

the diffusion equation, using a simple finite-difference method. Alternatively, diffusion could be modelled as occurring in eight directions. Neighbours that lie outside the boundary of the object do not participate in the process. The solution of (2) is the convolution integral

$$\begin{aligned} I(x,y,t+dt) &= K \int\int I(x-u, y-v, 0) G(u,v,t+dt) du dv \\ &= K \int\int I(x-u, y-v, t) G(u,v,dt) du dv \ , \end{aligned}$$

where $I(x,y,0)$ is the initial condition.

The computational solution is usually

$$N_{i,j}(t+dt) = K \sum_k \sum_l N_{i-k,j-l}(t) G_{kl}(dt) \ .$$

The numerical accuracies of the two methods are similar [3, 4].

The diffusion equation is governed by the maximum principle [12], which states that all the maxima of the solution to the equation belong to the initial condition (in this case, the boundary of the image), and to the boundary of the domain of interest, if the diffusion constant is positive.

In a real diffusion process of matter or heat, the diffusion constant K would depend on the nature of the material and/or experimental data, and the isotropic assumption (realized only approximately in this discrete case) might not apply. In this simulated process, however, the only constraint is that K be appreciably smaller than the reciprocal of the number of directions (4 or 8) in which diffusion is permitted to occur. This ensures that the outflow from a boundary pixel does not cause computational problems early in the process. Extensive experimentation has shown that variations in K have virtually no effect on the shape-describing power of the method. There is an inverse relationship between K and t: increasing K results in the achievement of a desired stopping criterion at a smaller value of t; but small variations along the boundary are then lost because of the larger amounts by which pixel contents change at each time step.

When should the process be terminated? If the peaks are important, then a reasonable stopping rule would be the one that maximizes the amplitude difference between the peaks; that is, it would compute the difference between maximum and minimum along the boundary at each t, and the point at which that difference was maximized would be the stopping time. This yields the maximum signal-to-noise ratio for the extrema. Typically, however, that rule will

lead to a relatively early termination of the process, before appreciable smoothing of the boundary occurs. The set of boundary values, for each iteration of the process, constitutes a scale-space representation at a "scale" given by the time t. Figure 4 offers four kinds of measures to be considered as inputs to stopping rules, based on the goals of expressing the variation in pixel contents along the boundary and of determining when near-stability has been reached. The measures (all computed using only the boundary pixels), and the criteria for choice, are:

1. standard deviation: use point at which maximum occurs, and/or the point at which the value stabilizes;
2. coefficient of variation (standard deviation divided by mean): use point at which maximum occurs, and/or the point at which the value stabilizes;
3. mean: use knee of curve;
4. max-min difference: use maximum, or knee of curve.

Figure 4 provides these measures for the example of the N SKIRT (noisy) object shown in Fig. 3.

2.2 Objects Studied

A set of images of objects was generated that included a representative group of segment and extrema types — straight lines, concave and convex corners, and curves of large and small curvature. Examples of the objects appear in Fig. 3. Noise was introduced by generating a white Gaussian sequence and adding one of its elements to each pixel; then boundary pixels and their neighbours were compared to a threshold and rendered as part of the object boundary, or as part of the background. Pits and bumps along the boundary, and connected to it, were thereby created; they were one pixel deep or high, respectively, and one or more pixels long. The noisy versions of the images are shown also in Fig. 3.

The basic idea of the process is described as follows. The object has $10,000$ particles placed in each of its boundary pixels at $t = 0$, and the diffusion process carried out with $K = 0.01$. At $t = 10$, the contents of the boundary pixels are as shown in Fig. 1. A plot of the boundary-pixel contents as a function of boundary position appears in Fig. 2, for $t = 3$ and $t = 10$. It is clear that the concave corners of the object correspond to the two large positive-going peaks, and that the convex protrusions into the object correspond to the two large negative-going ones. Straight sections of the boundary correspond to constant or nearly-constant values in the plot. It is often convenient for comparisons between plots to normalize them by subtracting the mean (computed along the boundary) from each boundary pixel, and dividing by the standard deviation (also computed using only the boundary pixels).

Diffusion was performed on N SKIRT and plots were generated (Fig. 5) that show normalized (as in Sect. 2.1) pixel contents at each of five instants, in accordance with the candidate procedures for stopping noted above. Maximum standard deviation occurred at $t = 2$; maximum difference at $t = 5$, maximum coefficient of variation (COV) at $t = 75$, stable mean at $t = 163$, and both stable

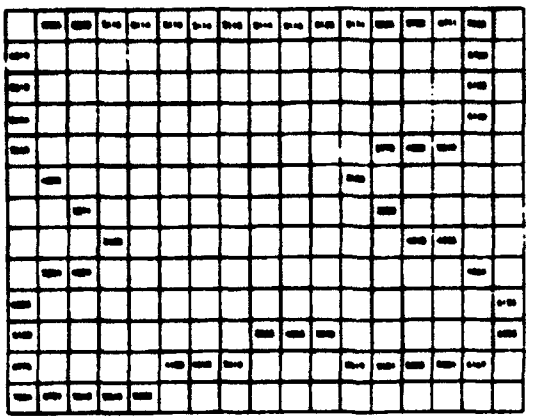


Fig. 1. An irregular shape at $t = 3$ with $k = 0.01$.

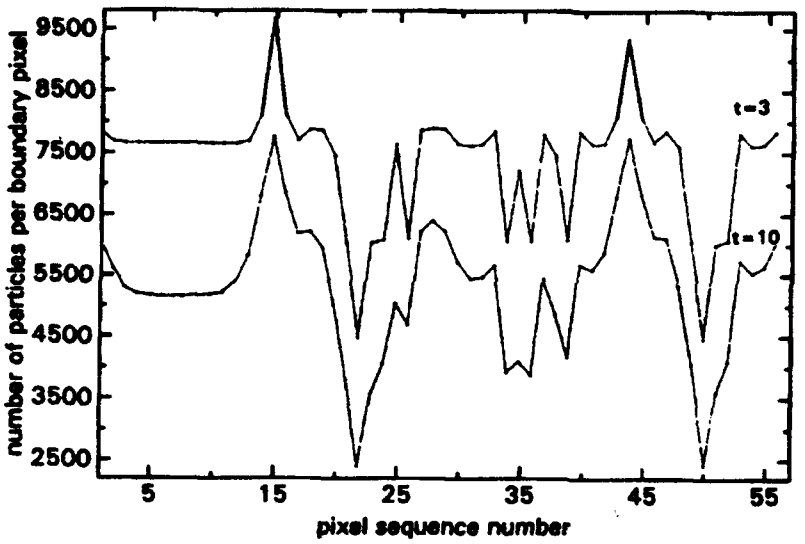


Fig. 2. Number of particles for concecutive boundary pixels for the shape of Fig. 1, at $t = 3$ and $t = 10$.

standard deviation and stable COV at $t = 212$. Numbers on the peaks of the plots in Fig. 5 correspond to the vertex numbers on the N SKIRT object in Fig. 3.

The distinct positive and negative peaks in the $t = 75, 163$, and 212 plots (called the "late" plots) correspond identically to the concave and convex corners, respectively, in the N SKIRT object. The horizontal and vertical straight lines between vertices in the object correspond to flat or nearly-flat segments of the diffusion plots. The "early" plots contain the same set of corner-related peaks, as well as peaks due to the noise bumps and pits evident in Fig. 3. Additionally,

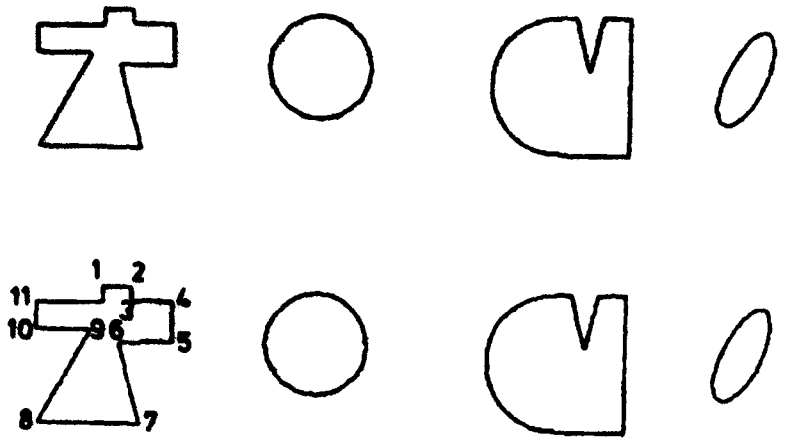


Fig. 3. Test objects (upper row without noise, lower row with noise $\sigma = 60$) left to right: SKIRT, CIRCLE1, CIRCLE3, OVAL.

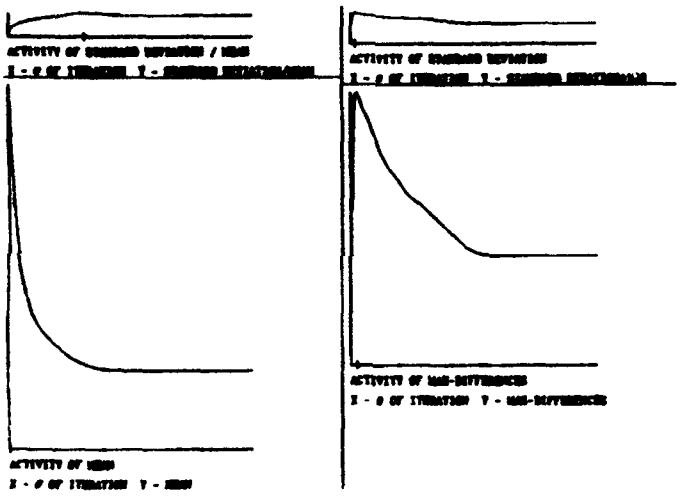


Fig. 4. Four stopping criteria for the noisy SKIRT object (N SKIRT).

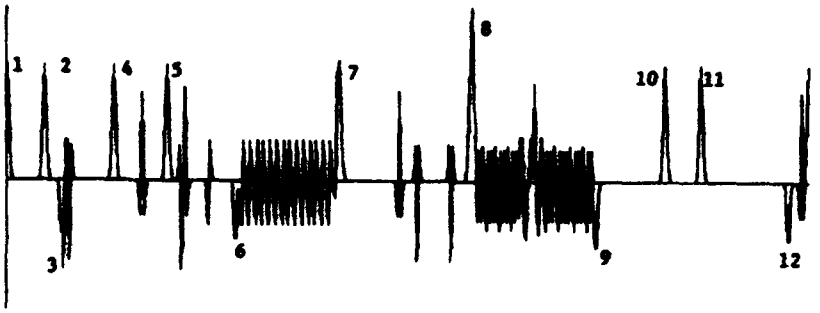


Fig. 5. N SKIRT at $t = 2$, corresponding to maximum standard deviation σ. Labelled peaks correspond to vertices in Fig. 3.

the early plots contain oscillations (between vertices 6 and 7 and also 8 and 9) arising from the staircase approximations to the diagonal edges in N SKIRT. Staircases can be viewed as sequences of alternating concavities and convexities, which explains the oscillatory behaviour. Both they and the noise are smoothed by the diffusion process during the interval between $t = 5$ and $t = 75$.

An oval and several kinds of circles were also examined; the diffusion results for those shapes are shown in Figs. 6, 7, and 8.

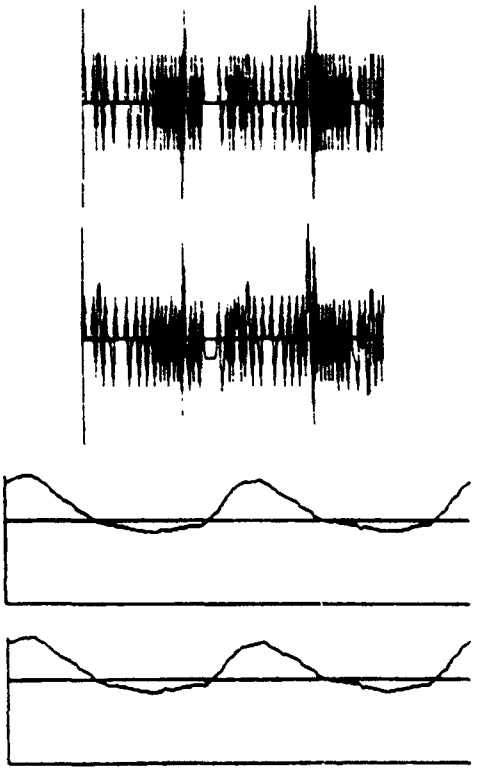


Fig. 6. (top to bottom) N OVAL at $t = 1$ (maximum σ), $t = 2$ (maximum difference), $t = 108$ (maximum σ/μ), and $t = 131$ (stable σ and stable σ/μ).

2.3 Properties of the Diffusion Process

It is the nature of a corner or of a general extremum of curvature that its shape and existence are defined by its set of neighbours. Noise, if uncorrelated (white), has no such local support, in general. It is therefore not unexpected that the diffusion process, acting in part as a local averager, suppresses small-scale variations while preserving major differences.

The peaks in the late plots of Fig. 5 have broader bases than had the peaks in the early plots that correspond to the same vertices in the original figure. This spreading of the peaks over time is a direct consequence of the diffusion process,

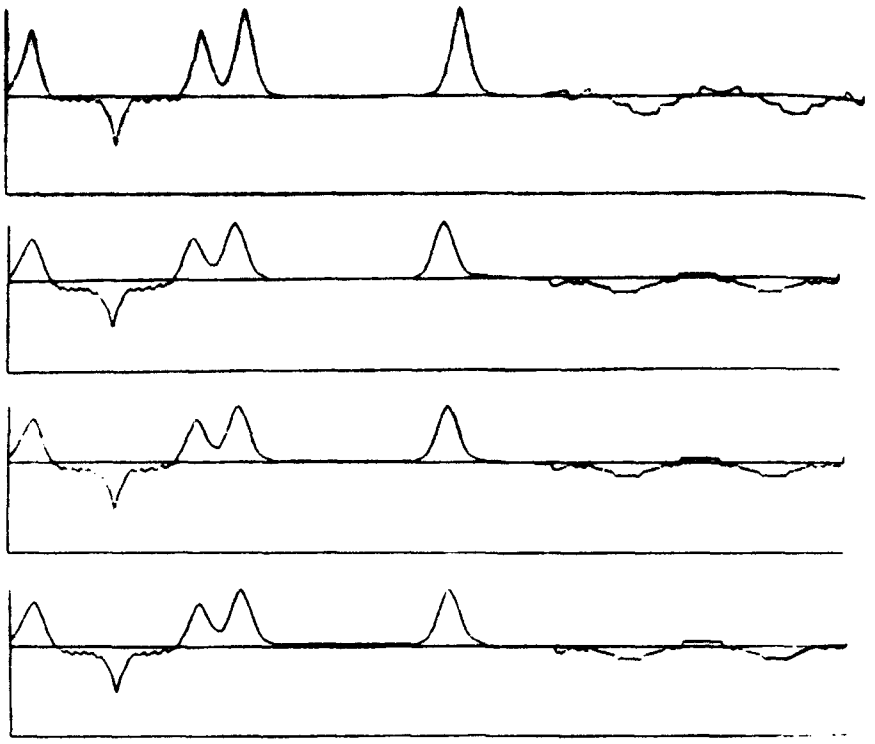


Fig. 7. (top to bottom) CIRCLE3 at $t = 91$ (maximum σ/μ), $t = 135$ (stable μ), $t = 146$ (stable maximum difference), and $t = 170$ (stable σ and stable σ/μ).

as the (relatively large) contents of the positive peaks' pixels spread into those neighbouring pixels that help to define the concave corner. Similarly, convex corners' negative peaks in the diffusion plot also spread out, as particles from more-distant pixels contribute to the (relatively small) contents of the pixels in the neighbourhoods of those corner pixels.

Peaks in the early plots, that arise from oscillations due to diagonal-line sampling and from noise, have no local support — there are no structural characteristics of the boundary to sustain the extreme amplitude. Rather, because the neighbouring pixels are of a different (and relatively uniform) structure, they serve to reduce that extreme amplitude by effectively averaging it with their own contents.

This behaviour may be summarized by the observation that structural peaks spread, while artefactual ones vanish. Examination of the late plots, then, would reveal the *structural* properties of the boundary — the existence of the important extrema — while the early plots would give precise *locations* of the extrema and information about the *kinds* of segments that lie between the extrema. So objects that appear in the late plots could point back to the early plots for the details that would contribute to a simulation of human object understanding. This is exactly the implication of the maximum principle, noted above, that accounts for the fact that no new peaks are generated after the first iteration of the diffusion

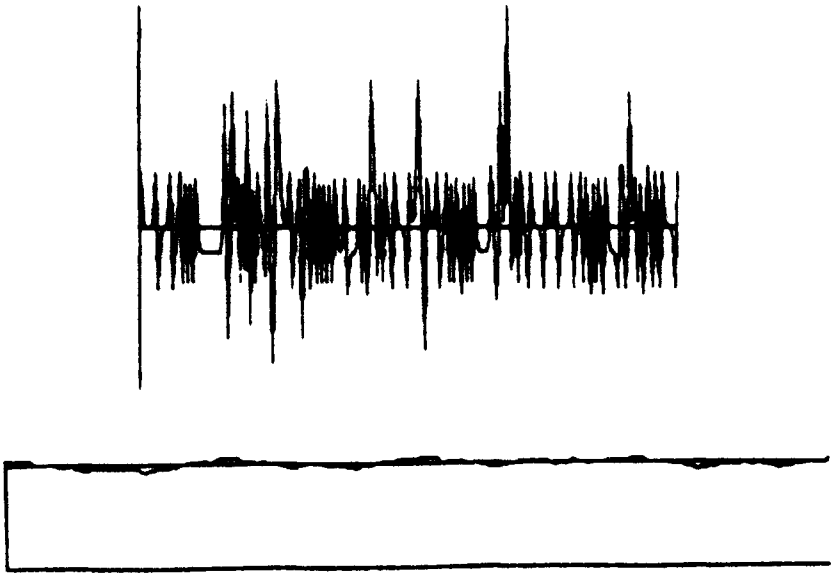


Fig. 8. (top to bottom) N CIRCLE1 at $t = 2$ (maximum difference), and $t = 139$ (stable μ and stable maximum difference).

process.

3 Discussion and Conclusions

This work bears some similarity to that of Witkin [13], who proposed scale-space filtering as a way of detecting and identifying visually significant features in a one-dimensional signal. Gaussian smoothings of the signal were performed using a series of values of the standard deviation, and the extrema then located by finding the zero-crossings of the second derivative of each smoothed signal. Contours connecting those extrema at the different scales (standard deviations) gave an indication of the location and visual importance of the extrema. No basis was given for choosing the range and increments of the standard deviations used, and there was no procedure recommended for extension to two-dimensional data. One approach to 2-D data was suggested by Rattarangsi and Chin [10], in which the boundary points coordinates were parameterized as $x(s)$ and $y(s)$, and addressed independently of one another.

The diffusion process described here is intrinsically 2-D and uses discrete time as a natural index of its progress; the stopping time is determined by the stabilization of simple measures taken around the boundary, and does not require that parameters be chosen. The procedure does require that the boundary of the object be determined in advance, and so this approach does not itself fully utilize the grey-scale information of the original image. It is easy to see, however, how the scale-space-dependent boundary detection of, for example, Perona and Malik could be augmented — at whatever scale — by the present procedure to

identify the significant extrema of curvature of the boundary. They are indicated by the local maxima in the diffusion plot when the process stops, or at any earlier time. The locations of those broad peaks point to precise values of the boundary positions of the extrema in the early plots. In addition, the nature of the diffusion waveform lying *between* each pair of extrema in the early plots provides a description of the kind of line-segment present: straight, or curved with a measure of the curvature. It therefore should be useful in extracting and identifying geons for subsequent object recognition. The modest amounts of noise used to date were successfully removed and did not affect the performance of the process.

References

1. Attneave, F. (1965). Some informational aspects of visual perception, Psychol. Rev. 61, pp. 183–193.
2. Biederman, I. (1985). Human image understanding: recent research and a theory, Computer Vision, Graphics, and Image Processing 32, pp. 29–73.
3. Burden, R. L., Faires, J.D. (1989). Numerical Analysis, 4th ed., PWS-Kent, Boston.
4. Ghez, R. (1988). A Primer of Diffusion Problems, Wiley, New York.
5. Koenderink, J. J., van Doorn, A.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
6. Lindeberg, T. (1990). Scale-space for discrete signals, IEEE Trans. on Pattern Analysis and Machine Intelligence 12, pp. 234–254.
7. Lindeberg, T. (1993). Scale-space for N-dimensional discrete signals, this volume, pp. 571–590.
8. Loew, M. H. (1987). A diffusion-based description of shape. In: Devijver, P.A., Kittler, J. (eds.), Pattern Recognition Theory and Applications, NATO ASI Series, Vol. F30, Springer-Verlag, Berlin, pp. 501–508.
9. Perona, P., Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion, IEEE Trans. on Pattern Analysis and Machine Intelligence 12 (7), pp. 629–639.
10. Rattarangsi, A., Chin, R.T. (1990). Scale-based detection of corners of planar curves, Proc. 10th Int. Conf. Pattern Recognition, Atlantic City, pp. 923–930.
11. Skliar, O., Loew, M.H. (1985). A new method for characterization of shape, Pattern Recognition Letters 3, pp. 335–341.
12. Widder, D. V. (1975). The Heat Equation, Academic Press, New York.
13. Witkin, A. P. (1983). Scale-space filtering, Proc. 8th Int. Joint Conf. Artif. Intell., Karlsruhe, Germany, pp. 1019–1022.

Towards a Morphological Scale-Space Theory

Rein van den Boomgaard and Arnold W.M. Smeulders

Department of Mathematics and Computer Science, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

Abstract. In this paper it is shown that erosions and dilations using increasingly larger quadratic structuring functions can be used to construct a morphological scale-space which is incrementally computable (the image at scale ρ can be calculated from the image at scale μ for $\mu \leq \rho$) and the (weak) solution of a differential equation. Furthermore it is argued that the morphological scale-space preserves causality in the resolution domain, in the sense that no spatial details are introduced by moving towards larger scales. This is illustrated with an example showing the singularity trace through scale-space.

Keywords: mathematical morphology, morphological scale-space, evolutionary systems, quadratic structuring function, skeleton, propagator, Burger's equation.

1 Introduction

The use of scale-space is nowadays well accepted in computer vision. Using a scale-space an image can be analyzed at all levels of resolution simultaneously. The way in which the visual details develop while going from low to high resolution defines the structure in the image.

In this paper the following requirements for the construction of a morphological scale-space are used:

1. The scale-space is a one-parameter family of images $F(x, \rho)$ with x an n-dimensional spatial vector and ρ the scale parameter. For $\rho = 0$ the original image f is obtained (i.e. $F(x, 0) = f(x)$).
2. The scale-space is incrementally computable, i.e. $F(x, \rho+d\rho)$ can be obtained from $F(x, \rho)$.
3. The scale-space $F(x, \rho)$ satisfies a differential equation linking an infinitesimal small change in scale (going from scale ρ to scale $\rho + d\rho$) with spatial properties of $F(x, \rho)$.
4. The scale-space "preserves causality in the resolution domain" (see [7]), i.e. it is required that by moving from high to low resolution no spatial details are introduced (only removed).

A well-known example of a scale-space is the one obtained by linearly smoothing the original image with increasingly larger Gaussian filters. Section 2 briefly states some of the properties of the Gaussian scale-space.

Instead of building a scale-space using the linear Gaussian filter, the use of morphological image transforms has also been proposed. Chen and Yan [2] investigated the "opening scale-space" obtained by opening a binary image with increasingly larger disks. They claimed that the morphological scale-space thus obtained preserves causality in the sense that zero-crossings in the contour are never created when moving towards larger scales. Nacken [9] has proved this claim to be incorrect by giving a counterexample. However, in our opinion this does not imply that morphological scale-spaces cannot be constructed. This paper indicates that whereas zero-crossings are a suitable starting point for Gaussian scale-spaces, this is not the case for morphological scale-spaces.

In this paper it will be shown that a morphological scale-space for grey-value images can be constructed which satisfies a differential equation. Kimia [6] has found similar differential equations describing the deformation of 2-D contours in scale-space. Considering only 2-D contours, Kimia then combines the morphological scale-space with the Gaussian linear scale-space.

Instead of Gaussian filtering, the erosion (dilation) of the original image by increasingly larger *quadratic structuring functions* is used. These structuring functions share many of the properties of the Gaussian filter and are called the morphological equivalents of the Gaussian functions [1]. The restriction to quadratic structuring functions is not essential. In [1] the general class of differential equations solved with morphological operations is considered.

2 Gaussian Scale-Space

An important property of the Gaussian function is that it is Green's function (or propagator) of the *diffusion equation*. Let $x \in \mathbb{R}^2$ and $\rho \in \mathbb{R}^+$ and let $F(x, \rho)$ be a real-valued function satisfying the diffusion equation:

$$F_\rho = \nabla^2 F$$

where $F_\rho = \partial F / \partial \rho$ and $\nabla^2 F$ is the Laplacian of F. With initial condition $F(x, 0) = f(x)$ (in computer vision f is the image being analyzed) the diffusion equation is solved using the Gaussian function to propagate the initial condition into the "scale-space" $F(x, \rho)$:

$$F(x, \rho) = (f * g^\rho)(x),$$

where $*$ denotes the convolution and g^ρ is the Gaussian function with scale parameter ρ:

$$g^\rho(x) = \frac{1}{4\pi\rho} e^{-\frac{\|x\|^2}{4\rho}}.$$

In computer vision the function $F(x, \rho)$ is interpreted as a family of images, where ρ indicates the level of resolution (or scale). The larger ρ is, the more

blurred the original image f will be, finally showing only the larger structures in the image, until ultimately any image detail disappears.

The diffusion equation serves as an apt starting point for constructing a scale-space (multiresolution representation) because it satisfies a *maximum principle* (see Hummel [5]). An immediate consequence of the maximum principle is that the scale-space "generated" by the diffusion equation *preserves causality in the resolution domain* [7], in the sense that by moving towards larger scales new details are never created. It is also the maximum principle which explains the unique properties of the zero-crossings in the Laplacian of F.

As pointed out by Perona and Malik [10] Gaussian blurring "does not respect the natural boundaries of objects". Objects that are better left unmerged *are* merged. Furthermore edge junctions (corners) are destroyed. They introduced an inhomogeneous blurring scheme in which the amount of (infinitesimal) isotropic blurring needed to obtain $F(x, \rho + d\rho)$ from $F(x, \rho)$ is determined by the magnitude of the gradient. They claimed that the scale-space thus obtained satisfies a maximum principle and thus guarantees preservation of causality in the resolution domain.

Most images analyzed by either the human visual system or a computer vision system are projections of three-dimensional reality on a two-dimensional retina. Projective image formation makes linearity of visual stimuli a questionable assumption. The scheme of Perona and Malik may be advantageous over the original scheme, but it still does not tackle the questionable assumption of linearity in the visual stimuli.

3 Elements from Mathematical Morphology

In this section the morphological notation used in this paper is given and some properties that are needed are discussed (see also [8, 4, 3]).

The erosion of a grey-value image f using *structuring function* g, using $\wedge$ as a shorthand for the infimum operator, is given by:

$$(f \ominus \check{g})(x) = \bigwedge_{y \in \mathbb{R}^n} [f(x+y) - g(y)]. \tag{1}$$

The dilation $f \oplus \check{g}$ is defined as:

$$(f \oplus \check{g})(x) = \bigvee_{y \in \mathbb{R}^n} [f(x+y) + g(y)], \tag{2}$$

where $\vee$ denotes the supremum operator. It should be noted that whereas in convolution kernels the pixels with zero value do not influence the convolution sum, in structuring functions these pixels have value $-\infty$. Table 1 summarizes the definitions of grey-value morphology needed in this paper.

Let g be a structuring function whose umbra $\mathbf{U}(g) = \{(x,t) \mid f(x) \geq t\}$ is a convex set. It is well known in mathematical morphology that any convex set S is divisible with respect to dilation, in the sense that $\alpha S \oplus \beta S = (\alpha + \beta)S$. This

Table 1. Notation and definitions of grey-value morphology. f,g and h denote functions (e.g. $f : \mathbb{R}^n \mapsto \mathbb{R}$) and x and y denote a position vector in $\mathbb{R}^n$. An element from $\mathbb{R}^{n+1}$ will be denoted as $(x;t)$, where $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Name	Notation	Definition
Translation	$f_{(y;t)}$	$f_{(y;t)}(x) = f(x-y)+t$
Complement	f^c	$f^c = -f$
Transpose	$\check{f}$	$\check{f}(x) = f(-x)$
Union	$f \vee g$	$(f \vee g)(x) = f(x) \vee g(x)$
Intersection	$f \wedge g$	$(f \wedge g)(x) = f(x) \wedge g(x)$
Dilation	$f \oplus \check{g}$	$(f \oplus \check{g})(x) = \bigvee_{y \in \mathbb{R}^n} [f(x+y)+g(y)]$
Erosion	$f \ominus \check{g}$	$(f \ominus \check{g})(x) = \bigwedge_{y \in \mathbb{R}^n} [f(x+y)-g(y)]$
Closing	$f \bullet g$	$(f \oplus \check{g}) \ominus g$
Opening	$f \circ g$	$(f \ominus \check{g}) \oplus g$

is also true for "convex" structuring functions. Define the family of structuring functions:

$$g^\rho(x) = \rho g\left(\frac{x}{\rho}\right).$$

Such a family of structuring functions is closed with respect to dilation [1]:

$$g^\rho \oplus g^\mu = g^{\rho+\mu}.$$

Among all convex structuring functions the quadratic structuring function (QSF)

$$q(x) = -\frac{1}{4}\|x\|^2,$$

is the only rotationally symmetric one which can be dimensionally decomposed [1] (the choice of the scaling factor 1/4 is irrelevant now, but will be clear later on). This means that a dilation $f \oplus q$ can be implemented by first dilating along the rows in an image, followed by a dilation along the columns in the image. The QSF q shares this property with the Gaussian function, which is the unique rotationally symmetric function that can be dimensionally decomposed with respect to convolution.

4 Morphological Propagators

Consider the dilation of an image f with the QSF q^ρ (see Fig. 1). The downwards pointing parabola positioned at point $(x, (f \oplus q^\rho)(x))$ hits the original function at the point y (which is called the *point-of-contact*). Note that the parabola, when placed at another point on the dilated function, may hit the original image at more than one point. These situations leading to singularities in the dilation result are treated later.

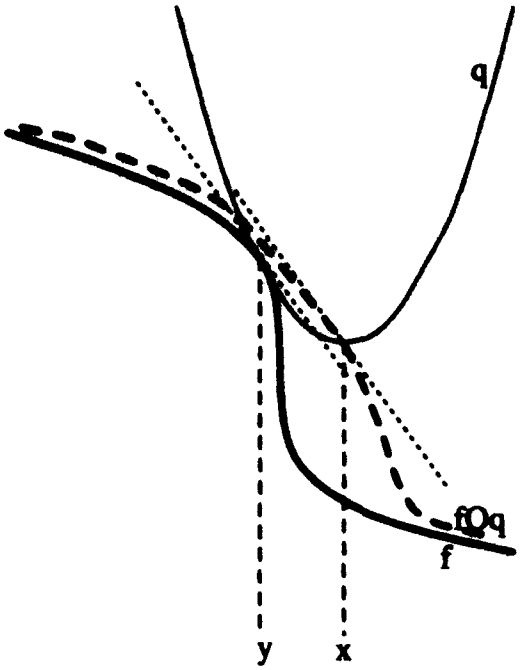


Fig. 1. Dilation with parabola

In [1] it is shown that if $f \oplus q^\rho$ is differentiable in x the point-of-contact y can be calculated:

$$y = x + 2\rho\nabla(f \oplus q^\rho)(x), \tag{3}$$

$$f(y) = (f \oplus q^\rho)(x) - q^\rho(x - y). \tag{4}$$

The above point-of-contact equations are easily proved by observing that the gradient of $f \oplus q^\rho$ in point x is equal to the gradient of the structuring function q^ρ in the point $x - y$, i.e. $\nabla(f \oplus q^\rho)(x) = (\nabla q^\rho)(x - y)$. The choice of a QSF q^ρ allows us to calculate an "inverse gradient function" (i.e. given $y = (\nabla q^\rho)(x)$, the function $(\nabla q^\rho)^{-1}$ such that $x = (\nabla q^\rho)^{-1}(y)$ exists). The inverse gradient function of q^ρ is given by $(\nabla q^\rho)^{-1}(x) = -2\rho x$.

It should be noted that an inverse gradient function does not exist for all structuring functions. For example all "flat structuring functions" are excluded from the analysis in this paper.

The point-of-contact equations easily lead to the differential equation linking an infinitesimal change in scale (going from scale ρ to scale $\rho + d\rho$) with spatial properties of the dilated function. Define the function $F^\oplus(x, \rho)$:

$$F^\oplus(x, \rho) = (f \oplus q^\rho)(x);$$

then with initial condition $F^\oplus(x, 0) = f(x)$, $F^\oplus$ is a weak solution of the differential equation:

$$F^\oplus_\rho = \|\nabla F^\oplus\|^2.$$

This can be proved by showing what happens when the scale is increased from ρ to $\rho + d\rho$. Because the QSFs are closed with respect to dilation it is true that:

$$F^{\oplus}(x, \rho + d\rho) = (f \oplus q^{\rho + d\rho})(x) = ((f \oplus q^{\rho}) \oplus q^{d\rho})(x) = (F^{\oplus}(\cdot, \rho) \oplus q^{d\rho})(x).$$

Using the point-of-contact equations (note that the dilation is done with $q^{d\rho}$) leads to:

$$\begin{aligned} F^{\oplus}(x, \rho) &= F^{\oplus}(x, \rho + d\rho) - q^{d\rho}(-2d\rho \nabla F^{\oplus}(x, \rho + d\rho)) \\ &= F^{\oplus}(x, \rho + d\rho) + d\rho \|\nabla F^{\oplus}(x, \rho + d\rho)\|^2. \end{aligned}$$

Neglecting all second- and higher-order terms in the Taylor expansion of $d\rho \|\nabla F^{\oplus}(x, \rho + d\rho)\|^2$ the above equation simplifies to:

$$F^{\oplus}(x, \rho) = F^{\oplus}(x, \rho + d\rho) + d\rho \|\nabla F^{\oplus}(x, \rho)\|^2;$$

rearranging terms and dividing by $d\rho$ gives:

$$\frac{F^{\oplus}(x, \rho) - F^{\oplus}(x, \rho + d\rho)}{d\rho} = \|\nabla F^{\oplus}(x, \rho)\|^2.$$

For $d\rho \to 0$ the differential equation:

$$F^{\oplus}_{\rho} = \|\nabla F^{\oplus}\|^2$$

is obtained. A formal proof can be found in [1], where the above differential equation is a special case of the class of differential equations obtained by considering the class of structuring functions that have an inverse gradient function.

Replacing dilation with erosion, $F^{\ominus}(x, \rho) = (f \ominus q^{\rho})(x)$, the differential equation becomes:

$$F^{\ominus}_{\rho} = -\|\nabla F^{\ominus}\|^2.$$

5 Morphological Scale-Space

The morphological scale-space $F(x, \rho)$ is defined by

$$F(x, \rho) = \begin{cases} F^{\oplus}(x, \rho) = (f \oplus q^{\rho})(x) : \rho \geq 0, \\ F^{\ominus}(x, \rho) = (f \ominus q^{\rho})(x) : \rho < 0. \end{cases}$$

In the previous sections it has been shown that

1. The scale-space is a one-parameter family of images. In contrast to the Gaussian scale-space the scale parameter ranges from $-\infty$ to $+\infty$. This reflects the fact that whereas linear operators treat object and background alike (a convolution is *self-dual*), in morphology the duality between object and background is explicitly dealt with. In a sense the scale-space defined above is comprised of two tightly linked scale-spaces: one for the objects and one for the background.
2. Because the QSFs are closed with respect to dilation (i.e. $q^{\rho} \oplus q^{\mu} = q^{\rho + \mu}$) the scale-space is incrementally computable.

3. The morphological scale-space satisfies a differential equation:

$$F_\rho = \begin{cases} \|\nabla F\|^2 : \rho \geq 0, \\ -\|\nabla F\|^2 : \rho < 0. \end{cases}$$

Comparing the above summary with the scale-space requirements stated in the introduction, the requirement of causality still has to tackled. As already said in a previous section, a parabola may hit the original function at more than one point. If it does a singularity in the dilation is the result (singularity in the sense that the dilation result $f \oplus q^\rho$ is not differentiable at that point). This also explains why $F^\oplus$ and $F^\ominus$ are *weak* solutions of the differential equations.

In order to show the importance of the singularities consider the function

$$\mu_X(x) = \begin{cases} +\infty : x \in X, \\ 0 : x \notin X \end{cases}$$

where X is a two-dimensional set. In [1] it has been shown that the erosion of μ_X with the QSF q^ρ results in squared distance transform of the set X, giving at each point $x \in X$ the square of the distance to the nearest point in X^c divided by 4ρ. The set of all non-differentiable points in the distance transform function forms the skeleton of the set X (provided the set X is sufficiently smooth). The skeleton points together with the distance value at those points provide a complete description of the original set. Let $S(X)$ be the skeleton of X (i.e. $f \ominus q^\rho$ is non-differentiable at all points of $S(X)$) and let s be the (squared) distance weighted skeleton:

$$s(x) = \begin{cases} (\mu_x \ominus q^\rho)(x) : x \in S(X), \\ 0 : x \notin S(X). \end{cases}$$

Dilation of s results in a function $s \oplus q^\rho$ such that all points x with $(s \oplus q^\rho)(x) > 0$ form the original set X. However, setting all points in s to zero where the original distance was less than λ gives:

$$s_\lambda(x) = \begin{cases} s(x) : s(x) \geq \lambda, \\ 0 : s(x) < \lambda; \end{cases}$$

then dilating s_λ with q^ρ results in the function $s_\lambda \oplus q^\rho$ such that all points x with $(s_\lambda \oplus q^\rho)(x) > 0$ form the *opening* of the original set $X \circ 2\sqrt{\rho\lambda}B$ where B is the disk with radius 1. Thus, the singularities of $\mu_X \ominus q^\rho$ do not only provide a complete description of the original set but also of all openings of that set.

It should be noted that because a "binary function" μ_X is used as the "original" image, all images in the dilation scale-space are scaled versions of the squared distance transform. Thus for binary images the singularities are independent of the scale ρ at which they are formed.

Considering arbitrary grey-value images to start with the analysis is somewhat different. However, it should be noted that there is a close correspondence between functions and sets (using the notion of an umbra). In fact the analysis in [1] started with sets and only at the end was the transition to functions

made. This implies that the above analysis for sets can also be used for arbitrary functions. It is evident that the positions of the singularities in this case are dependent on the scale ρ. It may be conjectured that:

- singularities (and the scale at which they are formed) provide a complete description of the original image;
- singularities form traces in the spatial-scale space such that traces only merge when moving from a lower to a higher scale (this is the morphological equivalent of the causality requirement);
- all singularities at scale ρ and higher provide a complete description of the closing $f \bullet q^\rho$ (considering the singularities in the dilation scale-space).

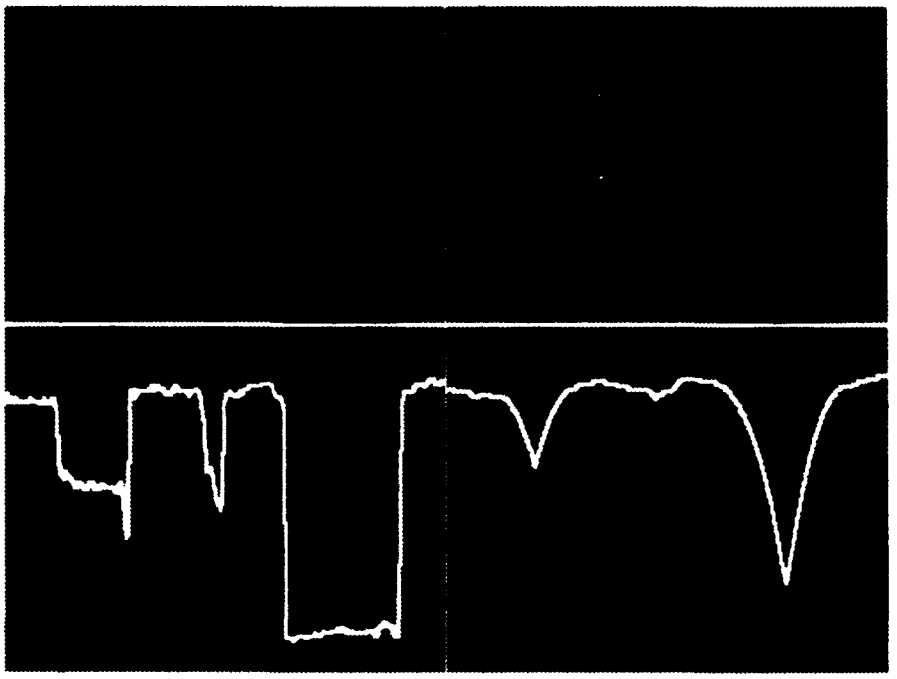


Fig. 2. Parabolic dilation

In Fig. 2 an example of a dilation of an image with a parabolic structuring function is shown. Superimposed on the images is the grey-value profile along the indicated line. Note the singularities in the dilation result. These singularities are easily detected by closing the dilated image with a small-scale parabola and taking the difference with the dilation.

In Fig. 3 an example of a singularity trace through the dilation scale-space is shown. The singularities at each scale level are superimposed on the original image (the singularities are drawn 3 pixels thick so that they become clearly visible). Note that the thin leg of the Rietveld chair only shows up at lower scales, it completely disappears at higher scales.

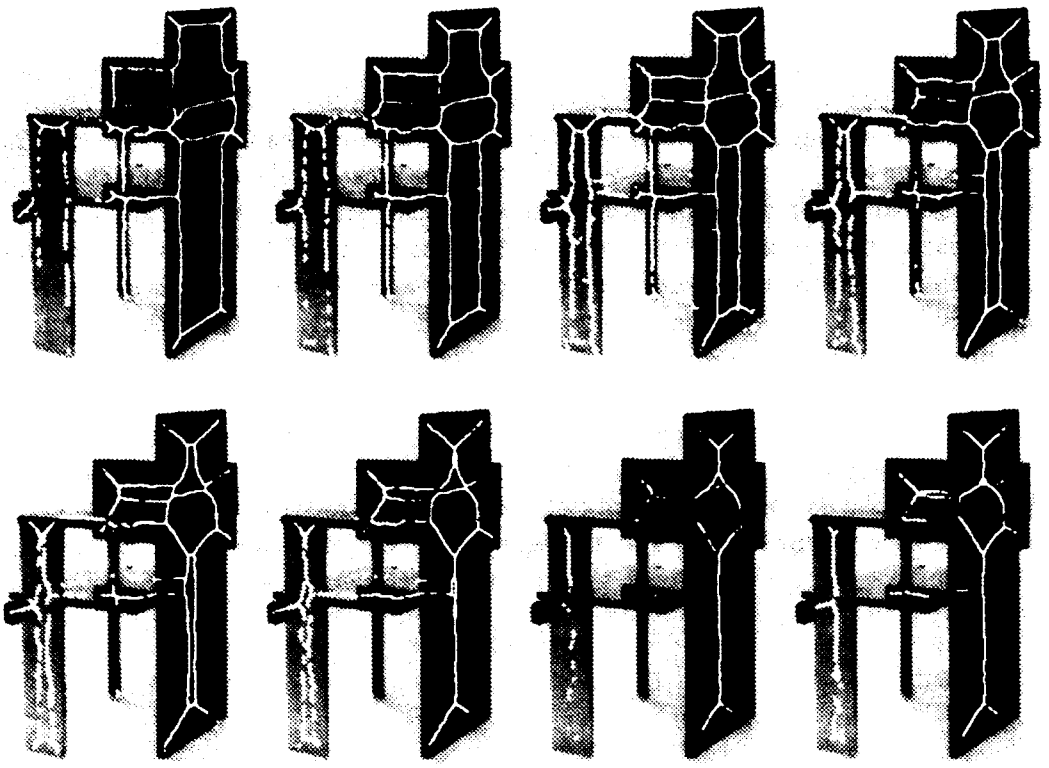


Fig. 3. Singularity trace through scale-space. From left to right, top to bottom, the scales range from 1 to 8 (integer valued).

6 Conclusion

In the introduction four requirements for a scale-space were given:

1. One-parameter family of images. In contrast to the Gaussian scale-space the scale parameter ranges from $-\infty$ to $+\infty$. This reflects the fact that whereas linear operators treat object and background alike (a convolution is *self-dual*), in morphology the duality between object and background is explicitly dealt with. In a sense the scale-space defined above is comprised of two tightly linked scale-spaces: one for the objects and one for the background.
2. The morphological scale-space is incrementally computable because the class of quadratic structuring functions is closed with respect to dilation. Furthermore quadratic structuring functions are of practical importance because they can be dimensionally decomposed.
3. The morphological scale-space satisfies a differential equation. The differential equation bears great resemblance to Burger's equation, which describes the propagation of shock-waves (see [1]).
4. The scale-space should preserve causality.

The results in this paper prove that morphological scale-spaces can be constructed using quadratic structuring functions, meeting the first three requirements.

The zero-crossings (providing the "signature" of the Gaussian scale-space) are of less importance for the morphological scale-space [2]. It has been argued that instead the singularities formed in the morphological scale-space provide the signature. For binary images the singularities are closely related to the skeleton which is known to be a complete description of the original set.

The use of the opening and closing is more often considered than the erosion and dilation to construct morphological scale-spaces. However, it is impossible to describe the closing function $f \bullet g^\rho$ with a differential equation. Furthermore it is conjectured that the dilation scale space (more specifically the singularity traces thereof) completely describes the closing "scale-space" as well. This can be interpreted as follows. In the same way as a 2-D set can be reconstructed from its distance weighted skeleton a function f can be reconstructed from its scale weighted singularity trace. Omitting all skeleton points with distance less than d from the reconstruction leads to the reconstruction of the opening (of size d). In the same way it is argued that this is true for the singularity reconstruction.

References

1. Boomgaard, R. van den. (1992). Mathematical Morphology: Extensions towards Computer Vision, Phd-thesis, University of Amsterdam.
2. Chen, M.H., Yan, P.F. (1989). A multi scaling approach based on morphological filtering, IEEE Trans. on Pattern Analysis and Machine Intelligence 11, pp. 694–700.
3. Giardina, C.R., Dougherty, E.R. (1988). Morphological Methods in Image and Signal Processing, Prentice Hall.
4. Haralick, R.M., Sternberg, S.R., Zhuang, X. (1987). Image analysis using mathematical morphology, IEEE Trans. on Pattern Analysis and Machine Intelligence 9, pp. 532–550.
5. Hummel, R., Muniot. R. (1989). Reconstructions from zero crossings in scale space, IEEE Trans. on Acoustics, Speech and Signal Processing 37, pp. 2111–2130.
6. Kimia, B.B., Tannenbaum, A.R., Zucker, S.W. (1993). Exploring the shape manifold: the role of conservation laws, this volume, pp. 601–620.
7. Koenderink. J.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
8. Maragos, P. (1987). Tutorial on advances in morphological image processing and analysis, Opt. Eng. 26. pp. 623–632.
9. Nacken, P. (1993). Openings can introduce zero-crossings in boundary curvature, IEEE Trans. on Pattern Analysis and Machine Intelligence, in press.
10. Perona, P., Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion, IEEE Trans. on Pattern Analysis and Machine Intelligence 12, pp. 629–639.

Geometry-based Image Segmentation Using Anisotropic Diffusion *

Ross T. Whitaker and Stephen M. Pizer

Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599-3175, USA

Abstract. Segmentations that are based on the geometry of the intensity surface of an image are useful for describing visual features in the image. Some visual features can be characterized as boundaries between geometric patches. Geometric patches are defined as regions in an image where geometric descriptors vary slowly over the support of the region. The boundaries between these patches represent discontinuities in the higher-order geometry of the intensity surface. Reliable and accurate characterizations of local geometry can be difficult. This is because stable measurements of the differential structure of images are obtained only by averaging over a finite aperture. In this work, the scale of differential measurements is established by using an anisotropic diffusion process that generates piecewise constant approximations to geometric descriptions of images. The types of the geometric features as well as the definition of boundaries depend on the types of visual features that one wishes to measure.

Keywords: differential geometry, segmentation, anisotropic diffusion, diffusion, scale, evolutionary systems.

1 Introduction

An image f is a smooth mapping from the image domain $I \subset \mathbb{R}^n$ to some subset of the set of real numbers $L \subset \mathbb{R}$, so that $f : I \mapsto L$; the range of f is called the *intensity* or sometimes *luminance*. The graph of this function is the *intensity surface*. The specification of contiguous regions in the image domain such that the regions depend on the shape properties or geometry of this graph is a *geometry-based image segmentation*. An example of this kind of segmentation is the partitioning of an image into regions that are based on slowly varying luminance or height. The resulting boundaries form luminance "edges" that are often detected as areas of high gradient magnitude. This example relies on the most basic property of the intensity surface – its height. Another example is the segmentation of images on the basis of extremum in height and the flanks or hillsides associated with these "peaks" and "pits". Such segmentations have been created

* This research has been supported by NIH grant #P01 CA47982.

on the basis of gradient directions and watersheds [8]. Second-order properties such as the mean curvature of the intensity surface or the second derivative in the gradient direction can also serve as the basis for such segmentations.

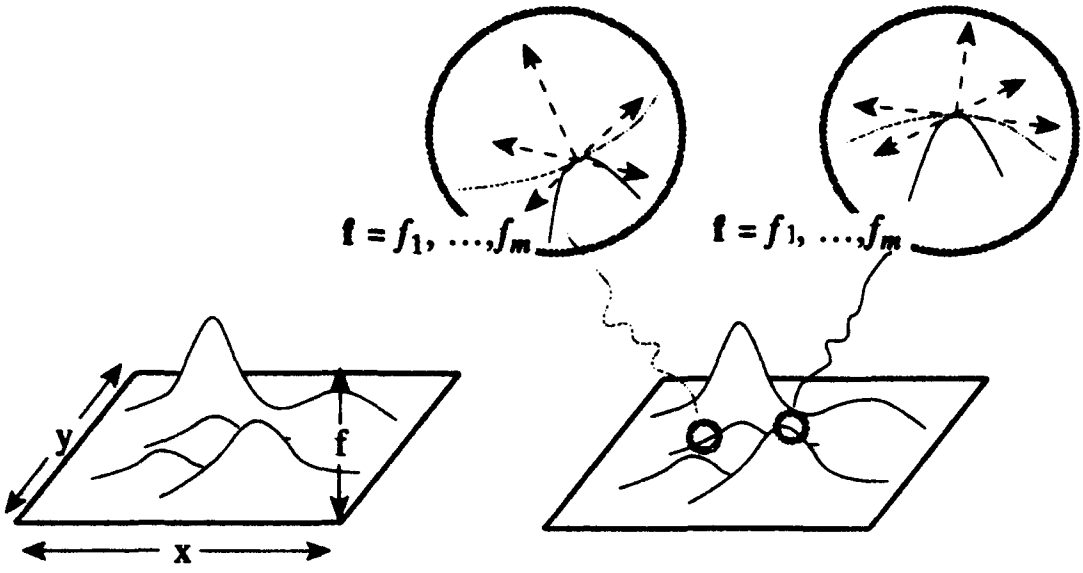


Fig. 1. Local shape of the image intensity surface represented by a finite set of local measurements made at every point in the image domain.

This research proposes a class of segmentation algorithms which group neighbourhoods in images into regions based on the homogeneity or similarity of the *local shape* within those neighbourhoods. The local shape at a point in the image is the differential geometry of the intensity surface about that point. In the "edges" example mentioned previously, local shape has implicitly been reduced to the height of this intensity surface – a zero-order property. In that case measuring the similarity among points consists of measuring the intensity at various places in the image and then comparing these values in local neighbourhoods. For higher-order properties of local shape the notion of similarity between the shape of two or more surface patches is ambiguous because there are many degrees of freedom which can be compared. This indicates that two important decisions must be made when constructing segmentations on the basis of geometry. The first decision is the *order* of the local geometry that is to be measured and compared. Representations of local surface structure must be finite and so comparisons are made between approximating functions to the intensity surface at each point in the image. The second decision is the nature of the comparison, or the *metric*, that is used in determining similarity. Such a metric quantifies the differences in shape between two surface patches. For example, suppose one decides to use second order information to describe local surface structure. Each point in the image is represented as a quadratic. Now define a metric to compare quadratics – there are a variety of options.

Regions are places in an image where variations in geometry are small within local neighbourhoods and boundaries are the loci of places where variations in geometry are large. Deciding how large variations should be in order to constitute a boundary can be difficult. In this paper we do not address this question directly but note that there are several widely used methods for making this decision in a discrete fashion if, indeed, a discrete segmentation is the goal. One option is to quantify the variation in local shape and then threshold this measure. Points where geometry changes at rates above the threshold are defined as boundaries. Another option is to specify boundaries as critical points in the rate of change of local geometry, as does the Canny edge operator [1].

In order to quantify differences in local shape, represent shape as a position in a finite-dimensional *feature space*. A feature space, as used in statistical pattern recognition [3], is the space consisting of a number of *features* or values that characterize points in an image. Typically, classifications of points in the image are made on the basis of proximity in the feature space. A set of geometric measurements of the intensity surface define a feature space as well as *geometric description* of the local surface shape. The similarity of shape between two points in the image is inversely related to the distance between the two corresponding points in this feature space. *Dissimilarity* is the amount of variation of geometric descriptions that exists within a neighborhood of a point in I. Thus, boundaries are regions of high dissimilarity.

2 Geometry-limited Diffusion

2.1 Describing Local Geometry

The local shape of the intensity surface at every point in the image is represented by a finite set of scalar values. These values constitute a set of geometric descriptors and an associated feature space. There are a variety of possibilities for such representations, but for this work we use sets of derivatives of intensity surface measured in Cartesian coordinates. We choose this representation for two reasons. First, derivatives of the intensity surface can be measured directly from a digital image using linear filters. Second, in analysing the behaviour of this system it is useful to be able to rely on the linearity of these operators and the orthogonality of these measurements. A set of derivatives on the original image create a multi-valued function $\mathbf{f} : I \mapsto F$ where $\mathbf{f} = f_1, \ldots, f_m$ designates this function as multi-valued and $F \subset \mathbb{R}^m$ is the feature space.

2.2 Scale and Anisotropic Diffusion

Stable differential measurements of discretely sampled (digital) images are obtained only through some neighborhood operator or kernel. Requiring the appropriate symmetries indicates that such kernels should be Gaussian blobs or derivatives of Gaussians. Koenderink [4] [illegible] that the use of Gaussian kernels as test functions introduces a free parameter, scale, into the measurements that

result from these kernels. Derivatives of sampled data are not calculated analytically but are measured by convolution with the appropriate kernel or test function. The scale of a particular measurement should be appropriate for the data and the task.

The convolution of the image with a test function is equivalent to blurring the original image before taking the measurement. Gaussian kernels of size s are the fundamental solutions to the diffusion equation also called the heat equation, where scale replaces the time or evolution parameter t.

$$c\nabla \cdot \nabla f = \frac{\partial f}{\partial t} \text{ and } t = \frac{s^2}{2}. \tag{1}$$

This blurring is essentially a low-pass filtering where the bandwidth of the filter is determined by the choice of scale in the Gaussian kernel. Indeed, the signal-to-noise ratio of measurements made with Gaussians can improve with increasing scale [7]. However, low pass filters such as the Gaussian can have adverse effects on the characterization of objects or features whose shapes depend on high frequency information.

Nonuniform diffusion has been proposed [6, 5] as an alternative to the uniform scale space as described by the diffusion equation. More specifically, edge-affected diffusion incorporates a variable conductance term which limits the flow of intensity according to local gradient information. Solutions to this equation have been shown to reduce uncorrelated noise while preserving, and even enhancing, edges. Several authors have proposed edge-affected diffusion in the form of the equation,

$$\nabla \cdot g(|\nabla f|)\nabla f = \frac{\partial f}{\partial t} \ , \tag{2}$$

where g, the conductance modulating term, is a bounded, positive, decreasing function of $|\nabla f|$. The solutions of this equation can be thought of as a kind of anisotropic or nonuniform scale space. When the original image is used as the initial condition, this process produces, over time, a set of smooth regions with nearly constant luminance, separated by step edges.

It has been shown that this process bears a strong mathematical resemblance to a class of relaxation algorithms which seek piecewise constant approximations to images [5]. Although this paper discusses primarily nonuniform diffusion equations as in (2), this is done with the understanding that this framework has similar implications for a broad range of regularization problems which are solved by means of a constrained smoothing process.

2.3 Scale Within Diffusion

Earlier work [9] has shown that uniform scale can be incorporated into the gradient measurement that is included in the conductance term of (2). This scale can account for the unreliability of local gradient measurements in the presence of correlated and uncorrelated noise. This approach is consistent with the notion that all image measurements have an associated scale, and that the appropriate scale or scales may depend on the data and the task at hand. Furthermore, by

decreasing the scale at which the gradient is measured over time, one can obtain boundary information that reflects both small-scale and large-scale gradient information. This results in the multi-scale anisotropic diffusion equation,

$$\nabla \cdot g(|\nabla G(s) \circ f|)\nabla f = \frac{\partial f}{\partial t} \ , \tag{3}$$

in which "$G(s)\circ$" denotes convolution with a Gaussian kernel of a particular size $s(t)$, which is itself a function of the time parameter and generally decreases as the process evolves. This process describes a continuous "trade-off" between the isotropic and anisotropic scale, where the scale of isotropic measurements used in the conductance term decreases as the image undergoes progressively more anisotropic blurring.

2.4 Multi-valued Diffusion

In this section, the anisotropic diffusion of (2) is generalized in order to incorporate representations of surface shape. Instead of a single intensity image, diffuse a multi-valued image, which is a smooth mapping from the image domain to the feature space. The diffusion process introduces a time or evolution parameter, $t \in T \subset \mathbb{R}^+$, into the function $\mathbf{f}: I \times T \mapsto F$ so that there is a multi-valued function at each point in time or each level of processing. The multi-valued diffusion equation is

$$\nabla \cdot g(\mathcal{D}(G(s) \circ \mathbf{f}))D\mathbf{f} = \frac{\partial \mathbf{f}}{\partial t} \ , \tag{4}$$

where $\mathcal{D}: F \mapsto \mathbb{R}^+$ is a dissimilarity operator. The composition $\mathcal{D}\mathbf{f} : I \mapsto \mathbb{R}^+$ assigns a degree of dissimilarity to every point in the image. The convolution $G(s) \circ \mathbf{f}$ incorporates the notion (from (3)) of time varying, uniform scale. The derivative $D\mathbf{f}$ of $\mathbf{f}$ is in the form of a matrix, also called the Jacobian. The conductance g is a scalar, and the operator "$\nabla\cdot$" is a vector that is applied to the matrix gDf using the standard convention of matrix multiplication. Equation (4) is a system of separate single-valued diffusion processes, evolving simultaneously, and sharing a common conductance modulating term. The boundaries are not defined on any one image, but are shared among (and possibly dependent on) all the images in the system.

2.5 Diffusion in a Feature Space

The behaviour of the system (4) is clearly dependent on the choice of the dissimilarity operator $\mathcal{D}$. In the single feature case, $m = 1$, the gradient magnitude proves to be a useful measure. That is, $\mathcal{D}f \equiv (\nabla f \cdot \nabla f)^{\frac{1}{2}}$. For higher dimensions, evaluate dissimilarity based on distances in F, the feature space. The dissimilarity operator is constructed to capture the manner in which neighbourhoods in I map into F. At a point $x_0 \in I$, the dissimilarity measures the *density* of space in the neighborhood of x_0 after it is mapped to F. If the resulting space is dense, it will indicate that the neighborhood of x_0 has low dissimilarity. We

propose a dissimilarity that is the Frobenius (root sum of squares) norm of the Jacobian. If J is the Jacobian of $\mathbf{f}$, with elements J_{ij}, then the dissimilarity is

$$\mathcal{D}\mathbf{f} \equiv \|J\| = \left(\sum_{j=1}^{n} \sum_{i=1}^{m} {J_{ij}}^2 \right)^{\frac{1}{2}} . \tag{5}$$

This norm has several advantages over alternatives. First, this matrix norm is induced from the Euclidean vector norm so that in the case of $m = 1$, this expression is the gradient magnitude, as in edge-affected diffusion. Second, the square of this norm (as it often appears in the conductance functions [5]) is differentiable. Finally, this norm when applied to geometric objects (tensors) always produces a scalar which is invariant to orthogonal coordinate transformations.

The dissimilarity can be generalized to account for local transformations in the feature space. If one considers coordinate transformations (rotations and rescaling of axis) to be changes in the relative importance of features in the calculation of distance, then the dissimilarity measure could allow the relative importance of various features to vary depending on the position in the feature space. If $\Phi(y)$, for $y \in \mathbb{R}^m$, is the local coordinate transformation, then the dissimilarity operator is generalized:

$$\mathcal{D}_{\Phi}\mathbf{f} \equiv \|\Phi(\mathbf{f})D\mathbf{f}\| . \tag{6}$$

The local metric $\Phi(y)$ provides a flexibility in defining distances in the feature space. Each position in the feature space represents a different shape as represented by the local Taylor series expansion. Thus, $\Phi(y)$ is precisely the shape metric that is described in Sect. 1. For feature spaces that consist of geometric features, this metric is essential for comparing incommensurate quantities associated with axes of these spaces. It is also necessary in order to construct dissimilarity measures that are invariant to one's initial choice of spatial coordinates.

2.6 Diffusion of Geometric Features

The differential measurements that comprise the Taylor series description of local surface shape form feature vectors that become the initial conditions of a multivalued diffusion equation. In the feature space of geometric features, each point corresponds to a different shape, and the metric Φ quantifies these differences in shape. One should choose Φ so that it creates boundaries that are of interest and choose the scale $s(t)$ in the diffusion equation so that unwanted fluctuations in the feature measurements (noise) are eliminated.

As anisotropic scale increases, measurements can be taken from the resulting set of features can be made with progressively smaller Gaussian kernels, and decisions about presence of specific geometric properties can be made on the basis of those measurements. This strategy is depicted in Fig. 2. The assertion is that information about $(N+1)$th-order derivatives can be obtained in three steps. First, make differential measurements of the image up to Nth-order using

derivative-of-Gaussian filters. Second, regularize these derivatives in a way that breaks these "images" into patches that are nearly piecewise constant. Third, study the boundaries of these patches and make decisions about the presence of higher-order features which include $(N+1)$th-order derivatives.

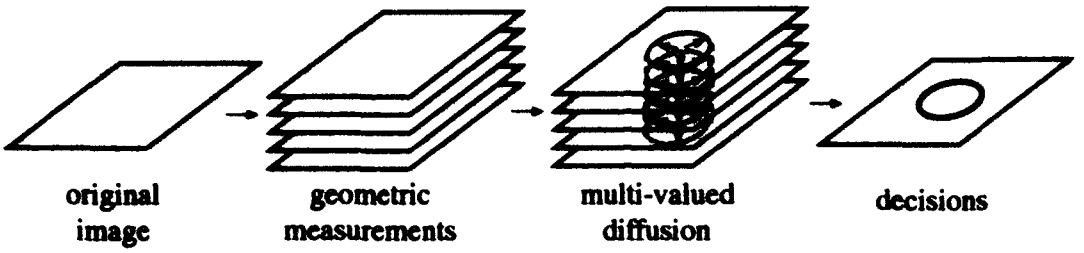


Fig. 2. Geometry-limited diffusion in which features are geometric measurements made on a single valued image.

2.7 An Example: First-order Geometry and Ridges

Such a system has been constructed in order to locate "ridges" or "creases" in digital images. The definition of ridges in two dimensions is described by Gauch [2] as loci of extremum of curvature in the level curves of the intensity surface. Level set curvature is a second-order property that describes the rate of change of direction of tangents to the level curves. Notice that the height of the intensity surface does not enter into this definition, so the local image geometry is encoded as two features

$$f_1(x,y,t) \equiv f^x(x,y,t) \text{ and } f_2(x,y,t) \equiv f^y(x,y,t) \ ,$$

where the raised indices are *labels* which denote that these features undergo nonlinear diffusion but have the image derivatives as their initial conditions. That is,

$$f^x(x,y,0) = \frac{\partial f(x,y)}{\partial x} \text{ and } f^y(x,y,0) = \frac{\partial f(x,y)}{\partial y} \ .$$

Effectively each local surface patch is modelled as a plane of a particular orientation which is specified by these two features. These two measurements comprise a two-dimensional feature space.

Define a dissimilarity operator which captures only changes in the direction of the vector created by this pair of features. This is done by using the local metric

$$\Phi = \begin{bmatrix} \frac{f^x}{(f^{x2}+f^{y2})} & \frac{-f^y}{(f^{x2}+f^{y2})} \\ 0 & 0 \end{bmatrix} \ , \tag{7}$$

or by computing the Jacobian of a function $\mathbf{g}$ that is normalized with respect to the Euclidean length $\mathbf{f}$, i.e. $\mathcal{D}\mathbf{f} \equiv \|D\mathbf{g}\|$ and $\mathbf{g} = \frac{\mathbf{f}}{|\mathbf{f}|}$. This normalization first maps all the points in F to the unit circle, making explicit the fact that

this distance measure, or dissimilarity, is invariant to any monotonic intensity transformation.

Experiments [10] show that this process forms sets of homogeneous regions which have similar gradient directions and which are the "flanks" or "hillsides" in the intensity surface of the original image. The boundaries of these regions appear to correspond well with the ridges in the original image. Because of the tendency of this process to produce piecewise constant solutions, the boundaries are distinct and allow for easy detection with one of the techniques already mentioned. The results of this process are robust with respect to noise, and the use of scale in measuring dissimilarity allows us to control the size of these patches without distorting the shapes of the boundaries between patches.

2.8 Time Behaviour of Invariants

We wish to construct a process which produces results that are independent of the coordinate system that is used to describe the image domain. In particular, the results should not be influenced by translations and rotations of the original image. The encoding of local geometry in terms of coefficients of Taylor series forces the choice of a coordinate system. The absolute position of points in the feature space will depend on this choice because the axes of the feature space are sets of directional derivatives that are aligned with the coordinate axis in the image domain. This presents a serious question. How can one construct a system which produces invariant results but which relies on a feature space that is intimately tied to the choice of coordinate systems?

The answer lies in the dissimilarity measure, which defines distance in the feature space. Positions in the feature space need not be invariant to orthogonal transformations in order to obtain invariant results. It *is* essential, however, that the *relative* positions of points and the distances between those points be invariant. To ensure this invariance, express the dissimilarity as a geometric invariant of the intensity surface of the original image. Given this, the dissimilarity will be invariant at the start of the process because it is a geometric invariant of the intensity surface. As the process progresses, however, the terms of the invariant will change so that they no longer resemble derivatives of the intensity surface. Will dissimilarity remain invariant as the nonlinear, nonuniform diffusion progresses? The following proposition shows that geometric invariants constructed from the original geometric features (derivatives of the intensity surface) remain invariant throughout the diffusion process. It concerns polynomial invariants as described by ter Haar Romeny and Florack [7], where polynomials are expressed using the Einstein summation convention of summing expressions with like indices over the basis. The lowered indices indicate a derivative in the direction associated with that axis. For example, the square of the gradient magnitude of a two-dimensional function $h(x, y)$ is expressed as

$$h_i h_i = \sum_{i \in x,y} h_i h_i = {h_x}^2 + {h_y}^2 = (\frac{\partial h}{\partial x})^2 + (\frac{\partial h}{\partial y})^2 \ , \tag{8}$$

and the Laplacian as

$$h_{ii} = \sum_{i \in x,y} h_{ii} = h_{xx} + h_{yy} = \frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} \cdot \tag{9}$$

Any expression of this form with the property that all indices are paired (no free indices) is a scalar that is invariant to orthogonal coordinate transformations (rotations and translations), and is thereby independent of one's choice of coordinates.

Proposition 1. *Given a geometry-limited diffusion system as described in Sect. 2.6 with:*

1. *a set of features* $\mathbf{f} = f_1, \ldots, f_m$ *with initial values that are derivatives of a smooth image,*
2. *a dissimilarity measure* $\mathcal{D}\mathbf{f}$ *that is an invariant expression of those features and their first-order derivatives, and*
3. *solutions which are analytic in time,*

then any function P *that depends on these features and has an initial value* $(t = 0)$ *that can be expressed as a polynomial invariant of the intensity surface is invariant for all* $t \geq 0$.

Proof. The proof relies on the analytic nature of the solutions to express the invariant as a Taylor series,

$$P(\mathbf{f}, t) = P(\mathbf{f}, 0) + \frac{\partial}{\partial t} P(\mathbf{f}, t)_{t=0} + \frac{1}{2} \frac{\partial^2}{\partial t^2} P(\mathbf{f}, t)_{t=0} + \ldots . \tag{10}$$

Using the chain rule and the diffusion equation, express each time derivative of the invariant in terms of spatial derivatives of the initial values $(t = 0)$ of the features,

$$\frac{\partial}{\partial t} P(\mathbf{f}, t) = \sum_{i=1}^{m} \frac{\partial}{\partial f_i} P(\mathbf{f}, t) \frac{\partial f_i}{\partial t} = \sum_{i=1}^{m} \frac{\partial}{\partial f_i} P(\mathbf{f}, t) \nabla \cdot g(D(\mathbf{f})) \nabla f_i \tag{11}$$

Inductively each term in this series in (10) is an invariant. Consider the kth term $(\partial^k / \partial t^k) P(\mathbf{f}, t)$ from (10) and assume (inductive assumption) that it is invariant and is also a function of the features and higher-order derivatives of those features. Then the time derivative $(\partial^{k+1} / \partial t^{k+1}) P(\mathbf{f}, t)$ must also be invariant. This is true because the time derivative of any of these terms can be converted as in (11) to a spatial derivative using the multi-value diffusion equation. Alternatively, expressing the invariant in terms of the initial values of the features allows the raised indices, which are labels, to be written as lowered indices, which indicate derivatives as used in the Einstein notation. With an invariant conductance term, the operator $\nabla \cdot g(D(\mathbf{f}))\nabla$ introduces indices into expressions only in pairs, thus maintaining the Einstein summation convention. The first term of the Taylor expansion, $P(\mathbf{f}, 0)$, is invariant by assumption because it is a polynomial invariant consisting of derivatives of the intensity surface. Inductively, the entire series of (10) is invariant, and by analyticity so is $P(\mathbf{f}, t)$. □

This result is important because it states that throughout the diffusion process the dissimilarity measure, as well as any other function of polynomial invariants, remains insensitive to the original choice of coordinates. The raised indices, which indicate the initial values of the features, are treated as derivatives when constructing differential invariants. For example, in the first-order process of Sect. 2.7, both the gradient magnitude $(f^i f^i)^{\frac{1}{2}} = ((f^x)^2 + (f^z)^2)^{\frac{1}{2}}$ and the Laplacian $f^i{}_i = f^z{}_z + f^y{}_y$ are invariant to orthogonal transformations for all $t \geq 0$.

3 Conclusions

Characterizing images in terms of the differential structure of the intensity surface allows one to examine the way shape varies across the image domain and to create segmentations on the basis of local shape. Regions in the image where local shape is homogeneous can be grouped on the basis of geometry. This approach requires a model of local shape that captures relevant properties with a finite representation as well as a definition of "distance" that allows one to quantify the similarity of two shapes. Anisotropic scale can be used as a means of localizing patch boundaries in a manner that appears to be accurate and robust. The geometry-limited diffusion process produces homogeneous patches with relatively well-defined boundaries while maintaining the geometric invariance of the measurements that are used to detect and characterize these boundaries.

References

1. Canny, J. (1987). A computational approach to edge detection, IEEE Trans. Pattern Analysis Machine Intelligence 8 (6), pp. 679–698.
2. Gauch, J.M. (1989). A multiresolution intensity axis of symmetry and its application to image segmentation, Report TR89-047, Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina.
3. Duda, R.O., Hart, P.E. (1973). Pattern Classification and Scene Analysis, John Wiley and Sons, New York.
4. Koenderink, J.J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
5. Nordstrom, N. (1990). Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection, Image and Visual Comp. 8 (4), pp. 318–327.
6. Perona, P., Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Analysis Machine Intelligence 12, pp. 429–439.
7. Romeny, T.H., Florack, L. (1991). A multiscale geometric approach to human vision. In: Hendee, B., Wells, P.N.T. (eds.), Perception of Visual Information, Springer Verlag, Berlin.
8. Rosin, P.L., Colchester, A.C.F., Hawkes, D.J. (1992). Early image representation using regions defined by maximum gradient profiles between singular points, Pattern Recognition, 25 (7), pp. 695–711.
9. Whitaker, R.T., Pizer, S.M. (1993). A multi-scale approach to nonuniform diffusion, CVGIP: Image Understanding 57 (1), pp. 99–110.
10. Whitaker, R.T. (1993). Geometry-limited Diffusion in the characterization of geometric patches in images, CVGIP: Image Understanding 57 (1), pp. 111–120.

Images: Regular Tempered Distributions *

Luc M. J. Florack[1], *Bart M. ter Haar Romeny*[1], *Jan J. Koenderink*[2], *and Max A. Viergever*[1]

[1] Computer Vision Research Group, University Hospital, Room E.02.222, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
[2] Department of Medical and Physiological Physics, University of Utrecht, Princetonplein 5, 3584 CC Utrecht, The Netherlands

Abstract. In this paper a framework is proposed for representing local image structure in an operationally well-defined and well-posed way. Its mathematical basis is the theory of *regular tempered distributions*. Under suitable physical constraints this theory turns out to be equivalent to scale-space theory for greylevel images.

Two formally equivalent representations of scale-space are presented. Apart from the familiar representation, which is based on a fixed-scale spatial integration using a Gaussian measure, an alternative representation is proposed that relies on a coarse-to-fine scale-integration. The potential functionality of this is explained in the context of active vision.

Keywords: local jet, metamerism, regular tempered distribution, scale-space, well-posed differentiation.

1 Introduction

The process of differentiation as defined in the standard way is known to be ill-posed in the sense of Hadamard. This means that differentiation, seen as a linear transformation in a Hilbert space of functions, is *discontinuous* (it is even unbounded). Consequently, conventional differential operators require a non-trivial modification in order to be of any practical use as a tool for describing local image structure. At this point it is important to stress that ill-posedness is indeed a problem inherent to the operators and *not* to the type of operands, that is images. It is still not commonly appreciated that "smoothing" prior to differentiation essentially does not remove the ill-posedness problem.

The ill-posedness problem was formally solved several decades ago in the theory of *tempered distributions* [8]. This theory allows for a significant relaxation of a priori constraints on the class of objects that are subject to differentiation. By basing the process of differentiation on an *integration*, rather than on an

* This work was performed as part of the 3D Computer Vision Research Program, supported by the Netherlands Ministry of Economic Affairs through a SPIN grant, and by the companies Agfa-Gevaert, Philips Medical Systems, and KEMA.

infinitesimal procedure, well-posedness is built in right from the start. In fact one may argue that images, by their very physical nature, are naturally modelled as *regular tempered distributions*, that is special kinds of *functionals*—as opposed to *functions*—related to an equivalence class of functions.

Whereas a function in the conventional sense has a structure "of its own", the structure of a distribution is defined in an operational way through its action on a dual space of *test functions*, called a *Schwartz space* (explained later in this paper). A regular tempered distribution associated with a real image is essentially a bilinear, symmetric form, or a scalar product. The "raw signal" serves to define its first argument, whereas its second argument can be any member of some physical Schwartz space. The abstract Schwartz space thus acquires a very vivid meaning as a set of smooth physical apertures or linear template kernels. This admits a well-posed differentiation scheme based on its well-defined adjoint operation on the Schwartz space. Clearly, this is the *only* way to make sense of "the differential structure of an image", which is a physical and mathematical non-entity in the conventional sense.

In this paper some of the basic results from the theory of regular tempered distributions will be reviewed and their relation to images will be pointed out. With some physical constraints on the admissible Schwartz functions, this will give us a set of well-posed *scaled differential operators*, apt for the description of local image structure up to any given order. An alternative representation of the fixed-scale differential operators will be proposed that relies on a scale-integration, yielding an alternative method for obtaining local image structure in a potentially even more robust way. Such an approach seems quite useful, in particular for an autonomous, active vision system. For such a system, a coarse-to-fine approach is likely to be needed for *accessing* or *reading* the scale-space data representation by high-level visual routines. This is particularly the case for a high-resolution system which is capable of sampling many more data simultaneously than these routines can handle. Permission for *writing* the scale-space data representation is naturally granted to the visual environment, not to the observing system. In other words, the distribution as such (that is the fixed first argument of the associated bilinear form) is set by the environment (as it should), and the visual system provides the dual space of test functions to extract measurements from it (*receptive fields*). The nature of the Schwartz space accounts for a *metameric* (many-to-one) mapping of a physical scene to a given data representation.

An overview of several alternative approaches to well-posed differentiation is given in this volume by Foster [3].

2 Regular Tempered Distributions and Scale-Space

Let us consider the class $\mathcal{P}(\mathbb{R}^d)$ of *functions of polynomial growth*. This class of (piecewise continuous, but generally not smooth) functions is sufficient for most physical applications, and more specifically, for most image analysis and computer vision purposes:

Definition 1. A function $g : \mathbb{R}^d \to \mathbb{R}$ is said to be of polynomial growth, notation: $g \in \mathcal{P}(\mathbb{R}^d)$, if there exists $c > 0$ and $m \geq 0$, such that for all $\mathbf{x} \in \mathbb{R}^d$

$$|g(\mathbf{x})| \leq c\left(1 + \|\mathbf{x}\|^2\right)^m .$$

For later convenience, let us also introduce the *multi-index* notation in d dimensions:

Definition 2. A multi-index $\mathbf{n}$ denotes any set of d non-negative integers, while $\|\mathbf{n}\|$ denotes its norm, that is the sum of all these integers:

$$\mathbf{n} = \{n_i\}_{i=1\ldots d} \quad , \quad \|\mathbf{n}\| = \sum_{i=1}^{d} n_i .$$

The multi-index notation will be used to abbreviate $D_{\mathbf{n}} = \partial^{\|\mathbf{n}\|}/\partial x_1^{n_1} \ldots x_d^{n_d}$, $\mathbf{x}^{\mathbf{n}} = x_1^{n_1} \ldots x_d^{n_d}$, etc. Also, $\phi_{\mathbf{n}}$ will be used as an alternative notation for $D_{\mathbf{n}}\phi$. Finally, $d\mathbf{x}$ will be used to denote the usual d-dimensional measure $dx_1 \ldots dx_d$.

For the class of functions of polynomial growth one then proceeds as follows in order to define their derivatives in a well-posed way: for each such function g one introduces a functional T_g, called a *regular tempered distribution*. This functional is defined so as to operate on the class of *smooth test functions* (also called *Schwartz functions*), which will be denoted by $\mathcal{G}(\mathbb{R}^d)$. A linear derivative $D_{\mathbf{n}}g$ of g is then likewise associated with a regular tempered distribution, whose action on a test function ϕ is expressed in terms of the action of T_g on the corresponding derivative $D_{\mathbf{n}}\phi$ of ϕ (which is well-defined). This is how it is defined:

Definition 3. The class of smooth test functions, $\mathcal{G}(\mathbb{R}^d)$, is defined by

$$\phi \in \mathcal{G}(\mathbb{R}^d) \Leftrightarrow \phi \in C^\infty(\mathbb{R}^d) \wedge \sup_{\mathbf{x}\in\mathbb{R}^d} |\mathbf{x}^\beta D_\alpha \phi(\mathbf{x})| < \infty ,$$

for all multi-indices α and β.

Definition 4. A linear functional $T : \mathcal{G}(\mathbb{R}^d) \to \mathbb{R}$ is called a tempered distribution if there exist $c > 0$, and multi-indices α, β, such that

$$|T(\phi)| \leq c \sup_{\mathbf{x}\in\mathbb{R}^d} |\mathbf{x}^\beta D_\alpha \phi(\mathbf{x})| ,$$

for all $\phi \in \mathcal{G}(\mathbb{R}^d)$.

Definition 5. Let $g \in \mathcal{P}(\mathbb{R}^d)$ be a function of polynomial growth, then its associated regular tempered distribution T_g is defined by the tempered distribution

$$T_g : \mathcal{G}(\mathbb{R}^d) \to \mathbb{R} \; : \; \phi \rightarrowtail \int g(\mathbf{x})\phi(\mathbf{x})d\mathbf{x} .$$

One may easily verify that a regular tempered distribution as defined by Definition 5 is indeed a tempered distribution according to Definition 4. The space of all tempered distributions is a linear space over $\mathbb{R}$ (with the usual definition of addition and scalar multiplication) and is denoted here by $\tilde{\mathcal{G}}(\mathbb{R}^d)$. Often, and for obvious reasons, a regular tempered distribution T_g is identified with the function g, and in this sense it is straightforward to define the derivative of g by identifying it with the derivative of its corresponding regular tempered distribution T_g:

Definition 6. The derivative $D_{\mathbf{n}}T$ of a tempered distribution T is defined by

$$(D_{\mathbf{n}}T)(\phi) \equiv (-1)^{\|\mathbf{n}\|}T(D_{\mathbf{n}}\phi) \ .$$

The conventional $-$ sign reflects the antihermitean nature of differentiation. Note the metameric nature of g alluded to in the introduction: since in practice one only has access to the values of $T_g(\phi)$ for some finite subspace of $\mathcal{G}(\mathbb{R}^d)$, $\bar{g}$ and g define metamers whenever $T_{\bar{g}}(\phi) = T_g(\phi)$ for all ϕ in that subspace.

For the purpose of removing the ill-posedness of differentiation, the above solution is completely to the satisfaction of the mathematician. In particular, there is no reason for constraining the space of smooth functions from this point of view. Physical considerations are required to *interpret* and to *constrain* the function space $\mathcal{G}(\mathbb{R}^d)$.

As far as the interpretation is concerned, the functionality of the Schwartz space $\mathcal{G}(\mathbb{R}^d)$ as a general bank of *linear templates* or, from a physiological point of view, as a stratification of *receptive fields*, has already been discussed. Thus a physical subspace of $\mathcal{G}(\mathbb{R}^d)$ is provided by the vision system as a prewired module of linear detectors (the "front-end"), the weighting profiles of which are matched against the image distribution to produce correlation numbers (several independent physical Schwartz spaces may exist for parallel front-end channels). Since the only way of accessing linear image information is via these correlation numbers, it is indeed quite natural to think of images as regular tempered distributions.

The constraints one would like to impose on $\mathcal{G}(\mathbb{R}^d)$ may of course depend on the application. If one considers the front-end as a general "read-only" data bank for *whatever* higher level image processing routine, one may want to impose constraints of a "universal" nature only. In the absence of a priori knowledge concerning the nature, location, orientation, and scale of image features that might be of interest, it is plausible to impose *translation*, *rotation*, and *scale invariance* [1].

This introduces *at least* two free parameters, $\mathbf{x}$ and σ say, corresponding to the base point and span of the linear aperture profiles ϕ (aptly called *local neighbourhood operators* for this reason). A (fuzzy) spatial neighbourhood of extent σ, centred at base point $\mathbf{x}$, will be indicated by $(\mathbf{x};\sigma)$. The freedom of choosing these parameters accounts for manifest shift and scale invariance. Although one could argue for the introduction of an angular parameter for incorporating manifest rotation invariance in a similar way, this would really be redundant (though for a biological system this kind of "redundancy" is quite natural).

Once one has established a basic physical kernel $\phi(\,.\,) \in \mathcal{G}(\mathbb{R}^d)$, one can scale and shift it so as to obtain a parameterized family of "zeroth order" operators $\phi(\mathbf{x},\,.\,;\sigma) \stackrel{\text{def}}{=} \sigma^{-d}\phi(\sigma^{-1}(\mathbf{x} - \,.\,))$, one for each local neighbourhood $(\mathbf{x};\sigma)$. Then one can consider all partial derivatives of this with respect to some arbitrarily chosen Cartesian frame. This yields a complete, hierarchical family of "scaled differential operators" (this procedure for constructing a physical Schwartz space can be generalized to account for other physical parameters as well: see [2]).

It is thus straightforward to reconcile image differentiation with the theory of regular tempered distributions: simply take the subclass of test functions to be the $\phi(\mathbf{x},\,.\,;\sigma)$ obtained by distributing scaled copies of a basic, physically admissible scaling function $\phi(\,.\,)$ over the entire image domain. If $\psi_0 \in \mathcal{P}(\mathbb{R}^d)$ is a given image and $\psi(\mathbf{x};\sigma) \stackrel{\text{def}}{=} T_{\psi_0}(\phi(\mathbf{x},\,.\,;\sigma))$, that is the smooth function obtained by freezing the scaled and shifted copy of ϕ into the regular tempered distribution associated with ψ_0, then a robust derivative of ψ_0 that depends continuously on ψ_0 is given by $D_{\mathbf{n}}\psi(\mathbf{x};\sigma) = D_{\mathbf{n}}T_{\psi_0}(\phi(\mathbf{x},\,.\,;\sigma))$. Despite the fact that the $D_{\mathbf{n}}$ on both sides of this equation are totally different operators (the one on the left-hand side is the conventional one, acting on smooth functions, whereas the one on the right-hand side is the new one, acting on distributions), the equality guarantees overall notational consistency (the left-hand side notation will be used henceforth). Moreover, if ψ_0 is sufficiently smooth, say $\psi_0 \in C^{\|\mathbf{n}\|}(\mathbb{R}^d)$, one has $D_{\mathbf{n}}T_{\psi_0} = T_{D_{\mathbf{n}}\psi_0}$. Using Definition 6 one obtains

$$D_{\mathbf{n}}\psi(\mathbf{x};\sigma) = (-1)^{\|\mathbf{n}\|}\int D_{\mathbf{n}}\phi(\mathbf{x},\mathbf{y};\sigma)\psi_0(\mathbf{y})d\mathbf{y}\;, \tag{1}$$

(the operator $D_{\mathbf{n}}$ is understood to apply to the dummy argument). In particular, using

$$\phi(\mathbf{x},\mathbf{y};\sigma) = \sigma^{-d}\phi(\sigma^{-1}(\mathbf{x}-\mathbf{y}))\;, \tag{2}$$

one gets

$$D_{\mathbf{n}}\psi(\mathbf{x};\sigma) = \int \sigma^{-d}\phi_{\mathbf{n}}(\sigma^{-1}(\mathbf{x}-\mathbf{y}))\psi_0(\mathbf{y})d\mathbf{y}\;, \tag{3}$$

in which $\phi_{\mathbf{n}}$ means $D_{\mathbf{n}}\phi$.

Note that one can define a unique image derivative only after fixing the inner scale σ on which one wishes to resolve the image's differential structure. The only smooth, self-similar test function ϕ that meets the previously mentioned front-end invariance requirements (translation, rotation, and scale invariance) is the normalized *Gaussian scale-space kernel* g.

Definition 7 Gaussian scale-space kernel. The Gaussian scale-space kernel g is defined by the normalized Gaussian

$$g(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^d}\exp\left(-\frac{1}{2}x^2\right)\;,$$

or, in Fourier representation

$$\tilde{g}(\omega) = \exp\left(-\frac{1}{2}\omega^2\right)\;.$$

Using (3) this immediately suggests a *complete* family of scale-space kernels:

Definition 8 Gaussian family. The Gaussian family is the set of all possible Gaussian derivatives:

$$g_{\mathbf{n}}(\mathbf{x}) = D_{\mathbf{n}} g(\mathbf{x}) ,$$
$$\tilde{g}_{\mathbf{n}}(\omega) = (i\omega)^{\mathbf{n}} \tilde{g}(\omega) .$$

On the basis of this family one may define the local jet for a given image ψ_0 and a given local neighbourhood $(\mathbf{x}; \sigma)$, which concisely captures all local image structure up to a given order:

Definition 9 Local jet. The local jet of order N for a given image $\psi_0 \in \mathcal{P}(\mathbb{R}^d)$ and a given local neighbourhood $(\mathbf{x}; \sigma)$ is the equivalence class of images $\chi_0 \in \mathcal{P}(\mathbb{R}^d)$ with spatial contact of order N at $\mathbf{x}$ on scale σ:

$$J^N[\psi_0](\mathbf{x}; \sigma) = \left\{ \chi_0 \in \mathcal{P}(\mathbb{R}^d) \mid \chi_{\mathbf{n}}(\bar{\mathbf{x}}; \bar{\sigma}) = \psi_{\mathbf{n}}(\mathbf{x}; \sigma) \quad \forall \mathbf{n} \quad \text{with} \quad \|\mathbf{n}\| \leq N \right\} .$$

For an extensive treatment, see [1, 9, 4, 5]. We will henceforth focus on Gaussian scale-space theory.

3 A Coarse-to-Fine Representation of Scale-Space

In Gaussian scale-space low-resolution information results from the fine-to-coarse propagation of higher resolution information along the scale dimension ("blurring"). The interactions that take place in this process are intrinsically irreversible and the density of structural degrees of freedom (and hence the natural sampling rate of local neighbourhood operators) decreases in a strict monotonic way. Since a high-resolution image, with inner scale σ say, is actually a multiresolution image, containing all coarse scale-information for $\hat{\sigma} > \sigma$ as well, one would expect it to be possible to retrieve the σ-details by a coarse-to-fine integration of this "deep structure".

In an active vision system this process may be under the control of an attention and foveation mechanism, allowing the system to roam about in scale-space. The advantage of this strategy is that if the system wants to read out certain interesting details (of a priori unknown scale), represented by the front-end local neighbourhood operators, it can start integrating on a coarse neighbourhood of the locus of attention (where the precise position of the base point of the local neighbourhood operators is not crucial), and decide "on the fly" how to adjust the base point of the integration path by switching to neighbouring points within the current $\hat{\sigma}$-neighbourhood (the path continuously bifurcates with decreasing scale because of the increasing sampling rate). As long as the initial coarse scale neighbourhood has sufficient overlap with the final base point of interest, and as long as the base point variations during the scale-integration process are small enough relative to the current inner scale (which decreases on integration), one may expect a final result rather independent of the continuous base point adjustments, that is close to the result that would have been obtained with a *fixed*

base point, frozen at the (a priori unknown) point of interest. To understand such an autonomous system to its full extent is of course far beyond the present state of the art. Instead let us concentrate on the scale-integration procedure for a fixed base point, showing that it is indeed possible to retrieve any structural image detail by such a coarse-to-fine approach.

Definition 10 Mexican-hat profile. A mexican-hat profile h is defined by the Laplacean of the Gaussian g:

$$h(\mathbf{x}) = \Delta g(\mathbf{x}) \quad \text{or, in Fourier representation:} \quad \tilde{h}(\omega) = -\omega^2 \tilde{g}(\omega) \, .$$

The following proposition shows how to obtain σ-scale details via a coarse-to-fine scale-integration. It involves a scale-ensemble of local neighbourhood operators. Compare this with Definition 9, in which the extraction of each jet component relies on a single local neighbourhood operator only.

Proposition 11. *The local jet* $J^N[\psi_0](\mathbf{x};\sigma) \mathrel{\widehat{=}} \{\psi_{\mathbf{n}}(\mathbf{x};\sigma) \mid \|\mathbf{n}\| \leq N\}$ *can be represented by a concatenation of a "mexican-hat" and a "geometric" front-end stage:*

$$\psi_{\mathbf{n}}(\mathbf{x};\sigma) = N^{-1} \int_0^\infty \int g_{\mathbf{n}} \left(\frac{\mathbf{x}-\mathbf{y}}{\sqrt{\sigma^2+\hat{\sigma}^2}} \right) \int h \left(\frac{\mathbf{y}-\mathbf{z}}{\hat{\sigma}} \right) \psi_0(\mathbf{z}) \frac{d\hat{\sigma}}{\hat{\sigma}} d\mathbf{y} d\mathbf{z} \, ,$$

in which the normalization constant N *is given by:*

$$N = \int_0^\infty \tilde{h}(\rho)\tilde{g}(\rho) \frac{d\rho}{\rho} \, .$$

Proof. Consider the "blurred" image $\chi(\mathbf{x};\sigma) = \int g(\mathbf{x},\mathbf{y};\sigma)\chi_0(\mathbf{y})d\mathbf{y}$, or, by diagonalising the kernel through Fourier transformation, $\tilde{\chi}(\omega;\sigma) = \tilde{g}(\sigma\omega)\tilde{\chi}_0(\omega)$. Multiply this by $\sigma^{-1}\tilde{h}(\sigma\omega)$ and integrate over $\sigma > 0$. This yields: $N\tilde{\chi}_0(\omega) = \int_0^\infty \tilde{h}(\hat{\sigma}\omega)\tilde{\chi}(\omega;\hat{\sigma})\hat{\sigma}^{-1}d\hat{\sigma}$, with N as given in the proposition (and in particular, $N = -1/2$ for our Gaussian-mexican-hat pair (g,h)). If one now replaces $\tilde{\chi}_0(\omega)$ by $(i\omega)^{\mathbf{n}}\tilde{\psi}(\omega;\sigma)$, and ergo $\tilde{\chi}(\omega;\hat{\sigma})$ by $(i\omega)^{\mathbf{n}}\tilde{g}((\sigma^2+\hat{\sigma}^2)^{\frac{1}{2}}\omega)\tilde{\psi}_0(\omega)$, then one ends up with $N(i\omega)^{\mathbf{n}}\tilde{\psi}(\omega;\sigma) = \int_0^\infty \tilde{h}(\hat{\sigma}\omega)\tilde{g}_{\mathbf{n}}((\sigma^2+\hat{\sigma}^2)^{\frac{1}{2}}\omega)\tilde{\psi}_0(\omega)\hat{\sigma}^{-1}d\hat{\sigma}$. Fourier inversion then completes the proof. □

Proposition 11 shows a two-stage front-end, consisting of a Laplacean preprocessing stage, comprising a dense set of mexican-hat profiles of various sizes, and a "geometric" stage, based on the Gaussian family (at this stage one finds orientation sensitive filters). A possible interpretation of the different functionalities of these stages is that the Laplacean preprocessor is essentially the *scale-integration measure* for the second stage (it yields the derivative with respect to scale), thus encoding the image's differential structure along the scale dimension, whereas the Gaussian family captures all differential structure on each fixed-scale spatial "slice".

The normalization constant N depends on the choice of the filter h: any filter is allowed for which N is well-defined and non-zero. The lowest order linear filter that is invariant under Cartesian coordinate transformations and meets

the requirement is the mexican-hat profile, for which $N = -1/2$. Other, similar operators are at least of fourth order (such as the Gaussian biharmonic $\Delta^2 g$).

Note that the scale-integration measure $\sigma^{-1} d\sigma$ equals $d\tau$ if one parameterizes scale as $\sigma(\tau) = \sigma_0 \exp(\tau)$, which is the natural parameterization for a self-similar sampling. Note also that, because the base point $\mathbf{x}$ is fixed, the effective scale-integration domain is the interior of a "fuzzy" hypercone $0 < \|\mathbf{x} - \mathbf{y}\| < \lambda\hat{\sigma}$, $0 < \hat{\sigma} < \infty$ (a cone in the conventional sense for 2-dimensional images), where $\lambda > 0$ is some scale-invariant fiducial constant that controls the effective neighbourhood of $\mathbf{x}$ taken into account in an approximating finite-volume integration; the approximation rapidly converges to the exact result of Proposition 11 in the limit $\lambda \to \infty$. The hypercone can also be clipped to some physically sensible minimum and maximum scales σ_- and σ_+, the exterior of which contributes only by a negligible amount.

In a foveal vision system, one can always arrange things such that $\mathbf{x}$ corresponds to the foveal centre (by introducing an extra shift degree of freedom for foveation). For a self-similar system with a finite read-out capacity one can stack the receptive fields of various scales in such a way that the σ-cross-sections through the hypercone are all scaled copies of each other, with a fixed density and relative overlap of receptive fields (once again: scale-invariance!). Regardless of scale-sampling, one thus obtains a vision system that is characterized by a *linear decrease of resolution as a function of eccentricity* (the *smallest* receptive fields at a given base point $\mathbf{y}$ correspond to the boundary of the hypercone). Such a system naturally emerges from optimising the trade-off between its finite read-out capacity and high-resolution requirements. It is implicit in Proposition 11. See [7] for further details.

Despite the fact that the first stage entails a second-order differentiation of the image data on various scales, zeroth and first-order *local* information is *not* filtered out. Only the *global* zeroth- and first-order image averages are lost. In other words, all images $\psi_0(\mathbf{x}) + \alpha + \beta \cdot \mathbf{x}$ for arbitrary α and β are mapped to the same representation. The image in Proposition 11 should be regarded as a representative member of this equivalence class.

4 Conclusion and Discussion

The intimate relation of scale-space theory and the theory of regular tempered distributions has been explained in detail. The physically constrained Schwartz space that forms the domain of definition for the regular tempered distribution associated with a given physical image is made up of translated and scaled copies of a single scaling function, the self-similar normalized Gaussian. It has been pointed out that this is a rather minimal choice: generalized Schwartz spaces are conceivable by tuning the basic Gaussian to particular observables. Such Schwartz spaces carry additional parameters, such as velocity or disparity parameters.

The case of time-varying images has not been addressed here. The requirement of temporal causality introduces a nontrivial complication for the con-

struction of a physically admissible Schwartz space. A possible solution to this problem has been proposed by Koenderink; see [6].

Whereas the usual representation of scale-space relies on a fixed-scale spatial integration for each level of scale, an alternative one has been proposed involving a coarse-to-fine scale-integration in addition. In this representation any given local image property on a given scale can be obtained by an integration over a fuzzy (hyper)cone in scale-space centred at the given base point and clipped at the bottom on the scale level of interest. The relevance of such a representation has been pointed out for an active vision system. A physical realization of such a system naturally yields a foveal mechanism characterized by a linear decrease of resolution as a function of eccentricity.

References

1. Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J., Viergever, M. A. (1992). Scale and the differential structure of images, Image and Vision Computing, Vol. 10, pp. 376–388. Special Issue: Information Processing in Medical Imaging. Also: 3DCV Technical Report no. 91-30.
2. Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J., Viergever, M. A. (1992). Families of tuned scale-space kernels, Proc. European Conf. on Computer Vision, (Santa Margherita Ligure, Italy), pp. 19–23. Also: 3DCV Technical Report 91-21.
3. Foster, D. H. (1993). Classical and fuzzy differential methods in shape analysis, this volume, pp. 319–332.
4. Koenderink, J. J. (1984). The structure of images, Biol. Cybern., Vol. 50, pp. 363–370.
5. Koenderink, J. J., van Doorn, A. J. (1990). Receptive field families, Biol. Cybern., Vol. 63, pp. 291–298.
6. Koenderink, J. J. (1988). Scale-time, Biol. Cybern., Vol. 58, pp. 159–162.
7. Lindeberg, T., Florack, L. M. J. (1992). In preparation.
8. Schwartz, L. (1950–1951). Théorie des Distributions, Vol. I, II of Actualités scientifiques et industrielles; 1091,1122. Paris: Publications de l'Institut de Mathématique de l'Université de Strasbourg.
9. Witkin, A. (1983). Scale space filtering, Proc. Int. Joint Conf. on Artificial Intelligence, (Karlsruhe, W. Germany), pp. 1019–1023.

Local and Multilocal Scale-Space Description *

Alfons H. Salden, Bart M. ter Haar Romeny, and Max A. Viergever

3D Computer Vision Research Group, University Hospital, Room E.02.222, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.

Abstract. A new method is presented for solving equivalence problems for the extended jet of finite order of the scale-space corresponding to a 2-dimensional input image and the groups of spatially homogeneous affine and orthogonal transformations of local cartesian coordinate frames. By means of this method complete and irreducible sets of algebraic invariants are found that may describe any local and (or) multilocal affine or orthogonal invariant of scale-space. Consequently these sets may form bases for topological descriptions of scale-space.

Keywords: shape description, scale-space, local and (or) multilocal, algebraic invariant.

1 Introduction

The differential structure of an input image may operationally be well defined by the extended jet of its scale-space [1, 2]. This structure is acquired by convolution of the input image with fuzzy-derivative operators, i.e. physical representatives of the partial derivative operators. The jet in turn forms the basis for equivalence problems: when are two given differential geometric or algebraic objects of the jet the same if we allow a specific group of transformations of the variables? Thus, which equivalence relations or invariants characterize the object as far as the equivalence problem is concerned? In order to solve equivalence problems for a jet and some transformation group, Cartan's method may be applied [3]. This method leads to a complete and irreducible description of scale-space in terms of a countably infinite number of local invariants [4, 5]. But Cartan's method presupposes a specific definition of a connection on the basis of the jet. What should be done if it is impossible to set up such a connection? Furthermore, Cartan's method restricts itself to a purely local description of geometric structures that are implicitly defined on that jet. But structures are also characterized by multilocal equivalence relations. What about a complete and irreducible set of

* This work was performed as part of the 3D Computer Vision Research Program, supported by the Dutch Ministry of Economic Affairs through a SPIN grant, and by the companies Agfa-Gevaert, Philips Medical Systems and KEMA.

(multi)local invariants of scale-space under a transformation group? Equivalence problems of this kind have only been sparsely addressed [14].

In Sect. 2 related equivalence problems are defined for the extended jet of finite order corresponding to a 2-dimensional input image and for the groups of spatially homogeneous affine and orthogonal transformations of local cartesian coordinate frames. A formal solution to these problems is presented in Sect. 3 by applying Hilbert's method and using Klein's adjunction theorem [6, 7]. The equivalence relations underlying those problems then define complete and irreducible sets of algebraic invariants that are necessary and sufficient to describe any (multi)local affine or orthogonal invariant of scale-space. In Sect. 4 this method of (multi)local scale-space description is applied to equivalence problems for the extended jet of fourth order and for the affine group and the orthogonal group respectively.

2 Definition of Equivalence Problems

The definition of an equivalence problem with respect to image structure requires a representation of the extended jet of finite order corresponding to an input image and specification of the group of transformations allowed. An extended jet $j^N L$ of order N corresponding to a set of $\mathbf{C}^\infty$-images L is defined as [1]:

Definition 1. Let D and R be $\mathbf{C}^\infty$-manifolds (infinitely differentiable) and define an equivalence relation on the set of all $\mathbf{C}^\infty$-images L of D into R by requiring with respect to a chosen cartesian coordinate frame all the partial derivatives up to order N of those images to be equal on D. An equivalence class formed by such images with respect to that relation is called an extended jet $j^N L$ of order N: it consists of all the local jets of order N on D.

In order to represent physically the equivalence relations of such a jet for a 2-dimensional input image $L_0 : \mathbb{R}^2 \to \mathbb{R}$, its scale-space $L : \mathbb{R}^2 \times \mathbb{R}_0^+ \to \mathbb{R}$ and partial derivatives have to be generated by means of the linear isotropic diffusion equation with initial condition $L(.,0) = L_0$ [8]. The similarity solutions of this initial value problem then form a complete representation of the equivalence relations of that jet [1, 2, 9]. This means that a n-th order partial derivative $L_{\alpha_1 \ldots \alpha_n}$ of scale-space is given by a convolution of the input image L_0 and a fuzzy-derivative operator $G_{\alpha_1 \ldots \alpha_n}$:

$$L_{\alpha_1 \ldots \alpha_n} = L_0 * G_{\alpha_1 \ldots \alpha_n} \quad , \tag{1}$$

with

$$G(\mathbf{x}; s) = \frac{1}{4\pi s} e^{-\frac{x_i x_i}{4s}} \quad \text{and} \quad G_{\alpha_1 \ldots \alpha_n} = \frac{\partial^n G}{\partial x_{\alpha_1} \ldots \partial x_{\alpha_n}}; \quad x_\alpha = x_i, s$$

being the Gaussian kernel and its partial derivatives respectively. Here $\mathbf{x} \in \mathbb{R}^2$ represents a spatial position in cartesian coordinates, whereas $s \in \mathbb{R}_0^+$ corresponds to a level of scale. Note that *Einstein's summation convention* is used

in the sequel: sum over spatial components when indices appear twice in products. Furthermore, Greek indices denote, unless otherwise indicated, also partial derivatives with respect to scale s. But because the partial derivative with respect to s is equal to the Laplacian operator it suffices for a complete representation of the extended jet of order N corresponding to a scale-space L to consider the spatial differential structure of the extended jet of order $2N$.

The considered groups of transformations of the local cartesian coordinate frames are given below by definition.

Definition 2. The group of spatially homogeneous affine transformations $A(2)$ and that of orthogonal transformations $O(2)$ of the local cartesian coordinate frames (ξ_1, ξ_2) with basepoints $(\mathbf{x}; s) \in \mathbb{R}^2 \times \mathbb{R}_0^+$ are defined by the following transformation rules:

$$\boldsymbol{\xi}' = R\boldsymbol{\xi} \;\; ; \quad \boldsymbol{\xi} = (\xi_1, \xi_2) \;\; ,$$

with

$$\begin{cases} det(J(R)) \neq 0 \quad \text{if } R \in A(2) \\ R^t R = R R^t = \mathbf{1} \text{ if } R \in O(2) \end{cases} \tag{2}$$

where $\mathbf{1}$ is the identity operator and J the Jacobian.

Consequently the representation of the extended jet of scale-space and the groups of transformations provide a natural context for the definition of equivalence problems:

Definition 3. An equivalence problem is the problem of finding a complete and irreducible set of invariants necessary and sufficient to describe any (multi)local feature of the extended jet $j^N L$ of scale-space that is invariant under a specific group of transformations. A (multi)local invariant I then satisfies the following condition:

$$I\{R(j^N)L\} = det^g(R) I\{j^N L\} \quad \forall R \in A(2), O(2) \;\; . \tag{3}$$

Here the weight $g \in \mathbb{R}$ is determined by the invariant I; for $g = 0$ the property I is an absolute invariant, else a relative invariant.

In the sequel the problems for the different groups are called the affine and orthogonal equivalence problem respectively.

3 Solution of Equivalence Problems

From classical invariance theory it is known that the local jet of finite order at $(\mathbf{x}; s)$ in scale-space has a finite, complete and irreducible set of polynomial affine invariants [6, 10]. On the basis of this notion complete and irreducible sets of (multi)local invariants for the extended jet of finite order of scale-space and for the groups of coordinate transformations mentioned in the previous section will be derived by extending Hilbert's method [6] and by using Klein's adjunction theorem [7].

First, in order to solve the affine equivalence problem it has to be adjoined to another problem on the basis of Klein's adjunction theorem [7]:

Theorem 4. *The affine equivalence problem may be defined as the problem of finding the equivalence relations for the algebraic variety V_N that is defined by the roots of all the local binary ground forms L_n up to order N connected to scale-space:*

$$V_N \equiv \bigcup_{n=1}^{N} \{(\mathbf{x}; s; \boldsymbol{\eta}) \in \mathbb{R}^2 \times \mathbb{R}_0^+ \times \mathbb{C}^2 \mid L_n(\boldsymbol{\eta})|_{(\mathbf{x};s)} = 0\} \ . \tag{4}$$

The orthogonal equivalence problem may be adjoined to that for the affine equivalence problem by extending it for every point in scale-space with the following local algebraic equation:

$$\frac{1}{2}\eta_i\eta_i = 0 \ , \tag{5}$$

where a local binary ground form L_n of n-th order is defined as:

Definition 5. A binary ground form L_n of n-th order at $(\mathbf{x}; s)$ is defined as the corresponding n-th order part of the local Taylor expansion with respect to the cartesian coordinate frame $(\xi_1, \xi_2) \in \mathbb{R}^2$:

$$L_n(\boldsymbol{\xi})|_{(\mathbf{x};s)} = \frac{1}{n!} L_{i_1 \ldots i_n} \xi_{i_1} \cdots \xi_{i_n} \ , \tag{6}$$

where the partial derivatives $L_{i_1 \ldots i_n}$ of scale-space are evaluated according to formula (1).

Hilbert [6] has proposed in turn that the above equivalence problem at any $(\mathbf{x}; s)$ is directly related to the algebraic conditions ensuring the degeneration of another algebraic variety, namely:

Proposition 6. *The affine equivalence problem may be adjoined to the problem of finding the algebraic conditions for the existence of an algebraic variety S:*

$$S \equiv \{l_n | n = 1, \ldots, N\} \ , \tag{7}$$

where local binary ground form l_n is a product of ξ_1^{h+1} and a local binary ground form μ_{n-h-1} of $(n-h-1)$-th order:

$$l_n = \xi_1^{h+1} \mu_{n-h-1}(\boldsymbol{\xi}) \ ; \quad n = 2h \quad or \quad n = 2h+1 \ . \tag{8}$$

The proof of this proposition follows from two other theorems of Hilbert [6], namely:

Theorem 7. *If the vanishing of a set of invariants implies the vanishing of all polynomial affine invariants related to the algebraic variety defined by (4) (and (5)), then all those affine invariants are algebraic functions on that set.*

And

Theorem 8. *If all the algebraic invariants of the binary ground form of degree $n = 2h + 1$ or $n = 2h$ are zero, then the ground form has a linear factor of multiplicity $h + 1$; and, conversely, if the ground form has a linear factor of multiplicity $h + 1$, then all the algebraic invariants are zero.*

In order to obtain the algebraic equivalence relations for a single local binary ground form (6) Hilbert started off with the construction of the set Z_n of transvectants $[L_n, L_n]^k$:

$$Z_n \equiv \{L_n, [L_n, L_n]^1, \ldots, [L_n, L_n]^g\} \ , \tag{9}$$

where $g = h,\ h-1$ for n odd and even respectively. The k-th order transvectant is defined as:

Definition 9. The k-th order transvectant $[\ .\ ,\ .\ ]^k$ of two local binary ground forms A_n of order n and B_m of order m is defined by:

$$[A_n, B_m]^k \equiv \lim_{\xi \to \eta} \prod_{l=1}^{k} \epsilon_{i_l j_l} \frac{\partial}{\partial \xi_{i_l}} \frac{\partial}{\partial \eta_{j_l}} A_n(\xi) B_m(\eta) \ ,$$

in which ϵ_{ij} is the parity of the ordered pair (ij). For convenience the zero-th order transvectant of two identical forms is defined to be equal to itself.

Next he proved the following proposition [6]:

Proposition 10. *The necessary and sufficient condition for the local binary ground form (6) of n-th order to have a root of multiplicity $h + 1$ is equivalent to requiring the set of transvectants (9) to have one linear factor in common.*

To verify this, M is defined to be the least common multiple of the numbers $n, \ldots, 2(n-2h)$:

$$M = mn = 2m_1(n-2) = \ldots = 2m_g(n-2g) \ , \tag{10}$$

with

$$g = \begin{cases} h & \text{if} \quad n = 2h+1 \\ h-1 & \text{if} \quad n = 2h \end{cases}$$

On the basis of the numbers above two forms U and V are introduced with undetermined linearly independent parameters $u_0, u_1, \ldots, u_h$ and $v_0, v_1, \ldots, v_h \in \mathbb{R}$, namely:

$$U = \sum_{k=0}^{g} u_k ([L_n, L_n]^k)^{m_k} \ , \tag{11}$$

$$V = \sum_{k=0}^{g} v_k ([L_n, L_n]^k)^{m_k} \ . \tag{12}$$

Finally from classical algebra [11] it is known that the set (9) has a common linear factor if and only if the resultant R of the forms U and V vanishes:

$$R(U,V) = J_1 P_1 + \ldots + J_\mu P_\mu \ ; \quad J_\nu = 0 \ . \tag{13}$$

Here each P_ν is a product of powers of the parameters u and v above and each $J_\nu = 0$ forms an algebraic equivalence relation of the local affine equivalence problem for ground form (6). Thus the J_ν form a complete and irreducible set of local relative affine invariants of the local binary ground form (6).

Knowing the equivalence relations at $(\mathbf{x}; s)$ for each ground form (6) of the algebraic variety (4), then the equivalence relations still should be derived that are related to the algebraic conditions for the coincidence of roots of the local binary ground forms constituting the degenerated algebraic variety (7). Hilbert has proved that these conditions correspond to the vanishing of the resultant of two different transvectants of the ground forms [6]. Examples of this are given in Sect. 4.

Finally a complete and irreducible set of (multi)local algebraic invariants for the extended jet of scale-space and the affine group are found on the basis of a third theorem of Hilbert [6]:

Theorem 11. *Let A_n and B_n be two local binary ground forms of equal order, A_n one of the algebraic variety (4) at $(\mathbf{x}_1; s_1)$ and B_n also but at $(\mathbf{x}_2; s_2)$. Repeated application of the Aronhold operator P:*

$$P \equiv B_{i_1 \ldots i_n} \frac{\partial}{\partial A_{i_1 \ldots i_n}} \tag{14}$$

on a complete and irreducible set of local algebraic invariants for the problem at $(\mathbf{x}_1; s_1)$ generates a complete and irreducible set of bilocal algebraic invariants comparing multilocal structures of scale-space. Applying again the Aronhold operators to the latter set results in a similar set of trilocal invariants, etc.

In order to solve the affine equivalence problem and that of zero weight, $g(I) = 0$, rational invariants have to be formed on the basis of the affine invariants above: a complete and irreducible set of absolute affine invariants will then come about.

Secondly, adjoining on the basis of Klein's adjunction theorem (4), algebraic relation (5) to the degenerated algebraic variety (7), the orthogonal equivalence problem turns out to be similar to the affine equivalence problem. All basic orthogonal invariants may be found on the basis of Theorem 11 by applying the adjoined Aronhold operator Π given by:

$$\Pi = \delta_{ij} \frac{\partial}{\partial L_{ij}} \ ; \quad \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} \tag{15}$$

to the complete and irreducible set of relative affine invariants.

4 Examples of Equivalence Problems

The aim is to describe the scale-space of a 2-dimensional input image in terms of a complete and irreducible set of (multi)local algebraic invariants for its extended jet of fourth order and different groups of transformations by means of the methods developed in the previous section. Successively the affine and the orthogonal equivalence problem are solved. First of all the local algebraic invariants for the affine equivalence problem are derived that have to vanish on the basis of Theorem 7 whenever the spatial image structure of the local 4-jet becomes:

$$L_1 = \lambda_1 \xi_1 \ , \tag{16}$$

$$L_2 = \lambda_2 \xi_1^2 \ , \tag{17}$$

$$L_3 = \lambda_3 \xi_1^2 \xi_2 \ , \tag{18}$$

$$L_4 = \lambda_4 \xi_1^3 (a_1 \xi_1 + a_2 \xi_2) \ ; \quad a_k \in \mathbb{R}. \tag{19}$$

Here $\lambda_n \in \mathbb{R}$ are arbitrary constants representing the differential structure of the n-th order. Next a complete and irreducible set of (multi)local relative affine invariants of the extended jet of fourth order is generated on the basis of Theorem 11. The affine equivalence problem and that of zero weight is consecutively exemplified by forming a rational invariant on the basis of the relative affine invariants derived. Finally the orthogonal equivalence problem is solved by applying Aronhold operators (15) to the relative affine invariants and using again Theorem 11.

Solving the affine equivalence problem for the conditions (16) and (17) gives a complete and irreducible set S_2^a of affine invariants for the local 2-jet:

$$S_2^a = \{[L_1^2, L_2]^2, D(L_2)\} \ ; \quad D(L_2) \equiv [L_2, L_2]^2 \ . \tag{20}$$

In this set the second and third irreducible invariants are equal to the resultant of the linear and the quadratic local binary ground forms L_1 and L_2 and the opposite of one half of the discriminant of L_2 respectively.

The former equivalence problem may be extended to the local 3-jet by imposing the additional condition (18). Requiring the cubic local binary ground form L_3 to have a root of multiplicity two is equivalent to the vanishing of its discriminant $D(L_3)$:

$$D(L_3) \equiv [[L_3, L_3]^2, [L_3, L_3]^2]^2 \ . \tag{21}$$

Whether L_1 and L_2 have at least one linear factor in common with L_3 depends on the vanishing of the resultants of those forms with L_3:

$$R(L_1, L_3) \equiv [L_1^3, L_3]^3 \ , \tag{22}$$

$$R(L_2, L_3) \equiv [[L_2, L_3]^2 [L_2, L_3]^2, L_2]^2 + D(L_2)[[L_3, L_3]^2, L_2]^2 \ . \tag{23}$$

Equivalence relations for the coincidence of L_1 and L_2 with the factor ξ_1^2 of L_3 still have to be derived. It was shown by Hilbert [6] that for this to be the case

the vanishing of the local algebraic invariants Q that measure the coincidence of L_1^2 and L_2 with the determinant of the Hessian $H(L_3)$ of L_3 [6] is required:

$$Q(L_1^2, H(L_3)) \equiv [L_1^2, [L_3, L_3]^2]^2 \ , \tag{24}$$

$$Q(L_2, H(L_3)) \equiv [L_2, [L_3, L_3]^2]^2 \ . \tag{25}$$

The union of the features (21), ..., (25) and (20) forms a complete and irreducible set S_3^a of affine invariants of the local 3-jet.

Adding the pure fourth-order local structure to the local 3-jet a complete and irreducible set S_4^a of affine invariants may analogously be found that fulfill the conditions (16) up to (19):

$$S_4^a = S_3^a \cup \delta S_4^a \ , \tag{26}$$

with

$$\begin{aligned} \delta S_4^a = \{ & i(L_4), j(L_4), R(L_1, L_4), Q^2(L_1, H(L_4)), R(L_2, L_4), \\ & R(L_3, L_4), Q^2(L_2^2, H(L_4)), Q^2(H(L_3)^2, H(L_4))\} \ . \end{aligned} \tag{27}$$

The listed algebraic invariants in (27) are defined by:

$$\begin{aligned} i(L_4) &\equiv [L_4, L_4]^4 \ , \\ j(L_4) &\equiv [[L_4, L_4]^2, L_4]^4 \ , \\ R(L_1, L_4) &\equiv [L_1^4, L_4]^4 \ , \\ Q^2(L_1, H(L_4)) &\equiv [L_1^4, [L_4, L_4]^2]^4 \ , \\ R(L_2, L_4) &\equiv [L_2^2, L_4]^4 [L_2^2, L_4]^4 + 2D(L_2)[[L_4, L_4]^2, L_2^2]^4 + \\ &\qquad \frac{1}{6} D(L_2) D(L_2) [L_4, L_4]^4 \ , \\ Q^2(L_2^2, H(L_4)) &\equiv [L_2^2, [L_4, L_4]^2]^4 \ , \\ R(L_3, L_4) &\equiv [\alpha, \beta]^2 \ , \\ Q^2(H^2(L_3), H(L_4)) &\equiv [[L_3, L_3]^2 [L_3, L_3]^2, [L_4, L_4]^2]^4 \ , \end{aligned}$$

with

$$\alpha = [\rho, \rho]^4 - \frac{6}{5}[\rho, \sigma]^3 + \frac{12}{25}[\sigma, \sigma]^2 - [\sigma, \tau] - \frac{1}{6}\tau^2 \ ,$$

$$\beta = [\rho, L_3]^3 + \frac{6}{5}[\sigma, L_3]^2 + \frac{1}{2}[\tau, L_3] \ ,$$

where

$$\rho = 3[L_4, L_3] \ , \quad \sigma = \frac{1}{2}[L_4, L_3]^2 \ , \quad \tau = [L_4, L_3]^3 \ .$$

This concludes our search for the local algebraic invariants that satisfy the affine equivalence problem for the local 4-jet. Applying now Theorem 11 gives a complete and irreducible set of (multi)local invariants for the considered problem.

The solution of the affine equivalence problem under the additional condition of absolute invariance comes about by deriving by means of the relative affine

invariants a complete and irreducible set of rational invariant features of zero weight. For example the following absolute affine invariant feature α may be constructed:

$$\alpha \equiv \frac{([L_4, L_4]^4)^3}{([[L_4, L_4]^2, L_4]^4)^2} .$$

This image property is directly related to the (an)harmonic division of the roots of the fourth-order local binary ground forms L_4 [10].

In order to solve the orthogonal equivalence problem the affine equivalence problem is extended by imposing algebraic condition (5). When the local relative affine invariants derived above are subjected to the Aronhold operator (15), a supplementary irreducible set δS_4^o of orthogonal invariants of the local 4-jet is obtained:

$$\begin{aligned} \delta S_4^o = \{ & L, \Pi^k[L_1^2, L_2]^2, \Pi^k[L_2, L_2]^2, \Pi^k R(L_2, L_3), \\ & \Pi^k Q(L_2, H(L_3)), \Pi^k R(L_2, L_4), \Pi^k Q^2(L_2^2, H(L_4)) \mid \quad k \in \mathbb{N} \} . \end{aligned} \tag{28}$$

A further application of again Theorem 11 gives rise to a complete and irreducible set of (multi)local orthogonal invariants.

5 Discussion

A method has been presented to derive a complete and irreducible set of algebraic invariants that may describe any (multi)local affine or orthogonal invariant of the scale-space that corresponds to a 2-dimensional input image. Such a set is of major importance for a number of reasons. It forms, due to its completeness and irreducibility, a solid basis for feature detection, patch classification and topological description of that scale-space: for example, corners, T-junctions, Euler numbers may be extracted [12, 13]. It may play a crucial role in studying and understanding the behavior of nonlinear observation processes [9]. Furthermore, it may be used to tackle stereo, multi-modality and optic flow problems. For example, actually existing 3-dimensional geometric structures may appear under restricted environmental conditions, e.g. variations in the light source distribution, as induced structures of the jet bundle of the scale(-time) spaces of a pair or a time sequence of 2-dimensional input images. As covariances of image structure can define a process of parallel transport related to a connection on scale(-time) space, so may (multi)local algebraic invariants play equally well a crucial role in operationalizing that process [14]. Finally, for nonlinear equivalence problems for 2-dimensional input images and arbitrary groups of transformations of the variables it is even possible to express the solution in terms of the algebraic invariants found above [4].

What about the solution of equivalence problems with respect to higher dimensional input images and other transformation groups? Again the method described above may turn out to be appropriate for solving similar equivalence problems for 3-dimensional input images [6]. The symbolic method may, however, also lead to such a description [10, 6]: form all possible invariant features

up to a certain weight; decompose them by means of certain reduction rules into a complete set of invariant features and find all irreducible syzygies, i.e. algebraic relations for the derived invariants. For nonlinear equivalence problems for the extended jet and an arbitrary group of transformations of the variables Cartan's method seems to be the most attractive approach to follow [5].

Another appealing question is how to describe implicitly defined structures of scale-space itself? The connection on scale-space may mainly be set by the projection fields of algebraic or differential geometric invariants [8, 15].

References

1. Koenderink, J. J., van Doorn A. J. (1987). Representation of local geometry in the visual system, Biol. Cybern. 55, pp. 367–375.
2. Koenderink, J. J., van Doorn A. J. (1990). Receptive field families, Biol. Cybern. 63, pp. 291–298.
3. Cartan, E. (1952). Les problèmes d'équivalence. In: Oeuvres Complètes 2, pp. 1311–1334, Gauthiers-Villars, Paris.
4. Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J., Viergever, M. A. (August 1991). General intensity transformations. In: Johansen, P. and Olsen, S. (eds.), Proc. 7th Scand. Conf. on Image Analysis, (Aalborg, DK), pp. 338–345.
5. Salden, A. H., Florack, L. M. J., ter Haar Romeny, B. M. (1991). Differential geometric description of 3D scalar images, Internal Report 3DCV # 91-10.
6. Hilbert, D. (1893). Ueber die vollen Invariantensystemen, Math. Annalen, 42, pp. 313–373.
7. Klein, F. (1893). Erlanger Programm, Math. Annalen, 43, pp. 63–100.
8. Koenderink, J. J. (1984). The structure of images, Biol. Cybern. 50, pp. 363–370.
9. Salden, A. H., Florack, L. M. J., ter Haar Romeny, B. M., Viergever, M. A., Koenderink, J. J. (April 1992). The nonlinear rescaling process, Internal Report 3DCV # 92-28.
10. Weitzenböck, R. (1923). Invariantentheorie, P. Noordhoff, Groningen.
11. Gordan, P. (1871). Über die Bildung der Resultante zweier Gleichungen, Math. Annalen, 50, pp. 363–370.
12. Koenderink, J. J. (1984). The concept of local sign. In: van Doorn, A.J., van de Grind, W.A. and Koenderink, J.J. (eds.), Limits in Perception, VNU Science Press, Utrecht, pp. 495–547.
13. Lindeberg, T. (May 1991). Discrete Scale-Space Theory and the Scale-Space Primal Sketch, PhD thesis, Royal Institute of Technology, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden.
14. Koenderink, J. J., van Doorn A. J. (1988). Operational significance of receptive field assemblies, Biol. Cybern. 58, pp. 163–171.
15. Salden, A. H., Florack, L. M. J., ter Haar Romeny, B. M., Viergever, M. A., Koenderink, J. J. (1992). Multi-scale analysis and description of image structure, Nieuw Archief voor Wiskunde.

List of Authors

Subject Index

Printing: Druckhaus Beltz, Hemsbach
Binding: Buchbinderei Schäffer, Grünstadt

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 1: Issues in Acoustic Signal - Image Processing and Recognition. Edited by C. H. Chen. VIII, 333 pages. 1983. *(out of print)*

Vol. 2: Image Sequence Processing and Dynamic Scene Analysis. Edited by T. S. Huang. IX, 749 pages. 1983.

Vol. 3: Electronic Systems Effectiveness and Life Cycle Costing. Edited by J. K. Skwirzynski. XVII, 732 pages. 1983. *(out of print)*

Vol. 4: Pictorial Data Analysis. Edited by R. M. Haralick. VIII, 468 pages. 1983.

Vol. 5: International Calibration Study of Traffic Conflict Techniques. Edited by E. Asmussen. VII, 229 pages. 1984.

Vol. 6: Information Technology and the Computer Network. Edited by K. G. Beauchamp. VIII, 271 pages. 1984.

Vol. 7: High-Speed Computation. Edited by J. S. Kowalik. IX, 441 pages. 1984.

Vol. 8: Program Transformation and Programming Environments. Report on a Workshop directed by F. L. Bauer and H. Remus. Edited by P. Pepper. XIV, 378 pages. 1984.

Vol. 9: Computer Aided Analysis and Optimization of Mechanical System Dynamics. Edited by E. J. Haug. XXII, 700 pages. 1984. *(out of print)*

Vol. 10: Simulation and Model-Based Methodologies: An Integrative View. Edited by T. I. Ören, B. P. Zeigler, M. S. Elzas. XIII, 651 pages. 1984.

Vol. 11: Robotics and Artificial Intelligence. Edited by M. Brady, L. A. Gerhardt, H. F. Davidson. XVII, 693 pages. 1984.

Vol. 12: Combinatorial Algorithms on Words. Edited by A. Apostolico, Z. Galil. VIII, 361 pages. 1985.

Vol. 13: Logics and Models of Concurrent Systems. Edited by K. R. Apt. VIII, 498 pages. 1985.

Vol. 14: Control Flow and Data Flow: Concepts of Distributed Programming. Edited by M. Broy. VIII, 525 pages. 1985. *Reprinted as Springer Study Edition 1986.*

Vol. 15: Computational Mathematical Programming. Edited by K. Schittkowski. VIII, 451 pages. 1985.

Vol. 16: New Systems and Architectures for Automatic Speech Recognition and Synthesis. Edited by R. De Mori, C.Y. Suen. XIII, 630 pages. 1985. *(out of print)*

Vol. 17: Fundamental Algorithms for Computer Graphics. Edited by R. A. Earnshaw. XVI, 1042 pages. 1985. *Reprinted as Springer Study Edition 1991.*

Vol. 18: Computer Architectures for Spatially Distributed Data. Edited by H. Freeman and G. G. Pieroni. VIII, 391 pages. 1985. *(out of print)*

Vol. 19: Pictorial Information Systems in Medicine. Edited by K. H. Höhne. XII, 525 pages. 1986.

Vol. 20: Disordered Systems and Biological Organization. Edited by E. Bienenstock, F. Fogelman Soulié, G. Weisbuch. XXI, 405 pages.1986

Vol. 21: Intelligent Decision Support in Process Environments. Edited by E. Hollnagel, G. Mancini, D. D. Woods. XV, 524 pages. 1986.

Vol. 22: Software System Design Methods. The Challenge of Advanced Computing Technology. Edited by J. K. Skwirzynski. XIII, 747 pages. 1986. *(out of print)*

Vol. 23: Designing Computer-Based Learning Materials. Edited by H. Weinstock and A. Bork. IX, 285 pages. 1986. *(out of print)*

Vol. 24: Database Machines. Modern Trends and Applications. Edited by A. K. Sood and A. H. Qureshi. VIII, 570 pages. 1986.

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 25: Pyramidal Systems for Computer Vision. Edited by V. Cantoni and S. Levialdi. VIII, 392 pages. 1986. *(ROB)*

Vol. 26: Modelling and Analysis in Arms Control. Edited by R. Avenhaus, R. K. Huber and J. D. Kettelle. VIII, 488 pages. 1966. *(out of print)*

Vol. 27: Computer Aided Optimal Design: Structural and Mechanical Systems. Edited by C. A. Mota Soares. XIII, 1029 pages. 1987.

Vol. 28: Distributed Operating Systems. Theory und Practice. Edited by Y. Paker, J.-P. Banatre and M. Bozyiğit. X, 379 pages. 1987.

Vol. 29: Languages for Sensor-Based Control in Robotics. Edited by U. Rembold and K. Hörmann. IX, 625 pages. 1987. *(ROB)*

Vol. 30: Pattern Recognition Theory and Applications. Edited by P. A. Devijver and J. Kittler. XI, 543 pages. 1987.

Vol. 31: Decision Support Systems: Theory and Application. Edited by C. W. Holsapple and A. B. Whinston. X, 500 pages. 1987.

Vol. 32: Information Systems: Failure Analysis. Edited by J. A. Wise and A. Debons. XV, 338 pages. 1987.

Vol. 33: Machine Intelligence and Knowledge Engineering for Robotic Applications. Edited by A. K. C. Wong and A. Pugh. XIV, 486 pages. 1987. *(ROB)*

Vol. 34: Modelling, Robustness and Sensitivity Reduction in Control Systems. Edited by R.F. Curtain. IX, 492 pages. 1987.

Vol. 35: Expert Judgment and Expert Systems. Edited by J. L. Mumpower, L. D. Phillips, C. Renn and V. R. R. Uppuluri. VIII, 361 pages. 1987.

Vol. 36: Logic of Programming and Calculi of Discrete Design. Edited by M. Broy. VII, 415 pages. 1987.

Vol. 37: Dynamics of Infinite Dimensional Systems. Edited by S.-N. Chow and J. K. Hale. IX. 514 pages. 1987.

Vol. 38: Flow Control of Congested Networks. Edited by A. R. Odoni, L. Bianco and G. Szegö. XII, 355 pages. 1987.

Vol. 39: Mathematics and Computer Science in Medical Imaging. Edited by M. A. Viergever and A. Todd-Pokropek. VIII, 546 pages. 1988.

Vol. 40: Theoretical Foundations of Computer Graphics and CAD. Edited by R. A. Earnshaw. XX, 1246 pages. 1988. *(out of print)*

Vol. 41: Neural Computers. Edited by R. Eckmiller and Ch. v. d. Malsburg. XIII, 566 pages. 1988. *Reprinted as Springer Study Edition 1989, 1990.*

Vol. 42: Real-Time Object Measurement and Classification. Edited by A. K. Jain. VIII, 407 pages. 1988. *(ROB)*

Vol. 43: Sensors and Sensory Systems for Advanced Robots. Edited by P. Dario. XI, 597 pages. 1988. *(ROB)*

Vol. 44: Signal Processing and Pattern Recognition in Nondestructive Evaluation of Materials. Edited by C. H. Chen. VIII, 344 pages. 1988. *(ROB)*

Vol. 45: Syntactic and Structural Pattern Recognition. Edited by G. Ferraté, T. Pavlidis, A. Sanfeliu and H. Bunke. XVI, 467 pages. 1988. *(ROB)*

Vol. 46: Recent Advances in Speech Understanding and Dialog Systems. Edited by H. Niemann, M. Lang and G. Sagerer. X, 521 pages. 1988.

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 47: Advanced Computing Concepts and Techniques in Control Engineering. Edited by M. J. Denham and A. J. Laub. XI, 518 pages. 1988. *(out of print)*

Vol. 48: Mathematical Models for Decision Support. Edited by G. Mitra. IX, 762 pages. 1988.

Vol. 49: Computer Integrated Manufacturing. Edited by I. B. Turksen. VIII, 568 pages. 1988.

Vol. 50: CAD Based Programming for Sensory Robots. Edited by B. Ravani. IX, 565 pages. 1988. *(ROB)*

Vol. 51: Algorithms and Model Formulations in Mathematical Programming. Edited by S. W. Wallace. IX, 190 pages. 1989.

Vol. 52: Sensor Devices and Systems for Robotics. Edited by A. Casals. IX, 362 pages. 1989. *(ROB)*

Vol. 53: Advanced Information Technologies for Industrial Material Flow Systems. Edited by S. Y. Nof and C. L. Moodie. IX, 710 pages. 1989.

Vol. 54: A Reappraisal of the Efficiency of Financial Markets. Edited by R. M. C. Guimarães, B. G. Kingsman and S. J. Taylor. X, 804 pages. 1989.

Vol. 55: Constructive Methods in Computing Science. Edited by M. Broy. VII, 478 pages. 1989.

Vol. 56: Multiple Criteria Decision Making and Risk Analysis Using Microcomputers. Edited by B. Karpak and S. Zionts. VII, 399 pages. 1989.

Vol. 57: Kinematics and Dynamic Issues in Sensor Based Control. Edited by G. E. Taylor. XI, 456 pages. 1990. *(ROB)*

Vol. 58: Highly Redundant Sensing in Robotic Systems. Edited by J. T. Tou and J. G. Balchen. X, 322 pages. 1990. *(ROB)*

Vol. 59: Superconducting Electronics. Edited by H. Weinstock and M. Nisenoff. X, 441 pages. 1989.

Vol. 60: 3D Imaging in Medicine. Algorithms, Systems, Applications. Edited by K. H. Höhne, H. Fuchs and S. M. Pizer. IX, 460 pages. 1990. *(out of print)*

Vol. 61: Knowledge, Data and Computer-Assisted Decisions. Edited by M. Schader and W. Gaul. VIII, 421 pages. 1990.

Vol. 62: Supercomputing. Edited by J. S. Kowalik. X, 425 pages. 1990.

Vol. 63: Traditional and Non-Traditional Robotic Sensors. Edited by T. C. Henderson. VIII, 468 pages. 1990. *(ROB)*

Vol. 64: Sensory Robotics for the Handling of Limp Materials. Edited by P. M. Taylor. IX, 343 pages. 1990. *(ROB)*

Vol. 65: Mapping and Spatial Modelling for Navigation. Edited by L. F. Pau. VIII, 357 pages. 1990. *(ROB)*

Vol. 66: Sensor-Based Robots: Algorithms and Architectures. Edited by C. S. G. Lee. X, 285 pages. 1991. *(ROB)*

Vol. 67: Designing Hypermedia for Learning. Edited by D. H. Jonassen and H. Mandl. XXV, 457 pages. 1990. *(AET)*

Vol. 68: Neurocomputing. Algorithms, Architectures and Applications. Edited by F. Fogelman Soulié and J. Hérault. XI, 455 pages. 1990.

Vol. 69: Real-Time Integration Methods for Mechanical System Simulation. Edited by E. J. Haug and R. C. Deyo. VIII, 352 pages. 1991.

Vol. 70: Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms. Edited by G. H. Golub and P. Van Dooren. XIII, 729 pages. 1991.

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 71: Expert Systems and Robotics. Edited by T. Jordanides and B.Torby. XII, 744 pages. 1991.

Vol. 72: High-Capacity Local and Metropolitan Area Networks. Architecture and Performance Issues. Edited by G. Pujolle. X, 536 pages. 1991.

Vol. 73: Automation and Systems Issues in Air Traffic Control. Edited by J. A. Wise, V. D. Hopkin and M. L. Smith. XIX, 594 pages. 1991.

Vol. 74: Picture Archiving and Communication Systems (PACS) in Medicine. Edited by H. K. Huang, O. Ratib, A. R. Bakker and G. Witte. XI, 438 pages. 1991.

Vol. 75: Speech Recognition and Understanding. Recent Advances, Trends and Applications. Edited by P. Laface and Renato De Mori. XI, 559 pages. 1991.

Vol. 76: Multimedia Interface Design in Education. Edited by A. D. N. Edwards and S. Holland. XIV, 216 pages. 1992. *(AET)*

Vol. 77: Computer Algorithms for Solving Linear Algebraic Equations. The State of the Art. Edited by E. Spedicato. VIII, 352 pages. 1991.

Vol. 78: Integrating Advanced Technology into Technology Education. Edited by M. Hacker, A. Gordon and M. de Vries. VIII, 185 pages. 1991. *(AET)*

Vol. 79: Logic, Algebra, and Computation. Edited by F. L. Bauer. VII, 485 pages. 1991.

Vol. 80: Intelligent Tutoring Systems for Foreign Language Learning. Edited by M. L. Swartz and M. Yazdani. IX, 347 pages. 1992. *(AET)*

Vol. 81: Cognitive Tools for Learning. Edited by P. A. M. Kommers, D. H. Jonassen, and J. T. Mayes. X, 278 pages. 1992. *(AET)*

Vol. 82: Combinatorial Optimization. New Frontiers in Theory and Practice. Edited by M. Akgül, H. W. Hamacher, and S. Tüfekçi. XI, 334 pages. 1992.

Vol. 83: Active Perception and Robot Vision. Edited by A. K. Sood and H. Wechsler. IX, 756 pages. 1992.

Vol. 84: Computer-Based Learning Environments and Problem Solving. Edited by E. De Corte, M. C. Linn, H. Mandl, and L. Verschaffel. XVI, 488 pages. 1992. *(AET)*

Vol. 85: Adaptive Learning Environments. Foundations and Frontiers. Edited by M. Jones and P. H. Winne. VIII, 408 pages. 1992. *(AET)*

Vol. 86: Intelligent Learning Environments and Knowledge Acquisition in Physics. Edited by A. Tiberghien and H. Mandl. VIII, 285 pages. 1992. *(AET)*

Vol. 87: Cognitive Modelling and Interactive Environments. With demo diskettes (Apple and IBM compatible). Edited by F. L. Engel, D. G. Bouwhuis, T. Bösser, and G. d'Ydewalle. IX, 311 pages. 1992. *(AET)*

Vol. 88: Programming and Mathematical Method. Edited by M. Broy. VIII, 428 pages. 1992.

Vol. 89: Mathematical Problem Solving and New Information Technologies. Edited by J. P. Ponte, J. F. Matos, J. M. Matos, and D. Fernandes. XV, 346 pages. 1992. *(AET)*

Vol. 90: Collaborative Learning Through Computer Conferencing. Edited by A. R. Kaye. X, 260 pages. 1992. *(AET)*

Vol. 91: New Directions for Intelligent Tutoring Systems. Edited by E. Costa. X, 296 pages. 1992. *(AET)*

Vol. 92: Hypermedia Courseware: Structures of Communication and Intelligent Help. Edited by A. Oliveira. X, 241 pages. 1992. *(AET)*

Vol. 93: Interactive Multimedia Learning Environments. Human Factors and Technical Considerations on Design Issues. Edited by M. Giardina. VIII, 254 pages. 1992. *(AET)*

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 94: Logic and Algebra of Specification. Edited by F. L. Bauer, W. Brauer, and H. Schwichtenberg. VII, 442 pages. 1993.

Vol. 95: Comprehensive Systems Design: A New Educational Technology. Edited by C. M. Reigeluth, B. H. Banathy, and J. R. Olson. IX, 437 pages. 1993. *(AET)*

Vol. 96: New Directions in Educational Technology. Edited by E. Scanlon and T. O'Shea. VIII, 251 pages. 1992. *(AET)*

Vol. 97: Advanced Models of Cognition for Medical Training and Practice. Edited by D. A. Evans and V. L. Patel. XI, 372 pages. 1992. *(AET)*

Vol. 98: Medical Images: Formation, Handling and Evaluation. Edited by A. E. Todd-Pokropek and M. A. Viergever. IX, 700 pages. 1992.

Vol. 99: Multisensor Fusion for Computer Vision. Edited by J. K. Aggarwal. XI, 456 pages. 1993. *(ROB)*

Vol. 100: Communication from an Artificial Intelligence Perspective. Theoretical and Applied Issues. Edited by A. Ortony, J. Slack and O. Stock. XII, 260 pages. 1992.

Vol. 101: Recent Developments in Decision Support Systems. Edited by C. W. Holsapple and A. B. Whinston. XI, 618 pages. 1993.

Vol. 102: Robots and Biological Systems: Towards a New Bionics? Edited by P. Dario, G. Sandini and P. Aebischer. XII, 786 pages. 1993.

Vol. 103: Parallel Computing on Distributed Memory Multiprocessors. Edited by F. Özgüner and F. Erçal. VIII, 332 pages. 1993.

Vol. 104: Instructional Models in Computer-Based Learning Environments. Edited by S. Dijkstra, H. P. M. Krammer and J. J. G. van Merriënboer. X, 510 pages. 1993. *(AET)*

Vol. 105: Designing Environments for Constructive Learning. Edited by T. M. Duffy, J. Lowyck and D. H. Jonassen. VIII, 374 pages. 1993. *(AET)*

Vol. 106: Software for Parallel Computation. Edited by J. S. Kowalik and L. Grandinetti. IX, 363 pages. 1993.

Vol. 107: Advanced Educational Technologies for Mathematics and Science. Edited by D. L. Ferguson. XII, 749 pages. 1993. *(AET)*

Vol. 108: Concurrent Engineering: Tools and Technologies for Mechanical System Design. Edited by E. J. Haug. XIII, 998 pages. 1993.

Vol. 109: Advanced Educational Technology in Technology Education. Edited by A. Gordon, M. Hacker and M. de Vries. VIII, 253 pages. 1993. *(AET)*

Vol. 110: Verification and Validation of Complex Systems: Human Factors Issues. Edited by J. A. Wise, V. D. Hopkin and P. Stager. XIII, 704 pages. 1993.

Vol. 111: Cognitive Models and Intelligent Environments for Learning Programming. Edited by E. Lemut, B. du Boulay and G. Dettori. VIII, 305 pages. 1993. *(AET)*

Vol. 112: Item Banking: Interactive Testing and Self-Assessment. Edited by D. A. Leclercq and J. E. Bruno. VIII, 261 pages. 1993. *(AET)*

Vol. 113: Interactive Learning Technology for the Deaf. Edited by B. A. G. Elsendoorn and F. Coninx. XIII, 285 pages. 1993. *(AET)*

Vol. 114: Intelligent Systems: Safety, Reliability and Maintainability Issues. Edited by O. Kaynak, G. Honderd and E. Grant. XI, 340 pages. 1993.

Vol. 115: Learning Electricity and Electronics with Advanced Educational Technology. Edited by M. Caillot. VII, 329 pages. 1993. *(AET)*

NATO ASI Series F

Including Special Programmes on Sensory Systems for Robotic Control (ROB) and on Advanced Educational Technology (AET)

Vol. 116: Control Technology in Elementary Education. Edited by B. Denis. IX, 311 pages. 1993 *(AET)*

Vol. 118: Program Design Calculi. Edited by M. Broy. VIII, 409 pages. 1993.

Vol. 119: Automating Instructional Design, Development, and Delivery. Edited by. R. D. Tennyson. VIII, 266 pages. 1994 *(AET)*

Vol. 120: Reliability and Safety Assessment of Dynamic Process Systems. Edited by T. Aldemir, N. O. Siu, A. Mosleh, P. C. Cacciabue and B. G. Göktepe. X, 242 pages. 1994.

Vol. 121: Learning from Computers: Mathematics Education and Technology. Edited by C. Keitel and K. Ruthven. XIII, 332 pages. 1993. *(AET)*

Vol. 122: Simulation-Based Experiential Learning. Edited by D. M. Towne, T. de Jong and H. Spada. XIV, 274 pages. 1993. *(AET)*

Vol. 123: User-Centred Requirements for Software Engineering Environments. Edited by D. J. Gilmore, R. L. Winder, F. Détienne. VII, 377 pages. 1993.

Vol. 124: Fundamentals in Handwriting Recognition. Edited by S. Impedovo. IX, 496 pages. 1994.

Vol. 125: Student Modelling: The Key to Individualized Knowledge-Based Instruction. Edited by J. Greer and G. McCalla. X, 383 pages. 1994.

Vol. 126: Shape in Picture. Mathematical Description of Shape in Grey-level Images. Edited by Y.-L. O, A. Toet, D. Foster, H. J. A. M. Heijmans and P. Meer. XI, 676 pages. 1994.

Vol. 127: Real Time Computing. Edited by W. A. Halang and A. D. Stoyenko. XXII, 762 pages. 1994.