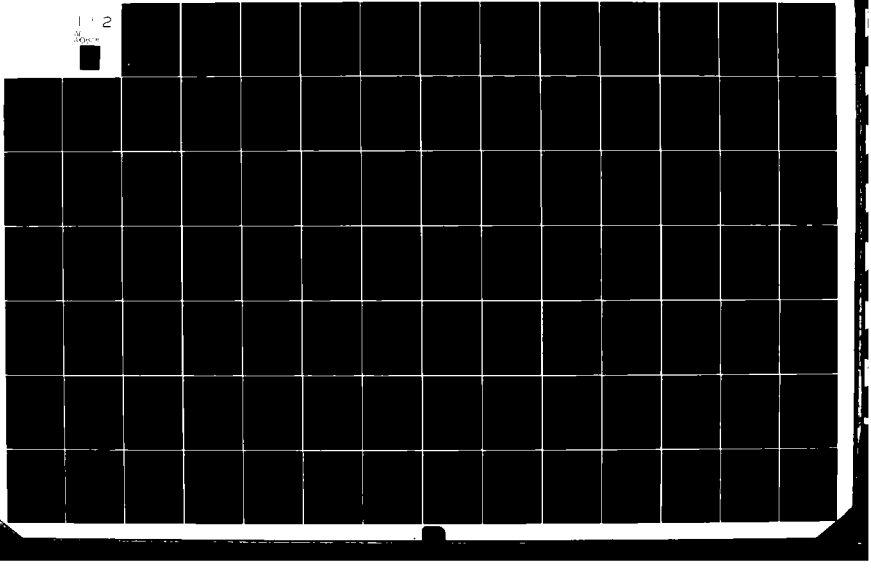
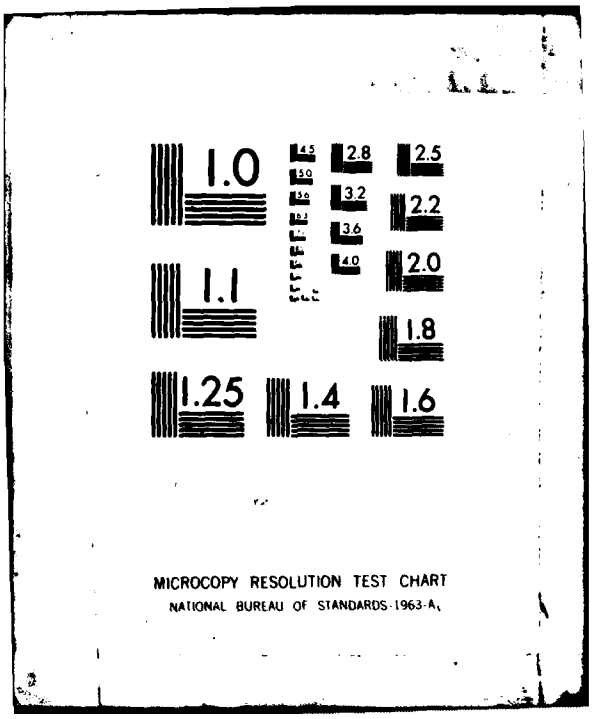
AD-A108 191 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 9/2
COMPUTATIONAL APPROACHES TO IMAGE UNDERSTANDING,(U)
OCT 81 M BRADY N00014-80-C-0505
UNCLASSIFIED AI-M-653 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A108191

DTIC FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

12

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AIM 653	2. GOVT ACCESSION NO.: AD-A108191
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Computational approaches to Image Understanding	5. TYPE OF REPORT & PERIOD COVERED: Memorandum
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Michael Brady	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0505
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, Massachusetts 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd, Arlington, Virginia 22209	12. REPORT DATE: October 1981 13. NUMBER OF PAGES: 186
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, Virginia 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Distribution is unlimited

ELECTE DEC 8 1981 S A

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Texture
Shape
Objects
Understanding

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent theoretical developments in Image Understanding are surveyed. Among the issues discussed are: edge finding, region finding, texture, shape from shading, shape from texture, shape from contour, and the representations of surfaces and objects. Much of the work described was developed in the DARPA Image Understanding Project. This paper will appear in Computing Surveys In Memory of Max Clowes and David Marr.

81 12 08 017

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601 |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 653 October, 1981

Computational approaches to Image Understanding

Michael Brady

ABSTRACT Recent theoretical developments in Image Understanding are surveyed. Among the issues discussed are: edge finding, region finding, texture, shape from shading, shape from texture, shape from contour, and the representations of surfaces and objects. Much of the work described was developed in the DARPA Image Understanding project.

This paper will appear in *Computing Surveys*

In memory of Max Clowes and David Marr

Acknowledgements. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's Artificial Intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505.

1. Introduction

One of the earliest applications of computers was the processing of visual data. With the benefit of *hindsight*, we can *see* that this *reflects* the importance of sight for humans, the difficulties faced by those lacking sight, and the continuing drive in computer science to automate human abilities.

There is currently a surge of interest in image understanding on the part of industry and the military. Interest seems certain to expand over the next several decades, as the following list of current applications indicates:

- AUTOMATION OF INDUSTRIAL PROCESSES.

Object acquisition by robot arms, for example by "bin picking".

Automatic guidance of seam welders and cutting tools.

VLSI-related processes, such as lead bonding, chip alignment and packaging.

Monitoring, filtering, and thereby containing the flood of data from oil drill sites or from seismographs.

Providing visual feedback for automatic assembly and repair.

- INSPECTION TASKS

The inspection of printed circuit boards for spurs, shorts, and bad connections.

Checking the results of casting processes for impurities and fractures.

Screening medical images such as chromosome slides, cancer smears, x-ray and ultrasound images, tomography.

Routine screening of plant samples.

- REMOTE SENSING

Cartography, the automatic generation of hill shaded maps, and the registration of satellite images with terrain maps.

Monitoring traffic along roads, docks, and at airfields.

Management of land resources such as water, forestry, soil erosion, and crop growth.

Exploration of remote or hostile regions for fossil fuels and mineral ore deposits.

• MAKING COMPUTER POWER MORE ACCESSIBLE.

Management information systems that have a communication channel considerably wider than current systems that are addressed by typing or pointing.

Document readers (for those that still use paper).

Design aids for architects and mechanical engineers.

• MILITARY APPLICATIONS.

Tracking moving objects.

Automatic navigation based on passive sensing.

Target acquisition and range finding.

• AIDS FOR THE PARTIALLY SIGHTED.

Systems that read a document and say what was read.

Automatic "guide dog" navigation systems.

Over the past decade there has been considerable growth in the theoretical base of image understanding (IU) by computer. This article surveys the current state of that theoretical base. As the intellectual climate for progress in IU improved, so funding became available for much needed basic research. Most of the work described in this survey was conducted under the Defense Advanced Research Project Agency's (DARPA) image understanding program at a small number of basic research centers: Carnegie Mellon University, the University of Maryland, Massachusetts Institute of Technology, the University of Rochester, SRI International, Stanford University, the University of Southern California, and Virginia Polytechnic and State University. The DARPA IU program has also produced a number of innovative applications oriented techniques. For reasons of space, these and other applications are omitted from the present discussion.

There is a considerable diversity of approaches to processing visual images by computer. As a result, the boundary between different thrusts is often vague, necessarily so. The characteristic feature of IU is the construction of rich descriptions from an image, an idea that is made more precise in the following pages. Of the many disciplines closely related to IU, four are of particular interest to the computer science community:

image processing, computer graphics, computer aided design and manufacture, and pattern recognition. image processing is primarily concerned with the transmission, storage, enhancement, and restoration of images. There are significant overlaps between IU and image processing, especially in the early processing operations of edge detection and region finding. William K. Pratt's book [PRAT78] is an excellent introduction to the subject. Computer graphics is concerned primarily with the display of visual information. Considerable attention has been given to representing points, edges, surfaces, and volumes to facilitate display. The geometry of perspective and parallel (or orthographic) projection has been studied in detail. Newman and Sproull's [NEWM73] book is a fine introduction. Computer aided design and manufacture (CAD/CAM) also gives attention to surface representations in order to define paths for numerically controlled tools and for making design by traditional techniques such as "lofting" amenable to mathematical analysis. The book by Faux and Pratt [FAUX79] introduces the mathematics of CAD/CAM. Although these three disciplines are closely related to IU, sometimes developing similar representations and uncovering similar constraints, they differ from IU in that they are not concerned *with the **interpretation or understanding of images.***

Pattern recognition is much more closely related to IU. Good introductions are available, including Duda and Hart [DUDA73] and Pavlidis [PAVL78]. The significant differences between IU and pattern recognition are the following:

- pattern recognition systems are concerned typically with recognizing the input as one of a (usually) small set of possibilities. IU aims to construct **rich descriptions** that can not be enumerated in advance but need to be constructed for each individual image. Three dimensional scenes, viewed from an arbitrary location, give rise to a wide variety of occlusion (overlap) relationships. One can hope to compute descriptions of three-dimensional layout but not to recognise it as an instance of one of a small number of stored prototypes.

- pattern recognition systems are mostly concerned with two dimensional images, such as leaf samples or fingerprints. When the images are of three-dimensional objects, such as engine parts, they are effectively treated as two dimensional, by treating each stable position as a separate object. IU has dealt extensively with three dimensional images.

• Most significantly, pattern recognition systems typically operate directly on the image. IU approaches to stereo, texture, shape from shading, indeed most visual processes, operate not on the image but on **symbolic representations** that have been computed by earlier processing such as edge detection.

Before we begin the survey proper, we note some common themes that have crystallized over the past decade.

• *Attention has shifted from restrictions on the domain of application of a vision system to restrictions on visual abilities.*

The most fundamental differences between image understanding as it is now, and as it was a decade ago, stem from the current concentration on topics corresponding to identifiable modules in the human visual system. Substantial progress has been made in, for example, binocular stereo, the extraction of important intensity changes from an image, the interpretation of surface contours, the determination of surface orientation from texture, the computation of motion, and the representation of three-dimensional objects. The focus of current research is defined more narrowly in terms of *visual abilities* than by restricting attention from the start to a *domain of application.* The depth of analysis is correspondingly greater. Increasingly, the progression is from general theoretical developments to specific practical applications. The alternative approach of inferring general principles from work in a limited practical domain is still present, but less so than formerly.

What identifies a particular operation as a distinguishable module in the visual system? Some of the most solid evidence for the claims of individual modules is offered by psychophysical demonstrations of human visual abilities. Care is taken, as far as possible, to isolate a particular source of information and show that the perceptual ability in question survives. One particularly intriguing source of evidence for modules in the human visual system comes from the study of patients with disabilities resulting from brain lesions (for example Weiskrantz, Warrington, Sanders and Marshall [WEIS74], Marshall and Newcombe [MARS 73], Stevens [STEV 76]. Many psychophysical experiments, seemingly isolating particular modules of the human visual system, have been reported in the literature. Notable examples include Gibson's demonstration of the perception of surface shape from texture gradients [GIBS50], Land's demonstration of the computation of

lightness [LAND71], [HORN74], and Julesz's demonstration of stereoscopic fusion without monocular cues [JULE71]. In some cases there is clear evidence of a human perceptual ability, although such evidence would hardly be referred to as psychophysical. Horn's work at MIT considers the highly developed human ability to infer shape from shading [HORN77, WOOD81, IKEU81]. Stevens considers the three-dimensional interpretation of surface contours by humans [STEV81]. On the other hand, it is equally clear that we do not have a specific module in our visual system to recognize "yellow Volkswagens" (see for example [WEIS73]. It is less clear whether we compute depth directly, as opposed to indirectly through integrating over surface orientations, or what use we make of directional selectivity or optical flow.

The change of focus from a narrowly specified domain of application to a particular module of the human visual system has had a number of far-reaching consequences for the way IU research is conducted. One consequence has been a sharp decline in the construction of entire vision systems that mobilize knowledge at all levels, including information specific to some domain of application. In order to complete the construction of such systems, it is almost inevitable that corners be cut and many overly simplified assumptions be made.

- *Representations have been developed that make explicit the information computed by a module.*

A number of representations are discussed in this survey, including the primal sketch, the reflectance map, intrinsic images, normalized texture property maps, and object representations based on generalized cones. A simple observation, which nevertheless has profound consequences, is that not all modules work directly on the image. Indeed, it seems that few do. Instead, they operate on *representations* of the information computed, or made explicit, by other processes. In the case of stereo, Marr and Poggio argue against correlating the intensity information in the left and right images [MARR79b]. Instead, they suggest that edge feature points are matched (see Section 4.1). Baker and Binford, Arnold, and Mayhew and Frisby argue that matching should in fact take place on a different representation, called the *primal sketch* [BAKE81, ARNO78, MAYH81].

Combining this observation with the previous point about modules of the visual system leads to a view of visual perception as the process of constructing instances of a sequence of representations. To each module

there corresponds a representation on which it operates, and a representation that it produces. The first of these representations, and the one whose structure is least subject to dispute, is the image itself. Not surprisingly, most attention has centered on those modules that operate upon the image (section 3). As we shall see, the further we progress up the processing hierarchy, the less secure the story becomes, as the exact structure of the representations becomes more subject to dispute. This is hardly surprising. The image aside, any representation is one module's input and another's output. Computer science teaches us that all of them shape its eventual structure.

For example, several modules of the visual system provide information about the layout of visible surfaces. Stereo provides disparity, from which local shape and relative depth can be computed. Motion, texture, and shading all provide evidence for shape. Barrow and Tenenbaum have suggested that a number of different viewer centered representations make explicit important information associated with surfaces [BARR78]. They call such representations **intrinsic images** and propose specific intrinsic images for depth, motion, surface topography, and color. The name intrinsic images stems from Barrow and Tenebaum's idea that the representations are addressed using the same coordinates as the image. For example the color at an image point whose coordinates are $\mathbf{p}$ might be found in representation C as $C(\mathbf{p})$. Others, notably Marr and Horn have suggested a single representation that makes explicit local surface orientation and discontinuities of depth [MARR78a, HORN82]. The precise details are uncertain at the time of writing.

- *The mathematics of image understanding are becoming more sophisticated.*

Mathematical analyses have been offered for some of the elements of visual perception, such as the relationship between image irradiance and scene radiance, the location of important intensity changes, and motion primitives. In each case, it is observed that the information in the image only partially constrains the interpretation of the image, and further constraints are sought. The additional constraints embody commitments about the way the world is, at least most of the time. For example, the world mostly consists of smooth surfaces, and scenes are mostly viewed from a position free of accidental alignments. Perceptual abilities such as stereopsis, lightness determination, and shape from shading and from texture, require that the appropriate

constraints be uncovered and appropriately expressed.

Most of the analyses to be discussed below begin with a precise description of the representations operated on and produced by the visual process under scrutiny. Increasingly, "precise" means "mathematically precise", as the technical content of image understanding has become steadily more sophisticated. Many observations about the world, as well as our assumptions about it, are naturally articulated in terms of the "smoothness" of some appropriate quantity. This intuitive idea is made mathematically precise in a number of ways in real analysis, for example in conditions for differentiability. Relationships between smoothly varying quantities give rise to differential equations, such as Horn's image Irradiance Equation. We shall discover the value of making the image forming process explicit. This in turn leads to a concern with geometry, such as the properties of the gradient, stereographic, and dual spaces. Combining geometry and smoothness leads naturally to multi-variate vector analysis, and to differential geometry. For the most part, a representation does not of itself contain sufficient information to guarantee that a module can uniquely arrive at the result computed so effortlessly by the human visual system. Additional assumptions, in the form of constraints, are required. This observation has led to application of constraint satisfaction and equation solving techniques from numerical analysis as well as various instantiations of Lagrange multipliers (especially in the form of the calculus of variations).

- *Locally parallel architectures have been developed.*

The majority of the work to be described here had its initial expression in the form of complex computer programs. A common complaint about artificial intelligence in general, and image understanding in particular, used to be that it not only did not run in *real time*, but inherently could not. To the extent that this referred to so-called "heterarchical" programs of the 1970's vintage, this was justified. However, artificial intelligence has been well advised not to make real time performance its most important metric of success, since such a metric often implicitly assumes a particular, usually sequential, model of computation.

Many recent vision algorithms take the form of parallel computations involving local interactions. Once the ideas are fully fixed in software, they are naturally realized in hardware. Davis and Rosenfeld review one

popular class of program structures, called "relaxation" [DAVI81]. In the case of edge finding, one algorithm has been implemented in TTL logic [NISH81], and several others in CCD[NUDD79]. The current rapid pace of developments in VLSI has further motivated research into local parallel programming architectures. It is likely that our concept of computation will change as a result of such developments. Vision will be one of the first areas to benefit from such advances. It seems that it will also be a continuing source of inspiration to VLSI designers [BATA81, NUDD79]. As more sophisticated ideas are embodied in hardware, new applications of image understanding will become feasible.

• *There are growing links between image understanding and theories of human vision.*

For many authors, the changing style of research in image understanding has not been simply a matter of a narrowing of attention and a more highly developed technical content. Instead, greater significance is attached to forging explicit links between IU and psychophysics and neurophysiology. From this perspective, image understanding aims at the construction of computational theories of human visual perception. In large part, this approach stems from a series of papers written by David Marr and his colleagues at MIT. Marr's work derives from a background in neurophysiology, and is expressly addressed to psychophysicists and neurophysiologists, among whom it has excited considerable interest. In particular, it is couched in terms they are accustomed to, and makes extensive reference to their literature, rather than that of computer vision. A book describing Marr's thoughts about human visual perception and incorporating summaries of the contributions he and his colleagues have made across the entire range of the subject is currently in press [MARR82].

It might be imagined that there would be considerable differences of emphasis, subject matter, and technical content between the work of those researchers who see themselves constructing a computational theory of human visual perception and those for whom human visual perception is at most a matter of secondary concern. This turns out not to be the case. For example, the ACRONYM system's representation of objects based upon generalized cones bears many similarities to that proposed by Marr and Nishihara, who relate their work to human perception[BROO79, MARR78]. Again, Horn and Schunck's work on the determination of optical

flow has intriguing similarities to the directional selectivity work of Marr and Ullman that was inspired by neurophysiology [HORN81c, MARR81].

Figure 1 shows some of the representations and modules to be discussed in the remainder of the paper. The figure is intended to make the organization of the paper easier to understand, but it should be treated with caution. The organization implicit in the figure is similar to that given in Barrow and Tenenbaum [BARR81b] and Marr [MARR78]. The representation referred to here as the "surface orientation map" is intended to cover what Marr calls the "$2\frac{1}{2}$D sketch" [MARR78a], Horn calls the "needle map" [HORN82], and Barrow and Tenenbaum call "intrinsic images" [BARR78].

The paper, and hence the figure, is limited in scope. As mentioned above, there is little discussion of applications. There is little if anything about color, and only cursory discussions of motion. The extraction of useful information from color is still extremely rudimentary. Motion has received some attention recently, but findings are preliminary. For example, it is far too early to know what information can be computed reliably from the changing patterns of brightness called the optical flow (see section 3.2). A pervasive view of motion perception is that it arises from temporal changes to the representations that are important for static vision. The Marr-Hildreth theory of edge detection inspired Marr and Ullman's work on directional selectivity, the primal sketch led to Ullman's work on long range motion, and Horn's work on shape from shading underlies the work of Horn and Schunck on the determination of optical flow.

Judged as a flow diagram, figure 1 suggests that the flow of information, and the construction of representations, is entirely sequential, proceeding from the lowest level operations on the image to more semantic higher level operations. Many authors have argued that perceptual processing cannot be so rigidly sequential. They suggest that perception is opportunistic, taking advantage of whatever information becomes available in an image. Natural scenes are normally highly redundant. Gibson [GIBS50] notes approximately 23 distinct cues for determining depth and surface layout, many of which are available in most images. However if only an unpredictable small selection of cues are available, vision is not usually impaired. Only when a single cue is present, as in the laboratory settings of experimental psychology, is our perceptual system easy to fool. Minsky

Figure 1. Some of the representations and modules discussed in the paper.

and Papert [MINS72] suggested that the flexible processing of information by the perceptual system might best be modelled by process interactions. This produced a rash of programs in which relatively high level

knowledge could actively intervene to modify the course of low level processing. Examples include [SHIR73, BAJC75, BAJC76B, TENE77, BRAD78, HANS77, BROO79, SELF81]. Similar "heterarchical" programs were experimented with in speech perception [LESS77]. The performance of such programs did not give cause for unbridled celebration. Some of the associated difficulties are reviewed in [BRAD79].

A rather different kind of flexibility is made available by local parallelism. [WALT72] showed how a variety of cues could be combined to yield an overall interpretation. [DAVI81] stress that an attribute of such process structures is their insensitivity to the sequence in which operations are performed. However, local parallel processes have their own problems. It is easy enough to start local parallel processes going. It is less easy to guarantee that they will stop (but see [HUMM80]), or to be able to make solid assertions about the final state of computation when they do stop. It may be that process structuring will become a key component of image understanding, but currently it is simply too early to be sure. For the moment it seems best to remain agnostic and concentrate on the solid achievements of the past decade, most of which are largely independent of process structuring.

Organization of the paper

In the next section we present a brief review of work in geometrically simple "microworlds". Some of the generally important ideas developed initially for the blocks world of line drawings of polyhedra are introduced. Kanade's extension to the world of origami, and Barrow and Tenenbaum's work on curved "play dough" figures is mentioned.

Section 3, by far the longest in the paper, discusses modules that operate directly upon the image. Subsection 3.1 concerns edge finding, 3.2 the determination of shape from shading, 3.3 texture, and 3.4 segmentation.

Section 4 discusses modules that operate on the output of section 3, which, following [MARR76a], we call the primal sketch. Subsection 4.1 discusses stereo, 4.2 shape from contour, 4.3 shape from texture and Kender's generalization to "shape from you name it". Finally, subsection 4.4 briefly discusses shape from motion.

Sections 5 and 6 discuss modules that operate on surface orientations and viewpoint independent representations.

2. Review of work on geometrically simple microworlds

Beginning with the seminal work of [ROBE62], much early attention of IU was devoted to interpreting line drawings of polyhedra automatically. This work marked a significant break from pattern recognition in that it emphasized **descriptions** of the objects present in a scene and the spatial relationships between them. For example, figure 2 might be described as a cube standing in front of a block. Clowes and Huffman stressed that the relationship between a scene and its image needs to be made explicit [CLOW71, HUFF71]. A line is the image of the edge of a polyhedron in the scene. They noted that lines can be labelled as convex, concave, or occluding(figure 3a). The interpretation of a line can not change along its length. A junction is the image of a three-dimensional vertex. Enumeration of the local volumes occupied by vertices, and the appearance of such vertices from all possible viewpoints gives rise to a set of labellings for junctions (figure 3b). Vertex labellings embody a local constraint: although there are three lines forming an arrow junction, and each line has four possible interpretations (counting the two senses of occlusion separately), there are not $4^3 = 64$ physically realizable labellings for an arrow vertex but only 3. Notice that every interpretation of a T-junction is assumed to signal an occlusion of the stem. Conversely, every scene occlusion gives rise to a T-junction. The constraints local to each junction propagate along the lines that connect them to adjacent junctions, possibly rendering some of the initial set of labellings at both junctions impossible. Clowes determined consistent interpretations by a search space technique. Surprisingly, many simple line drawings have many consistent interpretations, though occlusion often resolves ambiguity.

Despite the geometric restrictions imposed by Huffman and Clowes, their scheme had limited competence. First, as Kanade pointed out, the Huffman-Clowes scheme was essentially **qualitative** in that it could not distinguish between the truncated pyramid shown in figure 4a and the cube shown in figure 4b [KANA81]. Human perception is at least partly quantitative since we readily assign slopes to line drawn surfaces and

Figure 2. A typical line drawing of polyhedra studied by Huffman and Clowes.

estimate rectangularity of vertices from junctions. Since the line drawing in figure 4b can be the image of an infinite set of scenes, it is more precise to say that the Huffman-Clowes scheme could not determine that figure 4a has no interpretation for which vertex A is rectangular while figure 4b does. It is also interesting to ask why the cube is perceived as a cube. One proposal, due to Kanade, is sketched below.

A second manifestation of the qualitative nature of the Huffman-Clowes scheme is its inability to detect the impossibility of the line drawing shown in figure 5. Huffman's paper was principally concerned with "impossible objects" (such as that depicted in figure 5), and the consequent need for a more expressive representation. He proposed a representation called dual space and an orthographic projection of it called the dual picture graph. Mackworth [MACK73] developed the idea of a representation of surface shape further by introducing **gradient** space, an idea that was developed in [DRAP80, DRAP81, HORN77, KANA80, KANA81, KLND80, HUFF77, SUGI78, SUGI81].

Figure 3. a. The possible interpretations of an image line. b. The possible interpretations of a trihedral vertex.

Consider the imaging geometry depicted in figure 6: a surface $f(x, y) - z = 0$ is viewed from a great distance along the negative z-axis. Applying the chain rule,

Figure 4. The Huffman-Clowes scheme could not distinguish these line drawings.

$$\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy - dz = 0,$$

that is

$$(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1).(dx, dy, dz) = 0,$$

so that $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1)$ are the direction ratios of the surface normal or gradient. It is customary to denote $\frac{\partial f}{\partial x}$ by p and $\frac{\partial f}{\partial y}$ by q. The coordinate frame based on (p, q) is called **gradient space**. As an example consider a planar facet $ax + by + c - z = 0$. The gradient has $p = a, q = b$. The origin of gradient space corresponds to surface facets that point directly at the viewer. Moving away from the origin, it is easy to show that $(p^2 + q^2)^{\frac{1}{2}}$ is the slant of the surface normal. The angle τ whose tangent is q/p is the tilt of the surface normal (figure 1).

Figure 5. The Huffman-Clowes scheme could not determine that this line drawing depicts an "impossible object".

The coordinates can be aligned so that a vector $(x, y, z) = \underline{v}$ projects to $(x, y) = \underline{k} \times (\underline{v} \times \underline{k})$, where $\underline{k}$ is the unit vector in the z direction. In particular, the gradient vector $(p, q, -1)$ projects to (p, q). Suppose two planes P_1 and P_2 have surface normals $(p_i, q_i, -1)$, and suppose that they meet in a space vector $\underline{v}$. It is easy to show that the image l of $\underline{v}$ is perpendicular to the dual line connecting $g_1 = (p_1, q_1)$ to $g_2 = (p_2, q_2)$ [MACK73]. Furthermore, $\underline{v}$ is convex if and only if the order of the g_i across l is the same as the order of the images of P_i across l (figure 8). Mackworth exploited this observation in a program that was capable of determining the impossibility of the notched tetrahedron shown in figure 5. However, Mackworth's triangulation solution scheme could not determine the impossibility of the notched cube also shown in figure 5 [MACK73]. Draper [DRAP81] has analyzed the competence of Mackworth's gradient space scheme and an extension due to Huffman based on "dual space" [HUFF77].

The notched cube of figure 5 illustrates an assumption discussed by Kanade [KANA81], namely *lines that*

Figure 6. Viewing geometry for defining gradient space.

are parallel in the image are the images of vectors that are parallel in space. If lines l_1 and l_2 are the images of scene vectors $\underline{v}_1$ and $\underline{v}_2$, then it is easy to show that l_1 is parallel to l_2 if and only if the triple scalar product $[\underline{v}_1, \underline{v}_2, \underline{k}]$ is zero. It follows that Kanade's parallel line assumption fails only when $\underline{v}_1$, $\underline{v}_2$, and $\underline{k}$ are coplanar. Generally, people find it difficult to interpret such foreshortened figures properly [MARR78b, MARR78a].

Kanade [KANA81] has also studied an interesting assumption involving what he calls "skew-symmetry". Consider figures 9a, 9b and 9c. All three are interpreted as symmetric, planar figures viewed obliquely. As figure 9d shows, a skew symmetry defines two directions: the image of the axis of symmetry, called the skewed symmetry axis, and the image of the normal to the axis of symmetry that lies in the plane of the figure, called the skewed transverse axis. Skew symmetries feature prominently on the cube and truncated pyramid shown in figure 4. Kanade proposes that a skewed symmetry is always interpreted as the image of a real symmetry viewed obliquely. This assumption gives rise to a constraint, expressed in terms of the angles α and β defined

Figure 7. Slant and tilt in gradient space.

in figure 9d, relating the possible gradients of the surface containing the real symmetry. In fact, the possible gradients form the hyperbola shown in figure 10. Notice that the possible planes with least slant (the tips of the hyperbola) have a normal that projects into the bisector of the skewed symmetry axis and the skewed transverse axis. This accords with a heuristic finding of Stevens [STEV80].

It is important to realize that the parallelism and skew-symmetry assumptions apply beyond the blocks world. Kanade has shown how they can be combined with Huffman-Clowes style labelling and Mackworth-style algebraic analysis to give both a quantitative and a qualitative interpretation of line drawings in the microworlds of blocks and origami constructions [KANA81].

The junction labelling constraints of Huffman and Clowes are essentially local. The constraints of surface planarity, skew symmetry, and parallelism are less local and support more competent programs. However, none of the constraints are global in the sense that they apply simultaneously to all parts of the image. Waltz

Figure 8. Convexity preserves order across the gradient line.

investigated the global constraint afforded by the shadows cast by a single distant light source [WALT72]. The number of interpretations of a line rose from 4 to 12, with a consequent massive number of possible junction labellings. As Draper has pointed out the large (and probably unverified) labelling sets would be considerably larger without the assumption of general position of the viewer [DRAP80]. Waltz's line labels incorporate information about the surface geometry, illumination, and surface-object boundaries. The huge label sets precluded a tree search of the sort used by Clowes [CLOW71]. Instead, Waltz designed a filter program, potentially capable of running as a local parallel program, that usually converged to a single labelling in near linear time. The Waltz filter accelerated investigation of local parallelism. Line labelling is discussed by [ZUCK77, ZUCK81, HUMM80]. Waltz's program reaffirmed the value of redundancy when processing can make appropriate use of it. However, the complex line labellings confounded too much information from different levels of the visual system in an impoverished representation.

Figure 9. Skewed symmetry. a-c: examples of skew symmetry. d. definition of skewed-symmetry axis and skewed transverse axis. (Reproduced from [KANA81], figure 16)

The figures discussed in this section have all been images of objects with planar surfaces. Some authors have tried to relax this restriction. One difficulty with drawings of curved surfaces is that one of the basic assumptions of the Huffman-Clowes work no longer holds: a line can change its interpretation from one end to the other [HUFF71]. Turner [TURN74] noted that such changes of interpretation are not arbitrary, and he allowed a small number of transformations of a line label to arrive at an interpretation. Recently, Binford

Figure 10. A skewed symmetry defined by the angles α and β can be the projection of a real symmetry on a plane whose gradient is (p, q) if and only if the gradient lies on the hyperbola shown.(Reproduced from [KANA81], figure 17)

[BINF81] and Lowe and Binford [LOWE81] have suggested more general interpretations of curved lines that may enable labelling techniques to be extended to line drawings of arbitrarily curved surfaces (see also section 3.1.3).

Barrow and Tenenbaum [BARR78] have also studied a microworld of curved objects. They combine line

labelling techniques with Horn's work on shape from shading (see section 3.2) to interpret idealized images of "play dough" scenes.

Work in geometrically simple microworlds has played an important role in the development of image understanding. From the pioneering work of Roberts, Clowes, and Huffman to the present day, the goal has been to generate *descriptions* rather than transformed or classified images. The key has been to make the relationships between the scene and the image explicit. Examples include the interpretations of image lines as visible edges, and the analyses of skew symmetry and parallelism. Mackworth's development of gradient space points up the need for rich representations. Finally, Waltz's work shows that redundancy can be exploited by appropriate computing mechanisms.

Microworlds also set traps. It is irresistably tempting to deploy domain specific information at the earliest opportunity. Planar objects have a number of global properties that are not enjoyed by curved objects. For example, two planes intersect along a single straight edge in space, so that from any given viewpoint, one plane is always in front of the other on one side of the image of the edge, and always behind it on the other [DRAP81]. The labelling schemes of Huffman, Clowes, and Waltz, extended to idealised images of curved objects with reflectance patches and shadows, produce a vast number of labels that confound many distinct sources of information in a single label. It seems more fruitful to attempt to tease out the information provided by each of these sources separately.

3. Modules that operate on the image

3.1 Edge detection

A great deal of effort has been devoted to understanding how the significant intensity changes in an image can be extracted, and how the resultant information can best be represented. Marr coined the term *primal sketch* to describe such a representation [MARR76a]. Significant intensity changes correspond to a variety of events in a scene, such as depth, reflectance, and shadow boundaries, as well as discontinuities in

surface orientation. The image intensities I(x,y) form a surface that is a discrete approximation to one that is continuous nearly everywhere [ROSE76, PRAT79]. Quantization and sensor noise of various sorts complicate the formulation of a predicate that can completely reliably determine which intensity changes correspond to perceptible scene events (that is, which are "significant").

It has been observed repeatedly over the past twenty years that intensity changes correspond to maxima of the gradient of the image surface, equivalently a place at which the second derivative crosses zero and changes sign. Many local operators have been developed to approximate first and second directional derivatives by first and second differences. A representative sample is shown in figure 11. Mostly, such operators were developed and tuned for a limited domain of application.

Figure 12 shows an idealized step change in intensity and the response of first and second difference operators. In practice, gradient operators tend to produce a large response over a broad region flanking an edge (see figure 14, also [BINF81]), especially with intensity changes other than steps. As a result, feature points from a gradient operator have to be thinned, a process that makes it difficult to localize the position of the edge as accurately as with second difference operators. On the other hand, errors grow rapidly as differences are taken, so that second differences are much noisier than first differences.

A recent edge finder, which appears to work well on a range of natural images, is due to Nevatia and Babu [NEVA78]. It applies the six gradient operators shown in figure 13 to each point of an image and chooses the one giving the best response if (1) it is high enough and (2) it is not dominated by the responses at neighboring points in a direction which is normal to the same apparent edge. This process is followed by thinning, thresholding, and line fitting. Some indication of the performance of the Nevatia-Babu algorithm can be seen in figure 14.

Binford has argued that it is important to distinguish between the **detection** of an intensity change and its subsequent **localization** [BINF81]. He suggests that a maximum of a noisy signal is good for detecting change but not for isolation. Conversely, a zero crossing is ideal for localizing change but not for detection. MacVicar-Whelan and Binford find adjacent pixels between which a second differencing-like operator changes

Figure 11. A selection of masks from the image understanding literature used to compute approximations to the first derivative of an image in the x direction.

sign [MACV81]. Using linear interpolation they claim to be able to localize intensity changes with sub-pixel accuracy. Sub-pixel accuracy is also claimed by [MARR79] in the context of vernier acuity, where the eye is

Figure 12. The response of an edge and bar operator to an ideal step change in intensity. a. The intensity change. b. The response of a typical first difference edge operator such as that shown in figure 11a. c. The response of a typical bar operator such as that shown in figure 11e.

able to perceive breaks in lines that are more closely spaced than the physiology of the eye would seem to permit [MARR79].

Real images are further complicated by defocussing and the frequent occurence of slow intensity gradients across large portions of the image. Humans are largely unaware of slow linear intensity gradients [LAND71, MCCA74]. This seems to be because of "lateral inhibition", where the image is processed by "center surround" operators (figure 15) that resemble rotationally symmetric second differential operators.

Herskovits and Binford [HERS70] proposed an early taxonomy for the intensity changes they found in images of polyhedra, classifying them as "step", "roof", or "edge" changes (figure 16). As we shall elaborate below, they proposed different operators F_{step}, F_{roof}, and F_{edge} to detect each different type of intensity change. It is commonly supposed, especially in applications where scenes are effectively flat, that the majority of intensity changes are of the simple step type. Many detection schemes are predicated upon this assumption.

Figure 13. The masks used by [Nevu78] to compute first derivatives of an image at 30 degree intervals.

Figure 14. Sample results of running the Nevatia and Babu operator over a natural image.

Figure 15. A center surround operator.

Herskovits and Binford [HERS70] and Horn [HORN77] observe that step edges typically correspond to depth or reflectance boundaries, whereas the equally important class of intensity changes corresponding to surface orientation discontinuities often give rise to roof and edge transitions. Marr refined the Herskovits and Binford classification to include "extended edge", and "thin and wide bar" (figure 17) and proposed a variety of operators of different sizes to discriminate between them [MARR76a].

The construction of a primal sketch representation from an image has three distinguishable stages: (1) "feature points" are detected at which the intensity change is deemed to be significant; (2) feature points are grouped to form line segments, or small closed contours; (3) these line segments are interpreted as scene events, say as bounding contours or as true edges of visible surfaces. These three stages are discussed in turn in the following subsections.

The operators shown in figure 11 are directionally selective. Some authors have proposed the use of rota-

Figure 16. The taxonomy of intensity profiles proposed by Herskovits and Binford. a. idealization b. examples.

tionally symmetric operators, such as the Laplacian Δ, for edge detection [BRAD81b]. Several reasons have been advanced. Some authors prefer theoretical arguments, noting the (near) isotropy of human vision and the fact that the center surround operators giving lateral inhibition are rotationally symmetric. Others have stressed practical considerations. For example, in her discussion of the Marr-Hildreth theory of edge detection (to be discussed in section 3.1.1), Hildreth [HILD80,page 13] notes that "a number of practical considerations, which will be illuminated in the discussion of the implementation, suggested that the ... operators not be directional". Suppose instead that directional operators are used. Most algorithms for finding feature points have two stages: first, the image is convolved with directional operators in "sufficiently many" directions, and second, the outputs are combined to determine the orientation and extent of intensity changes. Regarding the first stage, both Marr and Hildreth [MARR80a, page 193] and Hildreth [HILD80, page 40] comment on the cost of convolving with a "sufficient" number of operators. They show that a single rotationally sym-

Figure 17. Marr's classification of the intensity changes that occur in natural images. After figure 2 of [MARR76a]

metric operator (the Laplacian) gives precisely the same results if a condition called "linear variation" holds. Regarding the second stage, Hildreth [HILD80, page 36] observes that edges in a direction close to that of the mask are elongated ("smeared") in the direction of the mask. She also notes that operators at several orientations give significant responses to any given edge, and that combining the responses is non-trivial. Other authors are less convinced of the need for rotationally symmetric operators for edge finding [BINF81].

The issue of control arises in edge finding as it does in all other areas of image understanding. It has been argued that it is not possible to find significant intensity changes, group them, or interpret them without engaging quite high level knowledge. Bajcsy and Tavakoli [BAJC75, BAJC76B] were early proponents of this view, as was Shirai [SHIR73]. Davis and Rosenfeld survey the application of relaxation processing to isolate feature points [DAVI81].

3.1.1. Finding feature points

Although many of the published schemes for detecting and isolating feature points were discovered empirically, there have been three main approaches to making edge finding more precise. The first consists of locally modelling the image by a parameterized analytic surface and determining the best fitting choice of parameters given the actual intensity distribution. The second is Binford's application of signal theory to edge finding. Finally, Marr [MARR76a] and Marr and Hildreth [MARR80] have developed a theory of edge finding in the human visual system that takes account of neurophysiology and psychophysics. We discuss each of these approaches in turn.

Surface fitting

The derivation of operators to approximate first and second differences by least squares surface fitting was introduced by Prewitt [PREW70], and Hueckel [HUEC71]. [BROO78, HUMM79, HARA80] give good introductions to the method. In the simplest case, where noise considerations are ignored, two things must be chosen: (1) the size of the local neighborhood or window in which the surface will be fit, and (2) the function to approximate the image surface in the window. For simplicity, we choose a window of size 2 by 2 and approximate the image surface in such a window by a plane $P(x, y) = ax + by + c$. Haralick [HARA80] calls this the "sloped facet" model. Assuming that the response of an edge operator is independent of the choice of coordinate origin, we assume that the window covers $x = 0, 1; y = 0, 1$ (figure 18). We determine the best fitting choice of parameters a, b and c by least squares minimization of the difference between the intensity values actually found in the window and those predicted by the function $P(x, y)$. The square of this difference is given by

$$\epsilon^2 = (a + b + c - I(1, 1))^2 + (a + c - I(1, 0))^2 + (b + c - I(0, 1))^2 + (c - I(0, 0))^2).$$

For a least squares fit, we first set

$$\frac{\partial \epsilon^2}{\partial a} = 0.$$

This implies

$$2a + b + 2c = I(1,1) + I(1,0).$$

Similarly, setting $\frac{\partial \epsilon^2}{\partial b}$ and $\frac{\partial \epsilon^2}{\partial c}$ equal to zero, we get

$$a + 2b + 2c = I(1,1) + I(0,1),$$

and

$$2a + 2b + 4c = I(0,0) + I(1,0) + I(0,1) + I(1,1).$$

Solving, we see that

$$2a = I(1,1) + I(1,0) - I(0,1) - I(0,0),$$

and

$$2b = I(1,1) + I(0,1) - I(1,0) - I(0,0).$$

The gradient of $P(x,y)$ in the x-direction is $\frac{\partial P(x,y)}{\partial x} = a$. Similarly, $\frac{\partial P(x,y)}{\partial y} = b$. We can depict the gradient operators a and b as in figure 18.

Haralick has extended the basic scheme illustrated above to model the effect of sensor noise [HARA80]. He adds a normally distributed noise term $\eta(x,y)$ to the function $P(x,y)$ and shows that an F-test is appropriate for deciding whether or not there is a significant change in the slope of adjacent sloped facets. Here "significant" is given its usual 1% statistical meaning.

Figure 18. a. The 2 by 2 window covering pixels (0,0) to (1,1). b and c. The gradient operators that result from best fitting a plane ("sloped facet") in the window shown in a.

Brooks [BROO78] considers fitting planes and quadratics to 3 by 3 windows. The best fit plane gives the Prewitt operator shown in figure 11, and the second derivative of the best fit quadratic gives the bar mask shown in figure 11. Brooks observes that the dot product of the gradient operators a and b in figure 18 is zero. This suggests that it may be possible to develop an orthogonal set of increasingly higher order masks. One natural choice for such an orthogonal set is the set of Fourier basis functions. Other choices are Walsh or Hadamard functions. The best fitting choice of Fourier basis functions was developed by Hueckel in an early application of the function fitting idea [HUEC71]. O'Gorman proposed the use of best fitting Walsh functions [OGOR76].

Binford's signal theory approach

Recently, Binford [BINF81] has outlined an approach to edge finding that has its roots in two early unpublished papers [HERS70, HORN73]. The details are not completely clear and would be a valuable addition

to the literature. It was noted above that image noise makes it difficult to determine reliably which intensity changes are significant. Herskovits and Binford showed how to estimate the signal to noise ratio for an image, and determined that the error is typically about 1% for a zero signal. They studied intensity profiles in scenes of polyhedra and proposed the classification shown in figure 16. The response of a bar mask to an ideal step edge is shown in figure 19 (see also [MARR76a]. Clearly, as the number of points in the bar mask increases, the operator can detect steps of lesser heights more reliably. Herskovits and Binford make this idea more precise by defining the sensitivity of an operator as the signal for which detection is 50% successful.

The intensity values determined by sensors are most reliable in the middle range. Accordingly, Herskovits and Binford [HERS70, page 36] suggest upper and lower thresholds u and l on intensity. The ideal step gives rise to a band of u's flanked by a band of l's. Define L to be the number of points at which the value is u in the left band minus the number of points at which the thresholded intensity is l. Similarly, R is the number of points in the right band at which the thresholded value is u minus the number at which the value is l. If $F_{step} = L - R$ is big enough, a local maximum is found. In this way the step is detected though not localized.

Figure 19 also shows the response of a bar mask to an ideal roof intensity change. Note that unlike step changes, the response reaches a maximum in the vicinity of the top of the roof. Accordingly an operator F_{roof} is defined as the difference $R + L$, that is the difference between the number of values u's and l's summed over both bands.

A refinement of the scheme is described in [BINF81]. The operator F_{step} approximates the derivative of the second derivative, or equivalently, detects the step intensity change by looking at the third derivative of intensity. The intensity change is then localized from the zero crossing of the second derivative. A roof change is detected from the maximum of the second derivative and localized from the zero crossing of the third derivative.

The operators F_{step}, F_{roof}, and a similar one for "edge effects" were incorporated in the Binford-Horn line finder [HORN73] and discussed retrospectively in [BINF81].

Marr's approach to edge detection by the human visual system

Figure 19. Response of a bar mask to an ideal step (a) and roof edge (b). 1. The intensity change. 2. Response to a lateral inhibition operator. 3. Derivative of 2.

A novel feature of Marr's development of the primal sketch [MARR76a] was its direct reference to neurophysiology and psychophysics, a commitment Marr continued to stress in later work. Marr's algorithm

for computing the primal sketch from an image had a number of interesting features. First, being inspired by neurophysiology, Marr applied the findings of Hubel, Wiesel, Barlow, and others, which seem to suggest that an early stage in the processing of visual information consists of convolving the image with edge and bar masks. As we observed above, such masks signal an approximation to the first and second (directional) derivatives of the intensity function. Marr based his algorithm on an analysis of the response of bar and edge masks to ideal instances of the scene events that give rise to intensity changes. The algorithm itself consisted of convolving an image with a number of edge and bar masks and then "parsing" the results by comparing the actual responses to those predicted for ideal scene events. It was noted that bar masks seemed to give more reliable information than edge masks, an observation whose explanation awaited the later development of ΔG operators which have a similar cross section (see below). The algorithm convolved the image with masks of different panel widths. Although the later justification for this would be in terms of separate processing channels, the original explanation was based on the need for noise reduction, although this idea was never formulated precisely. In any case, the outputs of the individual channels were combined, not only to reduce the effects of noise, but to compute measures such as the "fuzziness" of an edge. The idea of combining the outputs of independent channels remains an important goal of the work on zero crossings, but, with the singular exception of stereo (see below), it has not yet been worked out.

Marr and Hildreth [MARR80, page 189] point out that "a major difficulty with natural images is that changes can and do occur over a wide range of scales, so it follows that one should seek a way of dealing with the changes occuring at different scales." One way to do this, which has been proposed several times in the image processing literature, is to pass the image through a number of band limited filters. The difficult issues raised by the idea concern the choice of filters (bar mask, Fourier, Gaussian), the number of them, and the exact band pass characteristics of each.

Intensity changes are localized in space, a fact which derives from their physical causes [HORN77, MARR76, MARR80a]. Marr and Hildreth argue that they are also localized in the frequency domain. Marr and Hildreth [MARR80, page 191] note that "unfortunately, these two localization requirements, the one in

the spatial and the other in the frequency domain, are conflicting". The Fourier transform of a bar mask has components of arbitrarily high frequency. Similarly, the inverse transform of a bar-like band pass filter in the Fourier domain has significant "echoes"; [HILD80] gives examples. They point out that a Gaussian filter optimizes localization in both domains simultaneously, and so it is chosen as the band limiting filter in their theory.

For the practical considerations given in the introduction to this section, Marr and Hildreth propose the use of a rotationally symmetric operator to find feature points. An obvious candidate is the Laplacian Δ (see [BRAD81] for a discussion of rotationally symmetric operators). The Marr and Hildreth approach to edge finding follows Gaussian smoothing by convolving the image with a Laplacian, thus isolating the positions of zero crossings. In fact, by the convolution theorem [BRAC65, page 118],

$$\Delta(G * image) = (\Delta G) * image,$$

where G is a Gaussian operator, and * denotes convolution. Marr and Hildreth [MARR80, page 193] point out that the ΔG operator closely resembles the difference of Gaussian (DOG) operators proposed by Wilson and Giese [WILS77] (see also [WILS79]). Indeed they show that ΔG is the limit of a DOG, and that the DOG closely approximates it. The two-dimensional cross section of the ΔG operator is shown in figure 20a. It can be thought of as a smoothed version of a bar mask cross section, and may explain Marr's heuristic preference for bar masks over edge masks mentioned earlier. Wilson and Bergen's work suggests that there should be four bandpass channels at each retinal eccentricity, and that their characteristic sizes should scale linearly with eccentricity, being smallest in the fovea and doubling in size by about 4°.

Shanmugam, Dickey, and Green investigated the characteristics of the optimal frequency domain filter for edge detection [SHAN79]. By "optimal" they mean the filter that produces the maximum energy in the vicinity of the location of a (step) edge. Jernigan and Wardell [JERN81] have shown that there is no significant difference between the optimizing filter derived by Shanmugam, Dickey, and Green, and the difference of Gaussian filter proposed by Wilson and Bergen. The characteristics of the Shanmugan, Dickey and Green

filter are largely determined by a constant c that is the product of the frequency domain bandwidth of the optimal filter and its spatial interval. As c increases, the signal to noise ratio increases. However, for fixed bandwidth, the improved signal to noise ratio is achieved at the expense of resolution.

Recently, Marr, Hildreth, and Poggio have noted evidence for a fifth, smaller channel in the fovea [MARR79a]. Brady [BRAD80a] has shown how the Marr-Hildreth theory can be used to explain a number of psychophysical results about parafoveal processing in reading.

Figure 21 shows images of a leaf and a coffee jar which has been sprayed with black paint to provide a textured surface for stereoscopic fusion (see below). Figures 22 and 23 show the images in figure 21 filtered respectively through the coarsest and finest resolution channels in the fovea. Figure 24 shows the zero crossings of the Laplacian applied to the filtered images shown in figures 22 and 23.

One of the novel aspects of the implementation of the theory concerns the sizes of the ΔG operators. Edge finding operators are typically at most 7 pixels square; the *smallest* operator used in the implementation of the Marr-Hildreth theory at MIT is 35 pixels square. Not only are the resulting operators much closer approximations to the Gaussian (or any other filter for that matter), but the signal to noise characteristics of the smoothed images is vastly improved. One practical consequence of this seems to be that for computing the orientation of visible edges one can approximate differential operators by simple difference operators. Conventional edge finding operators confound filtering and differentiation, and have poor and essentially unpredictable filter characteristics. The first implemented version of the Marr-Hildreth theory took on the order of three hours to compute the zero crossings in the coarse channel of an image 512 pixels square. A prototype hardware implementation reduced this to 30 minutes. Nishihara and Larson report a TTL implementation that computes and displays the zero crossings in any channel of an image 128 pixels square in under 0.25 seconds [NISH81].

Directional selectivity for motion

Marr and Ullman [MARR81] investigate the possibility that the time rate of change of

Figure 20. (a) Two dimensional cross section of the ΔG operator, showing its resemblance to the center surround operators in the human fovea. (b) The cross section of a typical bar mask used by [MARR76a].

$$S(x, y, t) = (\Delta G)^* I(x, y, t).$$

Figure 21. images of (a) a leaf and (b) a coffee jar sprayed to produce a textured surface. (Reproduced from (a) [HELI80] and (b) [GRIM80])

Figure 22. The result of bandpass filtering the images shown in figure 21 to simulate the information available through the current channel in the human fovea.

Figure 23. The result of bandpass filtering the images shown in figure 21 to simulate the information available through the finest channel in the human fovea.

Figure 24. The edges isolated in the images shown in figures 22 and 23

can enable one to detect the direction of motion of zero-crossings. Define

$$T(x, y, t) = \frac{\partial S(x, y, t)}{\partial t},$$

so that

$$T(x, y, t) = \Delta G^* \frac{\partial I(x, y, t)}{\partial t}.$$

Figure 25 is based on [MARR81, figure 3]. It shows the response of $S(x, y, t)$ and $T(x, y, t)$ in the vicinity of an isolated intensity edge. Notice that for motion to the right, $T(x, y, t)$ is positive at the zero crossing, while for motion to the left it is negative. Marr and Ullman propose that motion to the right can be detected by the simultaneous activity of S^+, T^+, and S^-. On the basis of this analysis they find close agreement at moderate speeds between theoretical predictions and cell recordings (see figure 15). Richter and Ullman [RICH80] have accounted for the discrepancy at high speeds, and generally refined the model of directional selectivity, by noting that the two Gaussians whose difference approximates ΔG act like RC filters, composed of a resistor and a capacitor, with different time constants. This causes a slight delay in the onset of the negative outer part relative to the positive central part. Richter and Ullman's predictions show remarkable agreement with cell recordings for a wide variety of stimuli (see figure 26). Coincidentally, Richter and Ullman have proposed a theoretical structure for the outer plexiform layer of the human retina in which ΔG is computed. This suggests a particular VLSI implementation of ΔG. The general scheme is illustrated in figure 27.

3.1.2 Grouping feature points.

The methods of the previous section produce a set of feature points (figure 28) corresponding to places in the image at which the intensity change is considered significant. The next stage of processing imposes structure on the sea of individuated feature points by grouping them to form extended contours. Marr [MARR76,

Figure 25. Derivation of the STS operator proposed by Marr and Ullman for computing directional selectivity of motion. (a) The response of a vertical contrast boundary at time t to a ΔG operator, showing the position z of the zero crossing. (b) At time (t+dt) the edge has moved slightly to the right. Subtracting yields an approximation to T(x,y,t). Notice that T is positive at z. (c) analogously, an edge moving to the left is detected by a negative value for T at z. (Reproduced from [MARR81, figure 3]

Figure 26. Comparison of theoretical predictions for the response of an X-ganglion cell to moving stimuli using the models of Marr-Ullman and Richter-Ullman, and actual cell recordings. (a) Response curves taken from the neurophysiology literature for an edge, a wide bar, and a thin bar. (c) Theoretical predictions by the Marr-Ullman model. (b) Predictions by the Richter-Ullman model. (Reproduced from [RICH80, figure 13]

Figure 27. Spatial formation of a midget bipolar receptive field in the Richter-Ullman model. (a) The arrangement of cones and horizontal cells. Each horizontal cell covers a circle (the shaded area) with a radius three times larger than the cone pedicle (the dots). It contacts 7 cones. Thus seven horizontal cells contact each cone, connecting a total of 19 cones to create the surround area of a midget bipolar cell. (b) The contribution to the surround of the first, second, and third ring of cones. The receptive field of a midget bipolar cell resulting from the center contribution of one cone and the above surround is shown in 5 and a slice through its center is shown in 6. (figure reproduced from [RIC1180, figure 3]

page 501] argues that "grouping processes are available precisely because they are needed to help interpret the primal sketch; and furthermore that these symbolic processes, together with first order discriminations, operating recursively on the description of the primal sketch, are sufficient to account for most of the range of 'non-attentive' vision of which we are capable."

We may assume that there are few accidental alignments of object boundaries, shadows, reflectance boundaries, and surface discontinuities (also called "true edges") in the scene, that is, the image is taken from "general position". Then nearby feature points mostly arise from nearby scene points and for the same underlying physical cause. It follows that the descriptions associated with adjacent feature points that are perceptually grouped are very similar. If feature points have reliable and rich descriptions, perceptual grouping can be more effective. Similar considerations apply to other cases of local matching in vision such as stereo, motion computation, and the determination of texture.

Each of the methods for finding feature points described in the previous section has associated grouping processes. For example the Binford-Horn line finder compares feature points locally on the basis of the size of the contrast step across the intensity change, the type of intensity change, and the slope of the gradient [HORN73, page 7]. Marr [MARR76, page 503] also groups feature points on the basis of "orientation, contrast, type(EDGE, LINE, etc.), and fuzziness". He notes that "the first stage of grouping combines two elements only if they match in almost all respects, are very close to one another, and if there are no other candidates." Typical results of this process are shown in figures 29 and 30. Marr proposes a number of operations that group the short line segments produced by the first stage on the basis of collinearity, proximity, and similarity of slope [MARR76a]. The results of these operations are histogrammed locally and the dominant structures made explicit. Figure 29b shows the herring bone stripes computed from figure 29.

Many images contain extended straight contours, mostly corresponding to the straight edges that prevail in our man-made environment. Duda and Hart [DUDA73] and O'Gorman and Clowes [OGOR73] popularized a method introduced by Hough for finding straight lines in images. Ballard [BALL79] has extended the method considerably, and we follow his development here. Suppose that one is interested in discovering

instances of circles in an image. Ballard proposes to find the circles from the feature points that form their contours. Let there be a feature point at point (x, y), and suppose that the gradient of the intensity change is in direction θ. A circle is uniquely specified by three parameters: its center (a, b) and its radius r. To pass through the feature point (x, y), such a circle has to satisfy the constraint

$$(x - a)^2 + (y - b)^2 = r^2.$$

The gradient slope imposes the additional constraint $r = (y - b)\sec\theta$. It follows that each feature point constrains the circles passing through it with the given slope to a one parameter family. As before, adjacent feature points normally come from the same circle. There are two simple techniques for combining the additional constraint. First, one might intersect the one parameter families in the spirit of line labelling (see section 2). The noise inherent in the measurement of the center and radius suggests that something akin to a relaxation technique be used to find optimal circles. Several authors have suggested such an approach [ZUCK77, DAVI81]. Line labelling essentially combines evidence by an AND operation. Alternatively an OR operation can be used, corresponding to a summation or histogram. To accommodate noise, the range of possible values for the center and radius are quantized for each parameter to produce an "accumulator array". Each feature point contributes one vote to the (a_i, b_j, r_k) buckets in its one parameter family. Local maxima in the accumulator array are assumed to correspond to instances of circles.

Ballard has extended the Hough transform technique of combining constraints on defining parameter values to non-analytic functions and has shown how to estimate the effects of noise [BALL81].

3.1.3 Interpreting feature point segments as scene events

In the discussion of the microworlds in section 2, we noted the key contribution of Clowes and Huffman who stressed the need to make explicit the relationship between image fragments and scene events. The line labelling schemes of Huffman, Clowes, Kanade, Sugihara, and Waltz, and the surface labelling schemes of Mackworth, Huffman, and Draper all developed this fundamental idea. Generalizing from the blocks world,

Figure 28. image of a leaf and the feature points found in it using the Marr-Hildreth theory of edge detection. (Reproduced from [HILD80, figure 3]

Turner and Barrow and Tenenbaum developed labelling schemes that made explicit the possible interpretations of edges and surfaces in their microworlds.

Figure 29. image of a piece of herring-bone cloth and typical stripes extracted from it on the basis of slope of gradient at feature points. (Reproduced from [MARR76a, figure 19])

One would like to extend line interpretation to feature point segments. Elongated segments correspond to boundaries that mark important scene events: that is why feature points were isolated in the first place. The

Figure 30. a. An image of a piece of tweed and the feature points found in it using the Marr-Hildreth theory of edge detection. The figure illustrates grouping on the basis of orientation of the gradient of feature points. b. image of bricks and feature points grouped on the basis of contrast. Reproduced from [HILD80, figure 25]

first attempt to extend blocks world labelling schemes to real images seems to have been Bajcsy and Tavakoli's

model based interpretation of aerial photographs [BAJC76a].

Marr noted a correlation between different types of intensity change and the scene events that often gave rise to them. Entries in the primal sketch were marked with their interpretation in the scene, such as "edge", "shading edge", and "extended edge" [MARR76, page 490]. With the development of zero crossings, and the de-emphasis of bar and edge masks, it is unfortunately no longer obvious how to compute the assertions that Marr had previously advocated for inclusion in the primal sketch [HILD80, page 75]. The whole issue of constructing the primal sketch from zero-crossings is far from being resolved.

Binford [BINF81] and Lowe and Binford [LOWE81] have recently made an initial pass at the problem of interpreting feature point segments. Compared with the blocks world labelling schemes, the labellings that Lowe and Binford propose are very general. A segment is interpreted as a space curve, and constraints formulated on coincidence and the situations in which a curve corresponds to a bounding contour or true edge.

3.2. Determining surface shape from intensity values

Horn and his colleagues at MIT have studied the perception of shape from grey level shading. The input to the "shape from shading" process is the image and the output is some appropriate representation of surface shape. The exact form of the latter representation is not yet fixed, although [HORN82] offers some thoughts. Since we can perceive surface shape locally, in scenes with little or no semantic content, a reasonable first approximation is to represent the shape of a surface by its local surface normal. This requires two parameters, say p and q. The relationship between shape and the intensity I at a point (x, y) in an image takes the form

$$I(x, y) = R(p, q),$$

which Horn [HORN77] calls the *image irradiance equation.* Mathematically, the image irradiance equation is a nonlinear first order partial differential equation. Horn [HORN77] notes that the function R encodes the position of the viewer, the distribution of light sources (assumed to be fixed), and the reflectance characteristics of the surface material. Horn and Sjoberg [HORN79] derive the relationship between the function R and the bidirectional reflectivity functions used by photometrists, and they show how to calculate it in particular cases. One important special case is Lambertian reflectance, where the intensity varies as the vector dot product of the local surface normal and the direction of the light source.

One useful parameterization of the local surface normal uses the partial derivatives $p = \frac{\partial f}{\partial x}$ and $q = \frac{\partial f}{\partial y}$, where the viewed surface is $z = f(x, y)$. This gives rise to the representation introduced in Section 2 called *gradient space.* Two comments are in order. First, since slant and tilt (as defined by figure 7) have natural perceptual meanings, one might argue that the polar form of gradient space is preferred by the human visual system. Stevens [STEV80] develops this argument, and some further support for the position is provided by [WITK81].

Second, there is a basic problem with gradient space, namely its inability to represent occluding boundaries at which the surface turns away from the viewer. At occluding boundaries the slant angle is $\frac{\pi}{2}$, so that its tangent (s in figure 7) is infinite (note that this objection does not apply to using the angles σ and

τ as [STEV80] notes. Ikeuchi and Horn [IKEU81] introduce a different parameterization (f, g) of surface orientation that they call *stereographic space*. Formally, f and g are related to p and q by

$$f = \frac{2p(\sqrt{1+p^2+q^2}-1)}{p^2+q^2}$$

and

$$g = \frac{2q(\sqrt{1+p^2+q^2}-1)}{p^2+q^2}$$

Ikeuchi and Horn introduce the *Gaussian sphere*, and show that gradient space corresponds to projecting the Gaussian sphere onto the plane from its center, whereas stereographic space is the result of projecting from the north pole (when the viewing direction is from the south pole).

Although it cannot represent occluding boundaries, the mathematical development associated with gradient space is easier, and so it is used in most of this section. For a fixed distribution of light sources, and fixed reflectance characteristics, the image irradiance equation associates a brightness value with each surface orientation. Thus we can assign a brightness value to each point of gradient space. The representation is then called the *reflectance map*[HORN77]. It is convenient to scale brightness values to the range [0, 1], and to make iso-brightness contours explicit. Figure 31 shows the iso-brightness contours for a Lambertian reflector in the case of a single light source near the viewer. Figure 32 shows the result of moving the light source away from the viewer, while figure 33 shows the reflectance map for a gloss surface which approximates white paint.

Having set up the representation of the output of shape from shading, we now consider some of the algorithms that have been proposed for actually determining shape from an image. Recall that the image irradiance equation is a (usually nonlinear) first order partial differential equation. As such, it can be approached using one of the standard techniques for solving partial differential equations. Horn [HORN75] applied the characteristic strip method of solving partial differential equations to reformulate the image irradiance equation as a set of five ordinary differential equations. The solution surface is

$$f(x, y) - z = 0, \qquad (1)$$

Figure 31. Iso-brightness contours for a Lambertian reflector when the light source is near the observer. The brightness at a point is determined by the cosine of the angle between the local surface normal and the view vector. (Reproduced from [HORN77, figure 5]

Figure 32. Iso-brightness contours for a Lambertian reflector when the light source is removed from the observer. The brightness at a point is determined by the cosine of the angle between the local surface normal and the vector from the surface point to the light source.(Reproduced from [HORN77, figure 6]

Figure 33. Iso-brightness contours for a reflector that approximates white gloss paint. Notice the peak relative to the Lambertian reflector shown in figure 13, corresponding to the mirror like component of reflection of gloss paint. (Reproduced from [HORN77, figure 7]

and the image irradiance equation is

$$I(x, y) - R(p, q) = 0. \tag{2}$$

The surface normal has direction ratios $(p, q, -1)$. The characteristic strip method computes the solution surface by finding a family of space curves (strips) whose local tangents all lie in the tangent plane of the solution surface. Such a curve can be specified by a one parameter family of points $(x(s), y(s), z(s))$, where s corresponds to the distance traversed along the curve. Differentiating equation (1) with respect to s, we find:

$$p\frac{dx}{ds} + q\frac{dy}{ds} - \frac{dz}{ds} = 0.$$

It follows that $(\frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds})$ lies in the tangent plane of the solution surface. Since $pR_p + qR_q - (pR_p + qR_q)$ is identically zero, $(R_p, R_q, pR_p + qR_q)$ also lies in the tangent plane. Equating these two vectors gives the following three equations:

$$\begin{aligned}
\frac{dx}{ds} &= R_p, \\
\frac{dy}{ds} &= R_q, \\
\frac{dz}{ds} &= pR_p + qR_q,
\end{aligned}$$

Finally, differentiating equation (2) with respect to x gives:

$$I_x = R_p p_x + R_q q_x.$$

Since $p_y = I_{xy} = q_x$, we find

$$I_x = R_p p_x + R_q p_y,$$

and so

$$I_x = \frac{dp}{ds}.$$

Similarly,

$$I_y = \frac{dq}{ds}.$$

The characteristic strip formulation was used by Horn [HORN75] as the basis of an iterative computation as follows. Suppose that we know that image point (x_n, y_n) corresponds to a surface point at which the surface gradient is (p_n, q_n). Refer to figure 34, which shows iso-brightness contours passing through (x_n, y_n) in the image and (p_n, q_n) in the reflectance map. Consider a step ds along the characteristic strip, from (x_n, y_n) to (x_{n+1}, y_{n+1}) and, correspondingly, from (p_n, q_n) to (p_{n+1}, q_{n+1}). The five ordinary differential equations given above show that the step in the image is in the direction (R_p, R_q), that is to say, along the normal to the iso-brightness contour in the reflectance map. Similarly, the step in the reflectance map is in the direction normal to the iso-brightness contour computed in the image. In this way, knowing the reflectance map, one can proceed to compute a sequence of points and local gradients along the characteristic strip starting from a point in the image at which the surface gradient is known. Figure 35 illustrates the results of applying Horn's algorithm.

One problem with this method concerns the choice of the **singular image point** (x_0, y_0) required to start the iterative process at which the surface gradient (p_0, q_0) is determined uniquely by the intensity data. A further problem is that Horn's algorithm depends on the assumption that the underlying surface is locally convex at the singular point. Finally, the class of image irradiance equations for which Horn's algorithm works was unknown. (The latter question has recently been answered by [BRUS81].) Consequently research was directed to discover the criteria under which the shape of a surface is uniquely determined by an image. One suggestion was that bounding or occluding contours provided such conditions. Along such contours, the surface normal can be computed exactly from the image. However, occluding contours pose a problem for

Figure 34. The basis of Horn's iterative computation of shape from shading by the characteristic strip method. The surface gradient at the image point (x_n, y_n) is known to be (p_n, q_n). Isobrightness contours are shown in the image and in the reflectance map. A short movement in the image along the characteristic strip is in the direction of the solid line, which is normal to the isobrightness contour in the reflectance map. The converse relation also holds, and is depicted by the dotted line.

Figure 35. A sample result of Horn's characteristic strip algorithm. The figure shows the picture of a nose with superimposed characteristic strips (top figure) and contours (bottom figure). Reproduced from [HORN75, figure 1].

the gradient parameterization of local surface orientation, namely that at least one of the gradients p or q is infinite. This led Ikeuchi and Horn [IKEU81] to propose *stereographic projection* as defined above.

Ikeuchi and Horn [IKEU81] note some additional problems with the characteristic strip method. First, since the iterative method outlined above proceeds unidirectionally along a characteristic strip, it cannot exploit boundary conditions at *both* ends of the strip. Second, the build up of numerical errors along any individual strip can be substantial. A novel feature of Horn's [HORN75] algorithm is the simultaneous development of several characteristics to control the build up of error in any one. Woodham [WOOD81] observes that one can solve for surface shape if one makes a global assumption about the surface type, for example that it is convex, a ruled surface, or the surface of a generalized cylinder(see Section 6). Other authors propose smoothness constraints derived from the fact that the integral of depth around a closed loop in the image is zero [BROO79, STRA79]. Ikeuchi and Horn [IKEU81] discuss a more direct formulation of a smoothness condition that they state in terms of the stereographic parameterization of surface orientation. This enables them to use the bounding contour of an object as a source of boundary values for an iterative computation which fills in the surface orientation in the interior. Formally, denote the nth iterative approximation to the value of $f_{i,j}$ at image point (i, j) by $f^n_{i,j}$ with an analogous formula for $g_{i,j}$. Letting the local (four point) average at the nth iteration be $\bar{f}^n_{i,j}$, Ikeuchi and Horn derive the following recurrence relation as the basis of an iterative algorithm [IKEU81]:

$$f^{n+1}_{i,j} = \bar{f}^n_{i,j} + \lambda[I_{i,j} - R_s(\bar{f}^n_{i,j}, \bar{g}^n_{i,j})]\frac{\partial R_s}{\partial f}.$$

Here, R_s is the partial derivative of the reflectivity function R in the case of stereographic projection, analogous to R_p which was used above in the characteristic strip method. The resulting algorithm has been tested on a variety of images and works well. In particular, it appears to degrade gracefully as errors are introduced to the placement of the light source, the surface orientation on the boundary, and the nature of the reflectivity assumed for the surface. Strong *empirical* evidence is provided that the algorithm converges, although no proof is demonstrated. In case the occluding contour is partially incomplete, Ikeuchi and Horn's

algorithm still appears to converge, though it is not known at how many points it is necessary to specify the stereographic parameterization of the surface normal.

Bruss [BRUS81] has recently studied some of the mathematical properties of the image irradiance equation. First, she has shown that discontinuous solution surfaces can arise from a continuous image irradiance equation. It follows that one cannot determine for a continuous image irradiance equation whether or not there is an edge. The curvature of a surface also cannot be determined in general from its image. As an example, the image irradiance equation $x^2 + y^2 = p^2 + q^2$ has two different solution surfaces, one of which $z = xy$ consists entirely of hyperbolic points, while the other $z = \frac{1}{2}(x^2 + y^2)$ consists entirely of elliptic points. However, Bruss has proved that there is only one solution that is convex. She has also shown that bounding contours can be determined from the image only when the image irradiance equation is *singular*. This means that the reflectance function R and its first order partial derivatives are continuous, while the intensity function I is singular in x and/or y. For any given singular image irradiance equation the points on the occluding contour can be found by inspection of the intensity function $I(x, y)$.

Bruss also studied singular "eikonal" image irradiance equations that are of the form $p^2 + q^2 = I(x, y)$. If the intensity function I(x,y) *vanishes to second order* at the singular point, that is to say has the form

$$I(x, y) = \alpha x^2 + \beta xy + \gamma y^2 + O(|x^3| + |y^3|),$$

then there is exactly one positive locally convex solution surface in the neighborhood of the singular point. This result is applied to show that if there is a closed bounding contour, the solution surface is unique (up to translation along the z axis). If either the reflectance function is not $p^2 + q^2 = I(x, y)$, the intensity function does not vanish precisely to second order, or there is not a smooth closed bounding contour, there is not a unique solution surface. The reflectance function $p^2 + q^2$ closely models a number of practical situations such as imaging with scanning electron microscopes.

Woodham and Horn, Woodham, and Silver have developed a rather different method for computing shape from shading that may prove very important in practice, even if it bears very little resemblance to the

processes of human vision [WOOD81, HORN78b]. Suppose that we fix the view (camera) position, and that we set up two light sources at different known points. Suppose that the intensity levels at any image point (x, y) in the first and second images are $I_1(x, y)$ and $I_2(x, y)$. The first of these restricts the surface orientation at (x, y) to the iso-brightness contour in the reflectance map corresponding to the brightness value computed from $I_1(x, y)$ (figure 36a). Similarly, the surface normal is constrained by the iso-brightness contour defined by $I_2(x, y)$ (figure 36b), and hence to their intersection (figure 36c). A third light source provides complete disambiguation. This process has been called ***photometric stereo***, and can be implemented very efficiently as follows. First, there is a calibration phase in which an object whose surface shape is known, such as a sphere, is illuminated in turn by the set of light sources and imaged. This generates a set of n-tuples of intensity values (n is the number of light sources), each of which is associated with a known local surface orientation on the known calibration object. The surface orientation distribution of an unknown object can then be computed by using the n-tuples of intensity values at each corresponding image point as a lookup key into a table. To keep the storage requirements of the algorithm within bounds, the intensity values are quantized. One current implementation quantizes intensity to ten values in each of three measurements. Intermediate intensity triples are handled by interpolation from the nearest entries in the table. The method, which has been implemented by Silver, is fast and remarkably accurate [SILV80]. Figure 37 shows the reconstruction of an egg after a calibration phase using a sphere. Figure 38 is the superposition of a cross section of the known surface onto one computed by photometric stereo. Photometric stereo has been extended to handle objects with specularities by Ikeuchi [IKEU81], and has recently been applied to the industrial problem of bin-picking [BIRK81].

Optical flow

In Section 3.1.1, we surveyed the work of Marr and his group based on the detection of the important intensity changes in an image. In particular, we mentioned the recent work of Marr, Ullman, and Richter on detecting the direction of motion of a zero crossing by taking the time differential of $\Delta G^* I(x, y, t)$. We conclude this section with a brief discussion of the work of Horn and Schunck [HORN81c] that proposes

Figure 36. An illustration of photometric stereo. Suppose (a) the the brightness measured at the point (x, y) in the first image is 0.6 and (b) in the second image the brightness at the same point is 0.2. (c) superposition of the first two constraints shows that there are at most two consistent surface gradients.

Figure 37. The reconstruction of an egg shape by Silver's implementation of photometric stereo after a calibration phase using a sphere. The reflectance of all surfaces was Lambertian. (Reproduced from [SILV80])

Figure 38. Comparison of the cross section of an egg and a knob shape computed by photometric stereo (solid lines) and the true cross sections extracted from photographs (dotted lines). (Reproduced from [SILV80])

a method for computing *optical flow* by differentiating the brightness distribution in the image with respect to time. Optical flow is the distribution of velocities of apparent movement caused by smoothly changing brightness patterns. It has been noted that optical flows encode rich information about a scene and observer motion, and it has been suggested that this information can be computed from the flow field. This position is particularly associated with the followers of J. J. Gibson, who first studied flow fields [GIBS50, GIBS66, CLOC80, KOEN75, KOEN76, KOEN77, PRAZ80]. In particular, it has been suggested that optical flow facilitates object segmentation [NAKA74, CLOC80], computation of the parameters of the observer's own motion relative to the scene [PRAZ80, LONG80], and the determination of visible local surface normals [PRAZ80].

The work on *interpreting* optical flow has generally assumed that the flow is given, that it is somehow computed automatically and sufficiently noise-free. "Velocity sensitive neurons" have been postulated to compute the optical flow in animate visual systems [NAKA74]. Horn and Schunck [HORN81c] have studied the generation of the optical flow from brightness patterns that vary smoothly with time. They restrict attention to imaging a flat surface with uniform incident illumination, and smoothly varying reflectance. The image brightness at point (x, y) at time t does not change, and so

$$\frac{dI(x, y, t)}{dt} = 0.$$

Expanding, by the chain rule we find

$$I_x u + I_y v + I_t = 0,$$

where (u, v) is the optical flow $(\frac{dx}{dt}, \frac{dy}{dt})$. This shows that the component of the flow field in the direction of the brightness gradient (I_x, I_y) is

$$\frac{-I_t}{\sqrt{I_x^2 + I_y^2}}.$$

It is not possible to determine the component of the flow field perpendicular to the intensity gradient, that is to say along the iso-brightness contours. In practice, quantization errors and noise imply that $\frac{dI}{dt}$ is not exactly zero. To account for this, an error term E_b is introduced and defined by:

$$E_b = I_x u + I_y v - I_t.$$

To compute the component of the flow field along iso-brightness contours requires an extra constraint. Horn and Schunck derive a measure of the departure from smoothness of the flow [HORN81c]. Smoothness can be estimated by the square of the magnitude of the gradient of the optical flow velocity:

$$E_c^2 = (\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2 + (\frac{\partial v}{\partial x})^2 + (\frac{\partial v}{\partial y})^2.$$

The estimate of the departure from smoothness and the change in brightness combine in a measure of the error:

$$E^2 = \alpha^2 E_c^2 + E_b^2.$$

Using the calculus of variations, Horn and Schunck eventually derive the iterative computation:

$$u^{n+1} = \bar{u}^n - \frac{I_x[I_x\bar{u}^n + I_y\bar{v}^n + I_t]}{(\alpha^2 + I_x^2 + I_y^2)},$$
$$v^{n+1} = \bar{v}^n - \frac{I_y[I_x\bar{u}^n + I_y\bar{v}^n + I_t]}{(\alpha^2 + I_x^2 + I_y^2)},$$

Initially, the components (u, v) of optical flow are assumed to be zero everywhere. The algorithm works well on synthetic patterns as figure 39 shows.

3.3 Segmentation

A great deal of effort continues to be expended on segmentation, a process that is essentially the dual of

Figure 39. Optical flow patterns computed by the Horn-Schunck algorithm. (Reproduced from [HORN81c, figure 10])

edge finding. Recall that edge finding has three stages. First, significant intensity changes are detected and localized. The feature points are then grouped to form linear segments. Finally, segments are interpreted as scene events, such as depth, reflectance, and shadow boundaries, as well as discontinuities in surface orientation (true edges). Analogously, the process of segmentation begins by isolating those regions of an image in which there are no significant changes of intensity, and adjacent regions are then grouped, or "merged". Finally, the regions are interpreted as scene events, typically visible surfaces, shadowed areas, or patches in which the reflectance is uniform. As in the case of edge finding, the difficult issue is to frame a precise definition of "significant" so that segmented regions correspond to the perceptual entities that are their interpretations.

Some authors [MARR78, page 64] have concluded that segmentation is an ill-defined operation, since regions do not always correspond to portions of visible surfaces. Certainly, simple schemes for segmentation produce many spurious regions, just as simple approaches to edge finding ascribe significance to spurious intensity changes. Several authors have pointed out that region finding is no more, and no less, difficult than edge finding [HARA79, BINF81]. If segmentation and edge finding differ at all, it is with respect to the descriptions naturally associated with two-dimensional regions and one dimensional segments.

Early work on segmentation implicitly modelled an image as a collage of regions that are homogeneous in intensity and separated by step changes. A slight refinement was to accommodate noise heuristically by merging across weakest contrast boundaries [BRIC70, BARR71].

One approach to improving segmentation schemes is to incorporate better models of edge finding. Each of the processes for discovering feature points outlined in section 3.1.1 can be adapted to segmentation. Haralick [HARA80, page 62] observes that two pixels are part of the same region if and only if there is no significant difference between their associated sloped facets. If every intensity change uncovered by the Marr-Hildreth theory of edge finding is significant then closed contours of zero-crossings correspond to regions.

An alternative approach to improving segmentation is to invoke domain specific semantic information either to encourage or inhibit the merging of regions [TENE77, SELF81]. Such schemes for segmentation are

analogous to the semantically guided edge finders advocated by [BAJC75, BAJC76b, SHIR73].

Horn's work on shape from shading discussed in the previous section implies that there can be significant variations in intensity within a perceptual surface. In general, only a planar surface produces a region that is uniform in intensity (ignoring noise). Segmentation on the basis of intensity values is a heuristic consequence of the early preoccupation with scenes composed of planar surfaces (see section 2). According to the image irradiance equation, intensity is uniform within the image of a planar region because the surface orientation is constant. Ballard [BALL80] suggests that the concept of segmentation is more naturally associated with representations based on surfaces: Marr's $2\frac{1}{2}$D sketch, Horn's needle map, and Barrow and Tenenbaum's intrinsic images. As before, segmentation is the dual of discovering significant changes, say of surface orientation or depth. Such processes await investigation. Ballard proposes that the Hough transform can be generalized for this purpose [BALL80].

Many surfaces have constant texture or color. Color may be perceptually uniform across a surface even if there is significant variation in intensity. Horn's work [HORN74], based on Land's retinex theory, embodied the idea of segmentation on the basis of "lightness" for a two-dimensional world of "Mondrians". Extending Horn's work to three dimensions would not be trivial. Tomita, Yachida, and Tsuji [TOMI73] also experimented with segmentation on the basis of color. Ohlander, Price, and Reddy [OHLA78] experimented with multi-spectral descriptions including hue, saturation, and brightness. Brady and Wielinga [BRAD78] note that the Ohlander program works well on "patchwork quilt" images that are composed of large regions that are uniform in one of its nine descriptors. Tenenbaum and Barrow [TENE77] observe that because it is based on this heuristic, the program is easily fooled, especially by regions of repeated texture.

3.4 Texture

Texture is a compelling visual cue to the properties of a surface. We can recognize a region of an image as grass or the foliage of a bush or tree, and often we can do so in a black-white image without the aid of color. We easily distinguish velvet, woollen weaves, herring bone, and raffia. Pebbled paths stand out

from the surrounding soil. It seems that most terrain classification from satellite images is based on texture discrimination and recognition.

Haralick [HARA79] points out that although hundreds of articles have been written on the subject of computer recognition and description of texture (mostly from the standpoint of pattern recognition), few precise definitions of texture have been given. As a result, texture discrimination techniques are largely ad hoc. Most accounts of texture are based on the idea that its distinguishing characteristic is regularity of the "primitive" elements, called *texels*, of which the texture is composed, and of the spatial relationships between texels. If there is wide variation in the size of individual blades of grass, or if the blades are sparsely and non-uniformly distributed in the image, the grassy texture appears "ragged". In general, the strength of a texture is determined by the regularity of its texels and regularity in the spatial relationships between the texels. Zucker proposes that ideal textures are completely regular and can be modelled by regular two-dimensional graphs [ZUCK76]. He suggests that naturally occurring textures are distortions of ideal textures.

We prefer a rather different view of texture, based on an idea of what purpose texture perception serves. A grassy lawn, the foliage of a tree, and a pebbled path are all perceived as surfaces. Microscopic variations in a surface determine its reflectance [HORN79], while large scale variations in a surface determine its topography. The processes of determining shape from stereo, contour, texture, and motion are discussed in section 4. Mostly they operate on isolated edges and regions found by one of the processes discussed in sections 3.1 and 3.3. We suggest that texture refers to surface variations intermediate between microscopic reflectance changes and topographical changes made explicit by edge finding and segmentation. It follows that descriptions of texture require the isolation of macroscopic surface facets and the determination of the spatial relationships between such facets. In order to be perceived as a single surface, surface facets (texels) that are physically close should have similar descriptions. Regularity is the physical basis for grouping facets as a single surface. Surface variations are labelled reflectance, texture, or topographic depending upon the resolution at which they are viewed. (See [MALE77] for similar remarks).

The twin themes of statistics and structure run through most of the literature on texture. We commented

above that regularity is central to texture. Inevitably, regularity has been modelled statistically; for example, the distribution of slopes of individual blades of grass has a strong peak and small variance. Statistics has been applied more or less uncritically to texture. Maleson, Brown and Feldman [MALE77] quip that "the problem with statistical analysis is that if an inappropriate set of statistical measures is used, the final results are meaningless. For this reason, it is important to base statistics on a reasonable model of the phenomena to be measured." One approach to a 'reasonable model' is to apply statistical analysis only to texels that carry significant information about surface structure, in particular, those isolated by edge finding and segmentation.

Haralick [HARA79] has presented a good survey of purely statistical approaches to texture. Simple ideas such as computing autocorrelation functions perform relatively poorly [WESK76]. Bajcsy [BAJC73, BAJC76] model regularity by periodicity as determined from features of the polar form $P(r, \phi)$ of the Fourier transform of subimages. Combining all r to show the dependence on ϕ, peaks in $P_r(\phi)$ give evidence of directional textures such as grass. If there are no peaks in $P_r(\phi)$, $P_\phi(r)$ is investigated for peaks that give evidence of blob-like textures. Textures need to be strongly periodic to be found by the method. A better model was introduced by Julesz [JULE62] and refined by several authors, including Rosenfeld and Troy [ROSE70] and Haralick [HARA71]. The co-occurrence $P(i, j, d)$ specifies the relative frequencies with which two grey levels i and j occur separated by a distance d. Haralick and Bosley [HARA73] computed a number of features from co-occurrence matrices and used them to classify terrain from satellite images, achieving success rates of over 80%. Julesz [JULE71] conjectured that textures can be discriminated by non-attentive vision if and only if they differ in their second order statistics (essentially their co-occurrence matrices). As originally formulated, co-occurrence matrices specify the relative frequencies of individual grey levels. Horn's work on shape from shading shows how much information is confounded in a single grey level. Only when surfaces are essentially planar, for example satellite imagery, is grey level a reliable basis for aggregation into regions corresponding to surfaces. Haralick [HARA79, page 787] notes that while co-occurrence based on grey levels captures spatial relationships it does not capture shape aspects and hence does not work well for textures composed of large-area texels. In short, individual pixels are poor descriptors of surface facets.

Co-occurrence is not restricted to grey levels, however. Maleson, Brown, and Feldman [MALE77] propose segmented regions as texels. They suggest region descriptors that are insensitive to scale, such as the orientation of the major axis and eccentricity of the best fitting ellipse to a region. Details of the performance of a system based on this technique on a range of textures has yet to be published. Marr [MARR76] suggests that texture discrimination based on co-occurrence matrices could be accounted for by discrimination on ordinary statistics applied to the primal sketch. The scheme was not implemented, nor were descriptions proposed for texture. To this end, the main advance has been due to Vilnrotter, Nevatia, and Price [VILN81]. Their work is based on the Nevatia and Babu edge finder (see section 3.1). Textures are detected from edge repetition arrays that specify the co-occurrence of edges in a particular direction at a particular spacing. Once detected, texels are described in terms of their average size and intensity. Spatial organization is found by relating texels in different directions. Figures 40 and 41 show the results computed by the system for raffia and brick textures.

Figure 40. a. image of raffia. b. Sample of output from analysis of edge repetition arrays. c. abstract representation of the texels found in the raffia image. d. Reconstruction of the raffia image using the abstract texels (Reproduced from [VILN81, figures 1-4]

Figure 41. a. Two images of brickwork. b. Illustration of abstract primitives found in the images of *a*. c. Illustration of the spatial organization found in the textures in a. (Reproduced from [Viln81 figures 6,8,9]

4. Determining shape from the primal sketch

4.1. Shape from stereo

The slight disparities in the images received by the left and right eyes enable humans to determine the shape and relative depth of visible surfaces. The importance of automating stereo, and the difficulty of the problem, is well stated in a recent overview of Defense Mapping Agency applications [MAHO81].

There have been several attempts to develop a computational theory of binocular stereopsis since Julesz's demonstrations in the early 1960's that it is possible to fuse images stereoscopically without extensive monocular processing. Julesz [JULE71] presented substantial experimental evidence regarding binocular fusion of random dot stereograms, a perceptual device that he originated(see figure 42). The essence of stereo vision is the matching of descriptions computed from the images presented to the left and right eyes. The Julesz demonstrations argue that the descriptions to be matched are available at an early stage of visual processing. Two candidate descriptions considered for matching to date are the image (*area correlation*), and a representation of intensity changes (*edge based stereo*).

Julesz conjectured that stereo is a local parallel process, and a number of algorithms have been designed with this conjecture in mind. The first of these is due to Dev [DEV75], closely followed by Marr and Poggio [MARR76b, MARR76c]. Marr and Poggio call their algorithm "cooperative" by analogy with boundary value computations in physics. The algorithm could equally well be called a relaxation process [DAVI81]. Marr [MARR78] notes a number of difficulties with such algorithms as a theory of *human* stereo vision, namely human tolerance for the defocussing of one image, and the apparent ubiquity of vergence movements of the eyes as two images are fused. Perhaps more important are the so-called hysteresis effects in which images are matched only after a delay, or remain fused when they are pulled apart by an amount greater than is apparently possible for matching. Marr and Poggio [MARR79b] argue that while hysteresis effects suggest cooperativity, the effect can also be achieved by postulating a dynamic memory in which intermediate results of stereo processing can be stored.

Figure 42. A random dot stereogram devized by [JULE71]. First, an image is produced for the left eye, composed of random dots. The view from the right image is determined by translating each dot in the random dot image leftwards by an amount that depends on the relative distance of the corresponding point in a conceptual scene. Some dots are occluded as a result. Other image points that could not be seen by the left eye are now visible in the right eye. Such points are randomly filled by new dots.

Most work on area correlation stereo [HANN74, QUAM71, HEND78] operates on a succession of small windows (typically 10 by 10) from one image. For each window in the left image, a search is conducted for that window in the right image that optimizes a suitable correlation relation between the grey levels in the two windows. Area correlation has proven to be particularly effective in textured or smoothly shaded areas. It has supported terrain following automatic guidance systems, and some automatic mapping systems where the goal is to generate a digital terrain model associating a height with each map point imaged. Area correlation implicitly assumes that the left and right images differ only in viewpoint, that is they only differ photometrically. As a result, area correlation performs poorly near surface discontinuities where this photometric assumption is false. Conversely, edge based stereo assumes that the invariance between the left

Figure 43. The zero crossings located in the four channels of the Marr-Hildreth theory for the random dot image shown in a. (Reproduced from Grimson's forthcoming book [GRIM81]).

and right images is geometric. Baker and Binford [BAKE81] observe that in general the geometric assumption implicit in edge based stereo is more realistic than the photometric assumption implicit in area correlation. A further shortcoming of current area correlation techniques is that their accuracy is limited to a fraction of the window size (typically 5 picture elements). Edges can normally be localized with subpixel accuracy [MACV81, MARR79a].

Implicit in the above remarks about the suitability of area correlation for stereo matching of textured areas is a model of texture based on grey levels. We found earlier (Section 3.4) that texture describes surface microstructure with texels corresponding to surface facets. The extension of the approaches to edge based stereo to densely textured areas awaits further work on edge and region based accounts of texture.

Edge based stereo is strong where area correlation is weak, and conversely. An additional advantage of edge based stereo is its potentially greater efficiency, as there are considerably fewer edges than grey levels.

Stereo rests upon, and provides a stiff test for, any account of edge finding. In section 3.1.1 we discussed a number of approaches to edge finding. Marr and Hildreth's approach to detecting feature points has been applied to stereo by Marr and Poggio [MARR79b]. The left and right images are convolved with ΔG operators as described in 3.1.1. Matching takes place between the paired sets of zero crossings. Figure 21 showed the image of a coffee jar sprayed with spots of paint to yield a Julesz-like random dot stereogram from a real scene, and figure 24 showed the zero crossings produced by each of the four channels proposed by the Marr-Hildreth theory. Figure 43 shows the zero crossings produced in each of the four channels for the random dot image shown in figure 43a. In both figures 24 and 43, it is evident that it is considerably more difficult to establish an optimal match between the output of the fine channel from the left and right images than between the outputs of the coarse channel. Exploiting this observation, matching proceeds from the coarsest channel, which makes explicit gross detail and establishes a rough correspondence, down to the finest resolution channel. This coarse-to-fine strategy, in which a rough plan is used to narrow the search space prior to more detailed processing, is a basic idea in artificial intelligence. The application of a coarse-to-fine strategy like that in the Marr-Poggio theory of stereo seems to have been used by Moravec [MORA80] in a system constructed at Stanford. Note that the coarse-to-fine strategy may have to be modified for closely spaced edges that occur with textured surfaces.

Once the match between the zero crossings in the two images has been established for the four channels, one can compute the angular disparities (or even distances) to matched zero crossings, [GRIM81] gives details. Figures 44 and 45 show the disparity values computed for the coffee jar and the random dot stereogram shown in figure 42. A disparity value is recorded only where zero crossings from the two eyes are matched, and so the disparity map is often discrete. Since we mostly perceive the world as composed of smooth surfaces, it is necessary to consider possible interpolation processes for smoothly completing the surface orientation map from the discrete set of disparity values. This is a general problem and is discussed in the next section. Grimson's reconstruction process computes the shape shown in figure 46. Grimson's implementation of the Marr Poggio stereo theory demonstrates all of Julesz's experimental findings. It has also been applied to a

Figure 44. The disparity map computed from the output of the stereo matcher for the coffee jar. (Reproduced from Grimson's forthcoming book [GRIM81])

small number of stereo pairs of natural images.

In section 3.1 we characterized edge finding as having three successive stages: determining feature points, grouping them on the basis of their attributes, and interpreting them as scene events. The Marr-Poggio theory matches feature point descriptions on the basis of the position and sign of the zero crossing, before the feature points are grouped into linear segments. Recent psychophysical findings of Mayhew and Frisby [MAYH81] seem to indicate that it is necessary to match richer descriptions than zero crossings. Baker and Binford [BAKE81] and Arnold [ARNO78] propose that ambiguities can be resolved more efficiently and successfully on the basis of the richer descriptions associated with points on linear segments. Baker and Binford [BAKE81] match points at various scales using the position, contrast, and slope of the segment in the image, and the intensities on both sides of the intensity change. These separate pieces of evidence are combined by a linear weighting function. The optimal match is found along horizontal scan lines using a fast linear programming

Figure 45. The disparity map computed from the output of the stereo matcher for the random dot stereogram shown in figure 42. (Reproduced from Grimson's forthcoming book[GRIM81])

technique. Once edges are matched, grey levels are correlated by a similar process. Figure 47 shows the results computed by Baker and Binford's program on an image with both texture and edges. Arnold [ARNO78] also filters putative matches according to the position, slope, and contrast of edge segments. The edge segments are found using Hueckel's surface fitting technique. Arnold claims that this is the program's main deficiency. It is interesting to speculate how the Baker and Binford or Arnold algorithm might perform if they had the Marr-Hildreth zero crossing data to work on. Alternatively, it is interesting to ask how the richer descriptions proposed by Baker and Binford, Arnold, and Mayhew and Frisby could be incorporated into the Marr-Poggio theory.

All of the programs discussed in this section, except Arnold's, assume that the left and right images have been rectified prior to stereo matching. That is, they assume that the images have been rotated, translated, and scaled so that corresponding feature points can be found on the same horizontal scan line. Arnold's

Figure 46. The reconstructed coffee jar interpolated by Grimson's program from the disparity map shown in figure 44. (Reproduced from Grimson's forthcoming book [GRIM81])

Figure 47. Example results of Baker and Binford's stereo program. a. Stereo pair of images of natural terrain. b. The edges found in the images by a simple differencing operation. c. Illustration of disparities computed for the images. (Reproduced from [BAKE81, figures 10,11, and 17.])

program relies upon a rectification procedure developed by Moravec and Gennery [MORA79, GENN79]. In this procedure, "interesting" points such as corners are found in both images, and an optimal match is found. The tentative match is refined using a high resolution area correlator. A camera model solver computes the direction of the stereo axis, the relative rotation, scale change, and lateral translation between the left and right views. The ground plane is also determined. Lucas and Kanade have recently explored the application of a Newton-Raphson like technique to solve for the camera parameters[LUCA81]. Rectification remains a difficult open problem.

4.2 Shape from contour

Witkin [WITK81] has make a start on what seems to be a promising approach to computing shape from a primal sketch. His work concerns the perceived slant and tilt of a line drawing lying in a plane, such as the

map outline shown in figure 48. Witkin's approach relies on making the image forming process explicit, and using it to derive a probability density function. Assume that the axes in the image and in the planar scene are aligned, and denote the tangent direction measured in the image by α^* and the tangent at the corresponding point in the scene by β. Image foreshortening gives the relation

$$\tan(\alpha^* - \tau) = \frac{\tan\beta}{\cos\sigma},$$

where τ is the tilt and σ is the slant of the planar scene. A collection of measurements of α^* taken throughout the image define a distribution of tangent directions. If we hypothesize particular values for σ and τ, the above relation establishes a distribution for β. Given an *expected* distribution for (β, σ, τ), the likelihood of any *observed* distribution of α^* can be evaluated. Witkin shows that the probability density function of (β, σ, τ) is $\frac{\sin\sigma}{\pi^2}$. It turns out that the relative likelihood of (σ, τ) given a set A^* of measurements of α_i^* is

$$\prod_{1 \le i \le n} \frac{\pi^{-2}\sin\sigma\cos\sigma}{\cos^2(\alpha_i^* - \tau) + \sin^2(\alpha_i^* - \tau)\cos^2\sigma}.$$

The value of (σ, τ) for which this estimator assumes a maximum is the maximum likelihood estimate for surface orientation. Figure 49 shows the results of this procedure applied to a variety of shapes, and compares it to the tilt as estimated by humans. Witkin found that tilt could be estimated considerably more accurately than slant, a result he and Stevens [STEV80] established independently. In further work, Witkin assumes that surfaces are locally planar and applies a similar analysis to compute local surface orientation [WITK81].

4.3 Shape from texture

Of the modules which seem to bridge the gap between the primal sketch and the surface orientation map, none has received quite as much attention from psychologists as the computation of surface orientation and depth from texture gradients. Ever since Gibson [GIBS50] drew attention to their importance for computing depth (figure 50), they have been a major concern of his followers. Stevens [STEV80] notes the simplifications

Figure 48. A geographic contour shown at various orientations, with the density function obtained at that orientation. The density function is plotted by iso-density contours, with (σ, τ) represented in polar form: σ is given by distance to the origin, τ by the angle. The sharp symmetric peaks clearly visible at higher slants are the maximum likelihood estimates for (σ, τ). Reproduced from [WITK81, figure 4]

Figure 49. Results of running Witkin's estimation strategy. A number of shapes are shown at left. The center column plots human estimation of the tilt of the shapes, and the right column shows the tilt vectors predicted by the estimation strategy. (Reproduced from [WITKIN, figure 5]

Figure 50. A texture gradient in a natural scene. (Reproduced from [GIBS50]

assumed by most published analyses of texture gradients in the psychological literature. Typically, a horizontal ground plane is assumed that stretches into the far distance. Stevens proposes a two step computation: (1) isolate "characteristic directions" in which there is no depth change, and (2) compute depth from the slant and tilt representation of surface orientation. The idea has not been implemented. It assumes that primitive texels can be computed for natural images with sufficiently precise descriptions that the characteristic directions can be computed accurately. Bajcsy and Lieberman [BAJC76a] base the computation of texture gradients on Bajcsy's applicaton of the Fourier power spectrum to describing texture (see section 3.4) [BAJC73]. All of the other methods for computing texture discussed in section 3.4 could be adapted to the determination of texture gradients.

Kender [KEND80] has considered the computation of shape from texture as an instance of a general methodology that yields "shape from" algorithms from a variety of image observables. The general plan of

Kender's approach has three parts:

• Primitive texels are extracted from the image. Kender assumes that texels are the image of planar surface facets, but he offers no guidance for computing them.

• Each texel is assigned a set of possible scene parameters. This is the core of the approach. He introduces a set of normalized texture property maps (NTPM) that generalize, for example, Horn's reflectance map (section 3.2).

• texels that are assumed to arise from neighboring surface facets in three space compare the constraints on their sets of possible parameters, casting out those that are inconsistent on some appropriate grounds of smoothness. As Kender points out, this step is similar to relaxation processing as advocated by Davis and Rosenfeld [DAVI81].

Ballard's parameter networks bear many similarities to Kender's scheme [BALL81]. Where Kender prefers intersecting constraints, Ballard prefers adding them in accumulator arrays as part of his advocacy of the generalized Hough transform.

Kender's NTPMs have four associated choices.

• Since the goal of a "shape from" algorithm is a precise description of surface shape, an appropriate parameterization of surface orientation needs to be chosen. Popular choices are gradient space (section 2, section 3.2), the Gaussian sphere [HORN82], and stereographic space [IKEU81] (see section 3.2). In the example presented below, we choose gradient space.

• The imaging geometry is a key component of texture gradients. The essential choice is between perspective and parallel (orthographic) projection. Kender shows that while the mathematics of perspective projection is more complex, the constraint it offers is considerably tighter. For mathematical simplicity, we choose parallel projection.

• Assuming that texels have somehow been made available, several texture measures can be computed and related to possible scene fragments. Popular choices are texel length (for example the length of the major axis of one of the barrels shown in figure 50), the slope in the image of some direction associated with the

Figure 51. A texture with an unusual relationship between facets and the underlying planar surface. (Reproduced from [KEND80, figure 3.4]

texel (compare [MALE77], the angle in the image between two directions associated with the texel (compare Kanade's work on skew symmetry discussed in section 2 [KEND80]), or dot or edge density (compare [ROSE70, ROSE71]. We consider length and slope in the example below.

• Finally, the way in which the facet that projects to the texel is connected to the underlying surface has to be assumed. In figure 51 the facets can be interpreted as lying in the plane or protruding from it.

As an example of Kender's approach, consider the abstract texture shown in figure 52. We shall make the following choices: gradient space representation of surface orientation, parallel projection, and length and image slope of texels. We shall assume that the texels all lie in a planar surface and form two mutually orthogonal sets. We shall show that the orientation of the surface is completely determined.

We first consider the NTPM corresponding to the length of a texel. Figure 53 shows a texel of length L and slope α in the image. Suppose that one end of the texel is at the image origin and that the corresponding

Figure 52. An abstract texture. The horizontal and texels slanted at 45° are assumed to have the same length in the image and in the scene. It is further assumed that the horizontal texels are orthogonal to the slanted texels in the scene. (Reproduced from [KEND80, figure 3.9]

scene point is $(0, 0, d)$. Suppose that the deprojection of the other end of the texel is $(L\cos\alpha, L\sin\alpha, e)$. Since the deprojection of the texel lies in the plane whose normal is $(p, q, -1)$, it follows that $e - d = pL\cos\alpha + qL\sin\alpha$. The length of the deprojected texel is therefore

$$L_\alpha = L[1 + (p\cos\alpha + q\sin\alpha)^2]^{\frac{1}{2}}.$$

Applying this to the texture shown in figure 52 we have $L_0 = L_{\frac{\pi}{4}}$, that is

$$(1 + p^2) = (1 + \frac{(p+q)^2}{2}),$$

or,

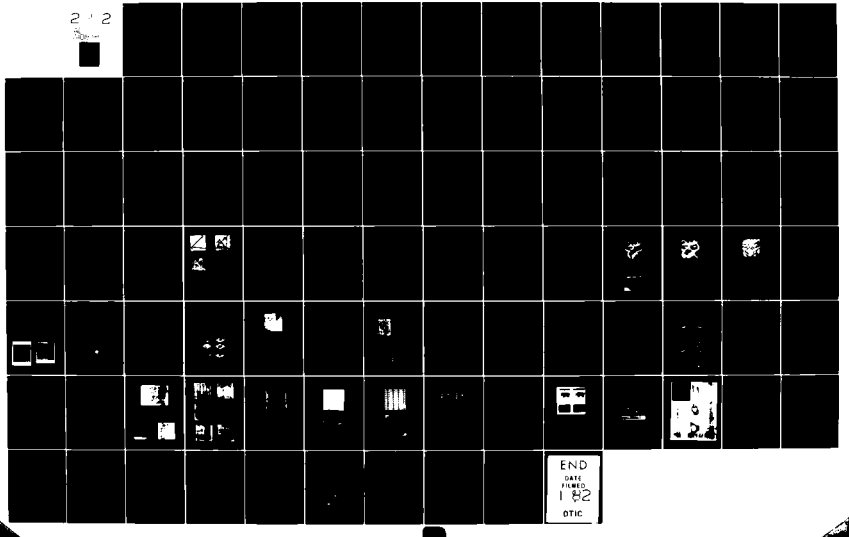
Figure 53. Length and slope of a texel in the image.

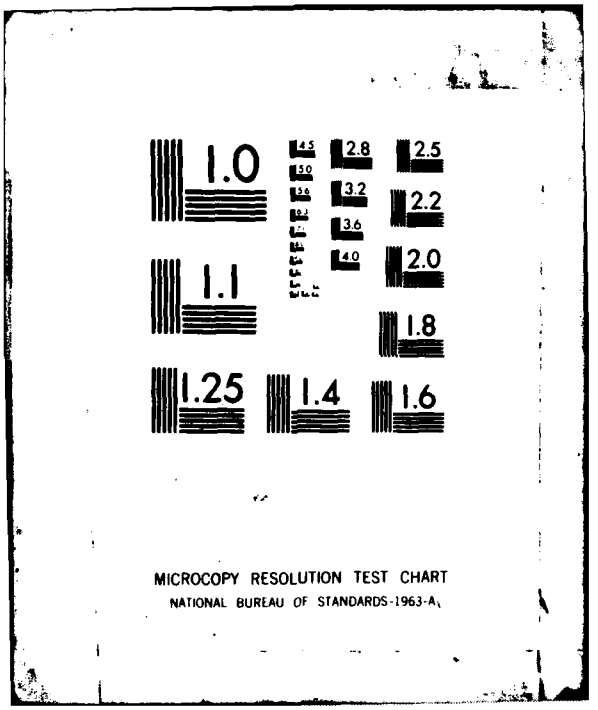
$$p^2 - q^2 - 2pq = 0.$$

We now consider the NTPM corresponding to image slope α of the texel shown in figure 53. Consider a scene-based coordinate system defined by the normal to the planar facet, the line of steepest descent of the facet, and a direction chosen to make a right handed system. The gradient line has direction ratios $\underline{l} = (p, q, p^2 + q^2)$. The normal to the plane is $\underline{n} = (p, q, -1)$, and so the third direction of the scene-based coordinate system is the cross product of these two, namely $\underline{m} = (q, -p, 0)$. Consider the deprojection $\underline{v} = (\cos\alpha, \sin\alpha, d)$ of the texel shown in figure 53. Kender [KEND80, page 114] defines the slope of $\underline{v}$ to be β, where

$$\tan\beta = \frac{\underline{v}.\underline{m}}{\underline{v}.\underline{l}}.$$

AD-A108 191 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 9/2
COMPUTATIONAL APPROACHES TO IMAGE UNDERSTANDING.(U)
OCT 81 M BRADY N00014-80-C-0505
UNCLASSIFIED AI-M-653 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

If we assume that $\underline{v}$ lies in the plane, so that $\underline{v} \cdot \underline{n} = 0$, we find

$$\tan\beta = \frac{q\cos\alpha - p\sin\alpha}{(p\cos\alpha + q\sin\alpha)(1+p^2+q^2)}.$$

Applying this to the texture shown in figure 52, the slope of the horizontal texels β_0 is given by

$$\tan\beta_0 = \frac{q}{p(1+p^2+q^2)}.$$

Similarly, the slope $\beta_{\frac{\pi}{4}}$ of the slanted texels is given by

$$\tan\beta_{\frac{\pi}{4}} = \frac{q-p}{(q+p)(1+p^2+q^2)}.$$

If we assume that the texels all lie in the plane and that they form two orthogonal sets, we have

$$\tan\beta_0 \cdot \tan\beta_{\frac{\pi}{4}} = -1.$$

Solving, we get another quadratic in p and q. When combined with the length constraint we can solve up to Necker reversal. Kender points out that if perspective projection is assumed the sense of the Necker reversal is often resolved.

4.4 Shape from motion

Just as the ideas about shape from shading and edge detection described in Sections 3.1 and 3.2 lead naturally to progress on motion perception, so do the developments surrounding the primal sketch. The first treatment of this issue is due to Ullman [ULLM78], who considered the problem of establishing a correspondence between the primal sketches in two successive image frames. Ullman also studied the problem of computing the structure of a rigid body from the correspondences of a small number of points in a number of views. It turns out that remarkably few of each are required to compute rigid three-dimensional structure. In

modelling normal vision of course, sparsity of information is manifestly not the problem! A different way to view such results is that they give information about how local an algorithm to determine three-dimensional structure can be. More recently, Webb [WEBB80, WEBB81], Hoffman and Flinchbaugh [HOFF80], and Rashid [RASH80] have considered the problem of reconstructing motion in depth from the output of the correspondence computation. Flinchbaugh and Chandrasekharan [FLIN81] coin the term "dynamic primal sketch" to describe the representation they compute, since it associates an image velocity measure with every primal sketch element. Flinchbaugh and Chandrasekaran [FLIN81] have proposed a number of grouping primitives to apply to the dynamic primal sketch, analogous to those discussed above for the (static) primal sketch.

5. Modules that operate on representations of surface shape

Many of the visual processes discussed in the previous sections compute the shape of a visible surface by finding the local surface orientation everywhere within its boundaries. This includes the work of Horn and his colleagues on shape from shading (Section 3.2), the computation of shape from contour investigated by Witkin (section 4.2), and the interpretation of optical flow [PRAZ80, CLOC80]. On the other hand, shape from stereo yields disparity only at the discrete set of zero crossings. A change of coordinates can convert the angular disparities to depths, but to compute the local surface normal everywhere on the visible surface it is necessary to interpolate a smooth surface from the discrete set of given points. We shall discuss this issue below. Binocular stereo is not the only module that generates an incomplete surface orientation map. Shape from texture (section 4.3) computations yield (constrained) surface orientations only at texture points, which may be more or less densely distributed. Stevens [STEV81] considers the interpretation of surface contours, and finds that they strongly constrain the perception of the underlying surface. Horn [HORN82] and Marr [MARR78a] suggest that in addition to local surface orientation, it is advantageous to make explicit the discontinuities in surface orientation and depth. It is not yet clear how surface normals should be parameterized, nor how accurately their values should be represented. Moreover, substantial advantages are likely to accrue from

attaching texture and color descriptors to visible surfaces, but the details are as yet unclear.

One might also consider maintaining separate representations corresponding to the four (or more) channels defined in the Marr-Hildreth theory of edge detection (described in Section 3.1.1 and used in the Marr-Poggio theory of stereo). This would enable the visible surfaces in a scene to be represented at different scales. It is clear that surface information needs to be made explicit at different levels of resolution: a pebbled path may be considered approximately planar by a human who is walking along it. On the other hand, an ant or person on roller skates may find the same path extremely difficult to navigate; in such cases the path is unlikely to be perceived as planar. As this example indicates, the level of resolution of a representation is determined largely by the process operating *upon* the representation, and there has been little investigation of such processes to date. Hinton shows that different representations of the same volume and set of surfaces can have a significant influence on the difficulty of perceptual tasks [HINT79]. Similarly, we have seen that grouping processes play an important role at several stages of visual processing, from edge finding to the interpretation of texture. Such processes have not yet been extensively investigated at the level of representations of surface orientations.

Perhaps the most important operation performed by any vision system is recognition. Representations below the level of surfaces are generally too unstructured to support recognition. One notable exception to this is recognition of surface type from texture information. Interestingly, we suggested in section 3.4 that texture is a form of surface representation. It has been argued that the surface orientation map is also inappropriate, in essence because it is viewer centered. Marr [MARR78a] notes that we are capable of recognizing objects from a wide variety of views, against a wide variety of backgrounds. To achieve this, he suggests a representation which makes explicit the three dimensional ("volumetric") nature of objects. We shall consider such representations in the next Section. For the moment we need only note that it is highly non-trivial to extract volumetric representations from a surface based representation, and so practical advantages might accrue from recognition based on the surface orientation map.

The case against surface based models of objects for recognition is essentially an argument against mul-

tiple views. Horn [HORN82] notes that irrespective of the force of the argument as regards general human vision, surface based models may still support important practical applications. For example, because of the limitations imposed by methods of manufacture, many industrial parts only assume a small number of stable configurations. Symmetry further reduces the number of substantially different views of a part. Since there are typically only a small number of parts in a parts mix, one can store a representation computed from the surface orientation map corresponding to each different view of a part in each configuration. Horn further suggests that it may be sufficient to throw away positional information and model an object by the distribution of its surface normals on the Gaussian sphere [HORN82]. Figure 54 illustrates the idea.

Perhaps the most difficult problem which sighted people constantly rely on their vision systems to help them to solve is the perception or planning of movements through cluttered space. The experience of programming robots to avoid obstacles and discover a satisfactory trajectory between two positions reveals the staggering difficulty of the geometric problems involved, problems which the human visual system solves effortlessly. Space, considered as an object, typically occupies a volume and consists of a surface whose descriptions push current representational frameworks to their limits, if not far beyond them. A solid start has been made on the problems of spatial planning by Lozano-Perez [LOZA81], who represents the set of possible configurations which an object can assume in the presence of obstacles and presents efficient algorithms for computing near optimal trajectories. A further important application lies in making precise the rather vague notion of *cognitive map*. It is usually supposed [LYNC60] that this only refers to object representations. Actually it seems that we have quite considerable navigational processes which operate on the surface orientation map.

We conclude this section with a discussion of the problem of interpolating a smooth surface from a discrete set of points, such as the disparity map computed by Grimson's implementation of the Marr-Poggio theory of stereo (section 4.1). One approach might be to apply the work on Coons patches, Bézier surfaces, and Ferguson surfaces developed for work in computer aided design (CAD) and computer aided manufacture (CAM) [FAUX79]. It is however worth asking whether the interpolated surface can be constrained by what we

Figure 54. Object representation in terms of the distribution on the Gaussian sphere of its local surface normals. (Reproduced from [HORN82])

know about *human* vision, by isolating constraints that have perhaps not figured largely in the development of CAD/CAM. Essentially, two such constraints have been uncovered, and are currently receiving attention.

The first was introduced by Grimson [GRIM81]. Suppose that D_{actual} is the disparity map from which we are to interpolate a smooth surface S. Horn's work on image formation tells us how to construct the image $Im(S)$, and this enables us to compute the set of zero crossings, and hence predict a disparity map $D_{predict}$. The actual and predicted disparity maps should agree everywhere, Actually, one does not explicitly construct the image of the interpolated surface and the predicted disparity map. Rather, it is used implicitly in deriving a number of theorems which constrain the surface S. Grimson has coined a suggestive slogan for this analysis: *no information is information*, since the *absence* of an initial value at the point (x, y) in the actual disparity map means that the gradient of the interpolated surface S *cannot change too rapidly there.*

The second constraint is based on the idea that the human visual system constructs *the most conservative* solution consistent with the data. Figure 55 is reproduced from [BARR81b], and shows a set of possible space curves, all of which produce an elliptical image. Significantly, we are unaware of most such possibilities, especially those that are discontinuous. We are able to interpolate smooth curves and surfaces without involving rich semantics. It also seems that the shape of the boundary plays the most significant role in determining the interpolated surface (see for example figure 56, which is reproduced from [BARR81b]. Taken together, these ideas suggest that the interpolation process can be modelled in terms of the calculus of variations (see for example [COUR37, volume I]).

The idea is to choose an appropriate "performance index" P and define the interpolated surface to be that which minimizes the integral of P subject to the boundary constraints. This idea has been explored by a number of authors. Unlike the ordinary differential calculus, it is not generally the case that a minimal surface exists, even for "plausible" performance indices. For example, it is not clear that there is a unique surface that minimizes the integral of the Gaussian curvature. Grimson [GRIM81] notes that the existence of a minimizing surface can be formally guaranteed if the performance index satisfies the technical condition of being a seminorm. He suggests the quadratic variation, which is defined to be $f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2$, and shows

how to construct the iteration operator shown in figure 57. The square Laplacian $f_{xx}^2 + f_{yy}^2$ also satisfies the seminorm condition. Brady and Horn [BRAD81b] show that any quadratic form in the second derivatives f_{xx}, f_{xy}, and f_{yy} is a seminorm and leads to a unique minimal surface. They further show that the rotationally symmetric performance indices form a vector space spanned by the quadratic variation and the square Laplacian. Since both operators satisfy the same Euler equation $\Delta^2 f = 0$, they cannot be distinguished away from given boundary points. Brady and Horn apply the statics of a thin plate to show that the quadratic variation provides the tighter constraint. Grimson notes that the null space of the quadratic variation is larger than that of the square Laplacian, containing for example the function $f(x, y) = xy$ [GRIM81]. He has worked out several examples showing that the quadratic variation leads to surfaces that accord better with human intuition. Brady and Grimson (forthcoming) use these ideas about surface interpolation to propose that subjective contours arise from surface perception.

Barrow and Tenenbaum [BARR81b] observe that in order to interpolate the circular cross section of a cylinder and sphere it is sufficient to assume that the curvature varies linearly in the image. They suggest that in general one should choose a linear expression for the curvature to minimize the least squares error. Brady, Grimson, and Langridge [BRAD80b] use an approximation to the one dimensional quadratic variation f_{xx}^2 to argue that subjective contours are cubics. The exact minimal integral curvature curve has recently been found by Horn [HORN81b].

6. Viewpoint independent representations of objects

The surface based representations discussed in the previous section are different for each particular viewpoint. Each viewpoint of each viewer in a scene defines a coordinate frame in terms of which the points that are visible from that viewpoint can be described. Other coordinate frames are naturally associated with the objects and surfaces in a scene, and it is often more convenient to describe relative positions and movements in those frames rather than in the ones lined up with a particular viewpoint. In many scenes there is a natural "global" coordinate frame that is independent of any viewpoint. For example, an airplane or ship has an

Figure 55. An elliptical image, and some of the space curves that might have generated it. (Reproduced from [BARR81b, figure 3-2]

Figure 56. Interpolation of a cylinder from a number of stimuli, including a silhouette, and half tone images produced from a variety of reflectance functions. (Reproduced from [BARR81b, figure 2.3]

Figure 57. The surface interpolation operator derived by Grimson from minimizing quadratic variation.

associated frame defined by its bow, stern, starboard, port, up, and down; rotations about those axes specify the yaw, roll, and pitch. A football field or a room has a natural frame defined by the sidelines or walls and by the gravitational vertical.

Points can be represented in homogeneous coordinates, for example, and frame transformations by 4×4 matrices that consist of a translation, a rotation, and a scale factor. This approach has proved valuable in

computer graphics [CARL78] and robotics [PAUL79]. Rotations can also be described as quaternions with a saving of storage [TAYL79, BROO80]. Frames can specify the transformation to scene coordinates, and hence by composition relate different viewpoints. Brooks and Binford [BROO80] note that one important use of inter-relating frames by composition is to make affixment relations explicit. The coordinate frame local to an airplane needs to be related to that defined by the runway on which it stands. The programming language AL [FINK74] was the first to provide a mechanism for the automatic maintenance of affixment relations.

Most objects are composed of connected parts, each of which can be described in its own local frame. A person has two arms, each of which is further subdivided into an upper arm, a forearm, and a hand. Like any structured representation, the important issues concern the choice of "primitives" and the means by which one part of a representation is related to another. Consider the latter issue first. Work in Robotics has adopted the Hartenberg-Denavit notation for kinematic chains to describe the geometric inter-relationships between successive links of an arm, a leg, or the several legs of a mobile robot [PAUL79]. Marr and Nishihara's suggestion [MARR78b] is a special case of this notation.

One approach to primitives is to consider objects to be composed of instances of a small set of prototype volumes, such as spheres, blocks, and triangular prisms [BRAI73]. This approach has been much used in CAD/CAM. The problem is that even simple objects have a complex description. One might add more and more primitives, such as truncated cones and pyramids, to reduce this complexity. Binford [BINF71] suggested another approach that has proved very fruitful. He introduced a more general class of volumes called *generalized cones* which includes as subclasses the primitive volumes mentioned previously. A generalized cone describes a volume by sweeping a cross section area along a space curve, called the "spine", while deforming it according to some sweeping rule. Figure 58 is reproduced from [BROO81] and shows a number of generalized cones. Notice that although elongation is the characteristic property of generalized cones, they are not necessarily elongated. Nor do they require a circular cross section. Nevertheless, generalized cones are particularly well suited to describing objects which have a natural axis. This certainly includes growth structures. Hollerbach [HOLL75] noted that Greek amphora are also well described by generalized cones, the

spine being a result of the process of manufacture on the potters wheel. Similar considerations apply to objects turned on a lathe or produced by extrusion. Conversely, objects produced by moulding, beating, welding, or sculpture tend to be awkwardly described in terms of generalized cones.

A major issue in description and recognition arises from the vast number of objects that we can distinguish. This leads to an enormous data base of models and makes the indexing process of crucial importance. The problem is ubiquitous in artificial intelligence and has produced a number of schemes for matching on the basis of partial descriptions. One recurrent theme is the use of abstraction to produce a smaller search space, the solution being used to guide further search in a less abstracted version. At a suitably high level of abstraction this can be recognized as the process which underlies the matcher in the Marr-Poggio theory of stereo described in Section 4.1. In the specific case of vision, Nevatia and Binford [NEVA77] and Marr and Nishihara [MARR78b] discuss various schemes for indexing. Agin [AGIN72], Nevatia and Binford [NEVA77], and Marr and Nishihara [MARR78b] note that a kinematic linkage can generally be approximated by a single cone. Such approximate descriptions provide for hierarchical descriptions at a useful variety of scales. Often, the most useful approximation is based on the most proximal link, more detailed descriptions deriving from applying the same process to the distal links of the chain. Brooks and Binford [BROO80] use subcategories of objects to achieve property inheritance and facilitate indexing. For example, they exploit the fact that a Boeing 747-SP is a special kind of Boeing 747 (with slight variations pertinent to recognizing one), and a Boeing 747 is a special kind of wide bodied jet (distinguished from other aircraft such as Boeing 727's on the basis of overall length and width to length ratio.)

Brooks and Binford [BROO80, BROO81] draw attention to the need to incorporate constraints into object descriptions. For example, a person has two legs which are of (roughly) the same length, and are roughly as long as the person's body. The actual sizes scale with (*a priori* unknown) camera position. As usual, constraints propagate. For example, the engine pods of a jet are deployed symmetrically on the front wings on either side of the fuselage. Finding an aircraft wing constrains the overall scale of the aircraft, and hence the length of the fuselage. Such constraints are represented naturally by numerical inequalities. Brooks [BROO81]

Figure 58. An indication of the range of objects which can be modelled simply using generalized cones. (Reproduced from [BROO81])

describes a program that determines the solutions of a set of such inequalities. If an object recognized as a person's body is much larger than one thought to be a tree, then the person is probably much nearer than the tree. Mechanisms for taking into account relatively remote possibilities such as giants and toy trees have been proposed (for example, [ANDE81].

Finally, we consider the process of extracting from an image the spine, cross section function, and sweeping rule which define a generalized cone. The work on this problem to date requires a number of simplifying assumptions. For example, Nevatia and Binford implicitly assume that the cross section function is circular [NEVA77]. Marr [MARR77] considered the problem in considerable detail and showed how, in a restricted case, a straight spine can be extracted from the inflection points on the bounding contour of an object. Brady showed that the spine can be extracted more reliably by using stationary points of curvature [BRAD79b]. Marr's work assumes that the bounding contour is planar, which is overly restrictive [BRUS81]. He also proposed a classification of the images of the joins between two straight spine cones.

Acknowledgements

The research for this paper was done at the artificial intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643. Larry Druffel encouraged me to write this paper and offered several insightful comments on an early draft. The referees made useful suggestions for improving and extending the paper. John Batali, Rod Brooks, Ken Forbus, Eric Grimson, Ellen Hildreth, and Bob Sjoberg made valuable comments on drafts. I gratefully acknowledge interactions with Harry Barrow, Tom Binford, Jerry Feldman, Eric Grimson, Ellen Hildreth, Berthold Horn, David Marr, Ram Nevatia, Marty Tenenbaum, and Patrick Winston during the preparation of this paper. Patrick Winston and Bob Berwick encouraged me to write English in the quaint style enjoyed by Americans. Al Borning made a number of editorial comments that improved the paper. I thank Judi Jones for preparing the figures.

References

[AGIN72] Agin G. J. Representation and description of curved objects, Stanford University, AIM 73, 1972.

[ANDE81] Anderson J. A. and Hinton G. E. Models of information processing in the brain, *Parallel models of associative memory* eds. G. E. Hinton and Anderson J. A., Erlbaum, 1981.

[ARNO78] Arnold D. Local context in matching edges for stereo vision, *Proceedings of the Image Understanding Workshop* ed. Lee Baumann S., May, 1978.

[BAJC73] Bajcsy R. Computer description of textured surfaces, Proc. Int. Jt. Conf. Artif. Int. 1973, 3, 572-579.

[BAJC76a] Bajcsy R., and Lieberman L. Texture gradient as a depth cue, Comp. Graphics and Image Proc. 1976, 5, 52-67.

[BAJC75] Bajcsy R., and Tavakoli M. Image filtering-a context dependent process, IEEE Trans. Circuits and Systs. 1975, CAS-22, 463-474.

[BAJC76b] Bajcsy R., and Tavakoli M. Computer recognition of roads from satellite pictures, IEEE Trans. Syst. Man and Cyb. 1976, SMC-6, 623-637.

[BAKE81] Baker H. Harlyn, and Binford T. O. Depth from edge and intensity based stereo, Int. Jt. Conf. Artif. Intell. 1981, 6, .

[BALL 79] Ballard D. H. Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 1979, , .

[BALL 80] Ballard D. H. Parameter networks: towards a theory of low level vision, University of Rochester, TR-75, 1980.

[BARR71] Barrow H. G. and Popplestone R. J. Relational Descriptions in Picture Processing, Machine Intelligence 1971, 6, 377-396.

[BARR78] Barrow H. G., and Tenenbaum J. M. Recovering intrinsic scene characteristics from images, *Computer vision systems* eds. Hanson A. and Riseman E. M., 1978.

[BARR81a] Barrow H. G. and Tenenbaum J. M. Computational vision, Proc. IEEE 1981, 6, 572-595.

[BARR81b] Barrow H. G. and Tenenbaum J. M. Interpreting line drawings as three dimensional surfaces, Artificial Intelligence 1981, 17, 75-117.

[BARR77] Barrow H. G. et. al. Interactive aids for Cartography and Photointerpretation: progress report, *Proceedings of the Image Understanding Workshop* ed. Baumann Lee, Science Applications, 1977.

[BATA81] Batali J. A vision chip, MIT, 1981.

[BINF71] Binford T. O. Visual Perception by Computer, Proc. IEEE Conf. Systems and Control 1971, , .

[BINF81] Binford T. O. Inferring surfaces from images, Artificial Intelligence 1981, 17, 205-245.

[BIRK81] Birk J., Dessimoz J. D., and Kelley R. B. General methods to enable robots with vision to acquire, orient, and transport workpieces, NSF Grantees Conference on Production Research and Technology 1981, 8, u1-u7.

[BOLL80] Bolles R. Locating partially visible objects: the local feature method, Proc. AAAI 1980, 1, 41-44.

[BOLL81] Bolles R. Three-dimensional locating of industrial parts, NSF Grantees Conference on Production Research and Technology 1981, 8, w1-w5.

[BOLL78] Bolles R., Quam L. H., Fischler M. A., and Wolf H. C. The SRI Road Expert: Image to Database correspondence, *Proceedings of the Image Understanding Workshop* ed. Baumann Lee, Science Applications, 1978.

[BRAC65] Bracewell R. The Fourier transform and its applications, New York, MacGraw-Hill, 1965.

[BRAD79a] Brady Michael The development of a Image Understanding system, Ricerche di Psicologica 1979,

II, 6, 113-148.

[BRAD79b] Brady Michael Finding the axis of an egg, Proc. Int. J. Conf. Artif. Int. 1979, 6, 85-87.

[BRAD80a] Brady Michael Toward a computational theory of early visual processing in reading, MIT, AI Memo 593, 1980.

[BRAD81a] Brady Michael The changing shape of Image Understanding, Artificial Intelligence 1981, 17, 1-17.

[BRAD81b] Brady Michael, and Horn B. K. P. Rotationally symmetric operators for surface interpolation, MIT, AI Memo ???, 1981.

[BRAD78] Brady Michael and Wielinga B. J. Reading the writing on the wall, *Image Understanding Systems* eds. Hanson and Riseman, Academic Press, 1978.

[BRAD80b] Brady Michael, Grimson W. E. L., and Langridge D. The shape of subjective contours, Proc. AAAI 1980, 1, 15-17.

[BRAI73] Braid I. Designing with volumes, Cambridge UK, Cantab, 1973.

[BRIC70] Brice C. R. and Fennema C. L. Scene Analysis using Regions, Artificial Intelligence 1970, 1, 205-226.

[BROO78] Brooks M. J. Rationalizing edge detection, Computer Graphics and Image Processing 1978, 8, 277-285.

[BROO79a] Brooks M. J. Surface normals from closed paths, Proc. Int. Jt. Conf. Artificial Intelligence 1979, 6, 98-101.

[BROO81] Brooks R. Symbolic reasoning among 3-D models and 2-D images, Artificial Intelligence 1981, 17, 285-349.

[BROO79b] Brooks R., Greiner Russell, and Binford T. O. The ACRONYM model based vision system, Proc.

Int. Jt. Conf. Artificial Intelligence 1979, 6, 105-113.

[BRUS81] Bruss Anna R. The image irradiance equation: its solution and application, MIT, 1980.

[CARL78] Carlbom I. and Paciorek J. Planar geometric projections and viewing transformations, Computing Surveys 1978, 10, 465-502.

[CLOC80] Clocksin W. F. Perception of Surface Slant and Edge Labels from Optical Flow: a computational approach, Perception 1980, 9, 253-269.

[CLOW71] Clowes M. B. On seeing things, Artificial Intelligence 1971, 2, 79-112.

[COUR37] Courant R., and Hilbert D. Methods of mathematical physics, New York, John Wiley, 1937.

[DAVI81] Davis Larry S. and Rosenfeld Azriel Cooperating processes for low-level vision: a survey, Artificial Intelligence 1981, 17, 245-265.

[DEV75] Dev P. Perception of depth surfaces in random-dot stereograms: a neural model, Int. J. Man-Machine Studies 1975, 7, 511-528.

[DRAP80] Draper S. W. Reasoning about depth in line-drawing interpretation, Sussex University, UK, 1980.

[DRAP81] Draper S. W. The use of gradient and dual space in line-drawing interpretation, Artificial Intelligence 1981, 17, 461-508.

[DUDA73] Duda R. O., and Hart P. E. Pattern classification and scene analysis, New York, John Wiley and sons, 1973.

[FAUX79] Faux I. D. and Pratt M. J. Computational Geometry for design and manufacture, Chichester UK, Ellis Horwood, 1979.

[FINK74] J Finkel R., Taylor R. H., Bolles R., Paul R., and Feldman J. A. Al., A programming system for automation, Stanford University, AIM 177, 1974.

[FISC80] Fischler M. A. and Bolles R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Proceedings of the Image Understanding Workshop* ed. Baumann Lee, Science Applications, 1980.

[FLIN81] Flinchbaugh B. E. and Chandrasekharan B. A theory of spatio-temporal aggregation for vision, Artificial Intelligence 1981, 17, 387-409.

[GENN79] Gennery D. B. Stereo-camera calibration, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 101-107, 1979.

[GIBS50] Gibson J. J. The Perception of the Visual World, Boston, Mass, Houghton-Mifflin, 1950.

[GIBS66] Gibson J. J. The senses considered as perceptual systems, Boston, Mass, Houghton-Mifflin, 1966.

[GRIM80] Grimson W. E. L. Computing shape using a theory of human stereo vision, MIT, 1980.

[GRIM81] Grimson, W. E. L. From images to surfaces: a computational study of the human early visual system, Cambridge, MIT Press, 1981.

[HANN74] Hannah M. J. Computer matching of areas in stereo imagery, Stanford, 1974.

[HANS77] Hanson Allen R., and Riseman E. M. VISIONS: a computer system for interpreting scenes, *Computer vision systems* eds. Hanson Allen R., and Riseman E. M., 303-335, 1977.

[HARA79] Haralick R. M. Statistical and structural approaches to texture, Proc. IEEE 1979, 67, 786-804.

[HARA80] Haralick R. M. Edge and region analysis for digital image data, Comp. Graphics and Im. Proc. 1980, 12, 60-73.

[HARA73] Haralick R. M., and Bosley R. Texture features for image classification, ERTS Symp. 1973, 3, 1929-1969.

[HEND78] Henderson R. L., Miller W. J., and Grosch C. B. A flexible approach to digital stereo mapping,

Photogrammetry Eng. and Remote Sensing 1978, 44, 1499-1512.

[HERS71] Herskovits A and Binford T. O. On Boundary Detection, MIT, AI Memo 183, 1970.

[HILD80] Hildreth E. C. Implementation of a theory of edge detection (MS dissertation), also AI-TR 579, MIT, 1980.

[HINT79] Hinton G. E. Some demonstrations of the effects of structural descriptions in mental imagery, Cognitive Science 1979, 3, 231-250.

[HOFF80] Hoffman D. D. and Flinchbaugh B. E. The interpretation of biological motion, MIT, AI Memo 608, 1980.

[HOLL75] Hollerbach J. M. Hierarchical Shape Description of Objects by selection and modification of Prototypes (MS Dissertation), also AI-TR-346, MIT, 1975.

[HORN73] Horn B. K. P. The Binford-Horn line-finder, MIT, AI Memo 285, 1973.

[HORN74] Horn B. K. P. Determining lightness from an image, Computer Graphics and Image Processing 1974, 3, 277-299.

[HORN75] Horn B. K. P. Obtaining Shape from Shading Information, *The Psychology of Image Understanding* ed. Winston P. H., McGraw-Hill, 1975.

[HORN77] Horn B. K. P. Understanding image intensities, Artificial Intelligence 1977, 8, 201-231.

[HORN81a] Horn B. K. P. Hill shading and the reflectance map, Proc. IEEE 1981, 19, 14-47.

[HORN81b] Horn B. K. P. The curve of least energy, MIT, AI Memo 610, 1981.

[HORN82] Horn B. K. P. Sequins and Quills - Representations for Surface Topography, *Representation of 3-Dimensional Objects* ed. Bajcsy R., Springer Verlag, 1982.

[HORN78a] Horn B. K. P. and Bachman B. L. Using synthetic images to register real images with surface

models, Comm. ACM 1978, 21, 914-924.

[HORN81c] Horn B. K. P. and Schunck B. G Determining optical flow, Artificial Intelligence 1981, 17, .

[HORN79] Horn B. K. P. and Sjoberg Robert W. Calculating the reflectance map, Appl. Optics 1979, 18, 1770-1779.

[HORN80] Horn B. K. P. and Sjoberg Robert W. Atmospheric modelling for the generation of albedo images, *Proceedings of the Image Understanding Workshop* ed. Baumann Lee, Science Applications, 1980.

[HORN78b] Horn B. K. P., Woodham R. J., and Silver W. M. Determining shape and reflectance using multiple images, MIT, AI Memo 490, 1978.

[HUEC71] Hueckel M. An operator which locates edges in digital pictures, J. Assoc. Comp. Mach. 1971, 18, 113-125.

[HUFF71] Huffman D. A. Impossible Objects as Nonsense Sentences, *Machine Intelligence 6*, eds. Meltzer B. and Michie D., Edinburgh University Press, 1971.

[HUFF77] Huffman D. A. A Duality Concept for the analysis of Polyhedral Scenes, *Machine Intelligence 8*, eds. Elcock E. W. and Michie D., Chichester: Ellis Horwood, 1977.

[HUMM79] Hummel Robert A. Feature detection using basic functions, Comp. Graphics and Im. Proc. 1979, 9, 40-55.

[HUMM80] Hummel Robert A., and Zucker Steven W. On the foundations of relaxation labelling processes, McGill University, Computer Vision and Graphics Lab., TR-80-7, 1980.

[IKEU81a] Ikeuchi K. Determination of surface orientations of specular surfaces by using the photometric stereo method, IEEE (accepted for publication) 1981, , .

[IKEU81b] Ikeuchi K. and Horn B. K. P. Numerical shape from shading and occluding boundaries, Artificial Intelligence 1981, 17, 141-185.

[JERN81] Jernigan Marvin E., and Wardell Ron W. Does the eye contain optimal edge detection mechanisms?, IEEE Trans. Syst. Man and Cyb. 1981, SMC-11, 441-444.

[JULE62] Julesz B. Visual pattern discrimination, IRE Trans. Inf. Theory 1962, 8, 84-92.

[JULE71] Julesz B. Foundations of Cyclopean Perception, Chicago, The University of Chicago Press, 1971.

[JOHA73] Johansson G. Visual perception of Biological Motion and a model for its analysis, Perception and Psychophysics 1973, 14, 201-211.

[KANA80] Kanade T. A theory of origami world, Artificial Intelligence 1980, 13, 279-311.

[KANA81] Kanade T. Recovery of the three dimensional shape of an object from a single view, Artificial Intelligence 1981, 17, 409-461.

[KEND80] Kender J. Shape from texture, Carnegie Mellon University, 1980.

[KEND80b] Kender J., and Kanade T. Mapping image properties into shape constraints: skew symmetry, affine transformations, and the shape-from-texture paradigm, Proc. AAAI, 4-7, 1980.

[KOEN75] Koenderink J. J., and van Doorn A. J. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer, Optica Acta 1975, 22, 772-791.

[KOEN76] Koenderink J. J., and van Doorn A. J. Local structure of movement parallax of the plane, J. Opt. Soc. Am. 1976, 66, 717-723.

[KOEN77] Koenderink J. J., and van Doorn A. J. How an ambulant observer can construct a model of the environment from the geometric structure of the visual inflow, *Kybernetik* eds. Hauske and Butenandt, 224-247, 1977.

[LAND71] Land E. H. and McCann J. J. Lightness and Retinex Theory, J. Optical Society of America 1971, 61, 1-11.

[LESS77] Lesser Victor R. and Erman Lee D. A retrospective view of the Hearsay-II architecture, Proc. Int. Jt. Conf. Artificial Intelligence 1977, 2, 790-800.

[LONG80] Longuet-Higgins H. C. and Prazdny K. F. The interpretation of moving retinal images, Proc. R. Soc. London 1980, 208, 385-387.

[LOWE81] Lowe D. G., and Binford T. O. The interpretation of geometric structure from image boundaries, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 39-46, 1981.

[LOZA80] Lozano-Perez T. Spatial Planning: a Configuration Space approach, MIT, AI Memo 605, 1980.

[LUCA81] Lucas B. D., and Kanade T. An iterative image registration technique with application to stereo vision, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 121-130, 1981.

[LYNC60] Lynch K. The Image of the City, Cambridge MA, MIT Press, 1960.

[MACK73] Mackworth A. K. Interpreting Pictures of Polyhedral Scenes, Artificial Intelligence 1973, 4, 121-137.

[MACV81] MacVicar-Whelan P. J., and Binford T. O. Line finding with subpixel precision, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 26-31, 1981.

[MAHO81] Mahoney William C. DMA overview of MC&G applications of digital image pattern recognition, Proc. SPIE conference, Washington, DC 1981, , .

[MALE77] Maleson J., Brown C., and Feldman J. Understanding natural texture, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 19-27, 1977.

[MARR76a] Marr D. Early processing of visual information, Phil. Trans. R. Soc. Lond. B 1976, 275, 483-524.

[MARR77] Marr D. Analysis of occluding contours, Proc. R. Soc. Lond. B 1977, 197, 441-475.

[MARR78a] Marr D. Representing Visual Information, *Image Understanding Systems* eds. Hanson and

Riseman, Academic Press, 1978.

[MARR82] Marr D. Vision, San Francisco, Freeman, 1982.

[MARR80a] Marr D. and Hildreth E. Theory of edge detection, Proc. R. Soc. Lond. B 1980, 207, 187-217.

[MARR79a] Marr D., Hildreth E., and Poggio T. Evidence for a fifth, smaller channel in early human vision, MIT, AI Memo 541, 1979.

[MARR78b] Marr D. and Nishihara H. K. Representation and recognition of the spatial organization of three dimensional structure, Proc. R. Soc. London B 1978, 200, 269-294.

[MARR76b] Marr D. and Poggio T. Cooperative computation of stereo disparity , Science 1976, 194, 283-287.

[MARR76c] Marr D. and Poggio T. Cooperative computation of stereo disparity, MIT, AI Memo 364, 1976.

[MARR79b] Marr D. and Poggio T. A theory of human stereo vision, Proc. R. Soc. Lond. B 1979, 204, 301-328.

[MARR80b] Marr D., Poggio T., and Hildreth E. Smallest channel in early human vision, J. Opt. Soc. Am. 1980, 70, 868-870.

[MARR81] Marr D. and Ullman S. Directional selectivity and its use in early visual processing, Proc. R. Soc. Lond. B 1981, 211, 151-180.

[MARS73] Marshall J. C. and Newcombe F. Patterns of Paralexia, Journal of Psycholinguistic Research 1973, 2, 175-199.

[MAYH81] Mayhew J. and Frisby J. P. Psychophysical and Computational studies toward a theory of human stereopsis, Artificial Intelligence 1981, 17, 349-387.

[MCCA74] McCann J. J., Savoy R. L., Hall J. A. Jr., and Scarpetti J. J. Visibility of continuous luminance gradients, Vision Research 1974, 14, 917-927.

[MINS72] Minsky, M. and Papert, S. Artificial Intelligence progress report, MIT, AI Memo 252, 1972.

[MORA79] Moravec H. Visual mapping by a robot rover, Proc. Int. Jt. Conf. Artif. Int. 1979, 6, 598-600.

[MORA80] Moravec Hans P. Obstacle avoidance and navigation in the real world by a seeing robot rover, Stanford University, AIM-304a, 1980.

[NAKA74] Nakayama K. and Loomis J. M. Optical velocity patterns, velocity sensitive neurons and space perception, Perception 1974, 3, 63-80.

[NEVA77] Nevatia R. and Binford T. O. Description and Recognition of curved objects, Artificial Intelligence 1977, 8, 77-98.

[NEVA78] Nevatia R. and Babu K. R. Linear feature extraction, *Proceedings of the Image Understanding Workshop* ed. Baumann Lee, Science Applications, November 1978.

[NEWM73] Newman W. M., and Sproull R. F. Principles of interactive computer graphics, New York, McGraw-Hill, 1973.

[NISH81] Nishihara H. K. and Larson N. G. Toward a real time implementation of the Marr-Poggio stereo matcher, *Proceedings of the Image Understanding Workshop* eds. Lee Baumann, 1981.

[NUDD79] Nudd G. R., Fouse S. D., Nussmeier T. A., and Nygaard P. A. Development of Custom-Designed Integrated Circuits for Image Understanding, *Proceedings of the Image Understanding Workshop* eds. Lee Baumann, 1979.

[OCAL74] O'Callaghan J. F. Recovery of perceptual shape organization from simple closed boundaries, Computer Graphics and Image Processing 1974, 3, 300-312.

[OGOR76] O'Gorman F. Edge detection using Walsh functions, Proc. second AISB Conf., Edinburgh 1976, , .

[OGOR73] O'Gorman F., and Clowes M. B. Finding picture edges through collinearity feature points, Proc. Int. Jt. Conf. Artif. Int., 543-555, 1973.

[OHLA75] Ohlander R. Analysis of Natural Scenes, Carnegie Mellon, 1975.

[OHLA78] Ohlander R., Price K., and Reddy D. Raj Picture segmentation using a recursive region splitting method, Comp. Graphics and Im. Proc. 1978, 8, 313-334.

[PAUL79] Paul R. P. Manipulator Cartesian Path Control, IEEE Trans. Sys. Man and Cyb. 1979, SMC-9, 702-711.

PAVL78] Pavlidis Theodosios Structured pattern recognition, Heidelberg, New York, Springer Verlag, 1978.

[PAVL80] Pavlidis Theodosios Algorithms for shape analysis of contours and waveforms, IEEE Trans. PAMI 1980, 2, 301-312.

[PRAT78] Pratt William K. Digital image processing, New York, Wiley Interscience, 1978.

[PRAZ80] Prazdny K. F. Egomotion and relative depth map fom Optical Flow, Biological Cybernetics 1980, 36, 87-102.

[PREW70] Prewitt J. M. S. Object enhancement and extraction, *Picture processing and psychopictorics* eds. Lipkin R. S., and Rosenfeld Azriel, 75-149, 1970.

[QUAM71] Quam L. H. Computer comparison of pictures, Stanford, 1971.

[RASH80] Rashid R. F. Towards a system for the interpretation of moving light displays, IEEE Pattern Anal. and Mach. Intell. 1980, PAMI-2, 574-581.

[REDD78] Reddy Raj Pragmatic aspects of Machine Vision, *Image Understanding Systems* eds. Hanson and Riseman, Academic Press, 1978.

[RICH80] Richter J. and Ullman S. A model for the spatio-temporal organization of X and Y-type Ganglion cells in the primate retina, MIT, AI Memo 573, 1980.

[ROBE62] Roberts L. G. Machine processing of three-dimensional solids, *Optical and electro optical informa-*

tion processing eds. Tippett M. , 1962.

[ROSE76] Rosenfeld Azriel and Kak Avinash C. Digital Picture Processing, New York, Academic, 1976.

[ROSE71] Rosenfeld Azriel, and Thurston M. Edge and curve detection for visual scene analysis, IEEE Trans. Computers 1971, C-20, 562-569.

[ROSE70] Rosenfeld Azriel, and Troy E. Visual texture analysis, Symp. Feature Extraction and Selection in Pattern Recog. 1970, 1, 115-124.

[SCHA77] Schatz Bruce R. The computation of immediate texture discrimination, MIT AI Memo, 426, 1977.

[SELF81] Selfridge P. G., and Sloan K. R. Reasoning about images: application to aerial image understanding, *Proc. Image Understanding Workshop* eds. Lee Baumann S., 1981.

[SHAN79] Shanmugam K. Sam, Dickey Fred M., and Green James A. An optimal frequency domain filter for edge detection in digital pictures, IEEE Trans. Patt. Anal. and Mach. Intell. 1979, PAMI-1, 37-49.

[SHAP79] Shapiro Linda A structural model for shape, IEEE Trans. PAMI 1980, 2, 111-126.

[SHIR73] Shirai Y. A Context-sensitive line finder for Recognition of Polyhedra, Artificial Intelligence 1973, 4, 95-119.

[SILV80] Silver W. M. Determining shape and reflectance using multiple images, MIT, 1980.

[SLOA80] Sloan K. R. and Ballard D. H. Experience with the generalized Hough transform, *Proceedings of the Image Understanding Workshop* eds. Lee Baumann, 1980.

[STEV80] Stevens K. A. Surface Perception by local analysis of Texture and Contour, MIT AI Lab. Technical Report, 512, 1980.

[STEV81a] Stevens K. A. Evidence relating subjective contours and interpretations involving occlusion, MIT AI Lab. Memo, 637, 1981.

[STEV81b] Stevens K. A. The visual interpretation of surface contours, Artificial Intelligence 1981, 17, 47-75.

[SPAC79] Spacek L.. A. Shape from Shading and more than one view, M. Sc. Thesis, University of Essex, UK, , 1979.

[STRA79] Strat T. M. A numerical method for shape from shading from a single image, MIT, 1979.

[SUGI78] Sugihara K. Picture language for skeletal polyhedra, Comp. Graphics and Image Proc. 1978, 8, 382-405.

SUGI81] Sugihara K. Algebraic and combinatorial approach to the analysis of line drawings of polyhedra, Electrotechnical Lab., Japan, RMI 81-02, 1981.

[TAYL79] Taylor Russell H. Planning and execution of straight line manipulator trajectories, IBM J. of Res. and Dev. 1979, 23, 424-433.

[TENE77] Tenenbaum J. M., and Barrow H. G. Experiments in interpretation guided segmentation, Artificial Intelligence 1977, 3, 241-274.

[TOMI73] Tomita F., Yachida M., and Tsuji S. Detection of homogeneous regions by structural analysis, Proc. Int. Jt. Conf. Artif. Int. 1973, 3, 564-571.

[TURN74] Turner K. J. Computer perception of curved objects using a television camera, Edinburgh University, 1974.

[ULLM78] Ullman S. The interpretation of visual motion, Cambridge, Mass., MIT Press, 1978.

[VILN81] Vilnrotter F., Nevatia R., and Price K. E. Structural analysis of natural textures, *Proc. Image Understanding Workshop* ed. Lee Baumann S., 61-68, 1981.

[WALT72] Waltz D. L. Generating semantic descriptions from drawings of scenes with shadows, MIT, 1972.

[WEBB80] Webb J. Static analysis of moving jointed objects, Proc. AAAI 1980, 1, 35-37.

[WEBB81] Webb J., and Aggarwal J. K. Visual interpretation of the motion of rigid and jointed objects, University of Texas, Austin, , 1981.

[WEIS74] Weiskrantz L., Warrington E. K., Sanders M. D., and Marshall J. Visual capacity in the hemianopic field following a restricted occipital ablation, Brain 1974, 97, 709-728.

[WEIS73] Weisstein N Beyond the Yellow Volkswagen Detector and the Grandmother Cell: A general Strategy for the Exploration of Operations in Human Pattern Recognition, *Contemporary Issues in Cognitive Psychology* ed. Solso R., W. H. Winston and Sons, 1973.

[WESK76] Weska J., Dyer C., and Rosenfeld Azriel A comparative study of texture measures for terrain classification, IEEE Trans. Syst. Man and Cyb. 1976, SMC-6, 269-285.

[WILS79] Wilson H. R. and Bergen J. R. A four mechanism model for spatial vision, Vision Research 1979, 19, 19-32.

[WILS77] Wilson H. R. and Giese S. C. Threshold visibility of frequency gradient patterns, Vision Research 1977, 17, 1177-1190.

[WITK80] Witkin Andrew P. A statistical technique for recovering surface orientation from texture in natural imagery, Proc. AAAI 1980, 1, 1-3.

[WITK81] Witkin Andrew P. Recovering surface shape and orientation from texture, Artificial Intelligence 1981, 17, 17-47.

[WOOD81] Woodham R. J. Analyzing images of curved surfaces, Artificial Intelligence 1981, 17, 117-141.

[ZUCK76] Zucker S. W. Toward a model of texture, Comp. Graphics and Im. Proc. 1976, 5, 190-202.

[ZUCK77] Zucker S. W., Hummel R. A., and Rosenfeld Azriel An application of relaxation labelling to line and curve enhancement, IEEE Trans. Computers 1977, C-26, 394-403, 922-929.

[ZUCL81] Zucker S. W., Leclerc Y., and Mohammed J. Relaxation labelling and local maxima selection:

conditions for equivalence, IEEE Trans. Patt. Anal. and Mach. Int. 1981, , .

fig 1

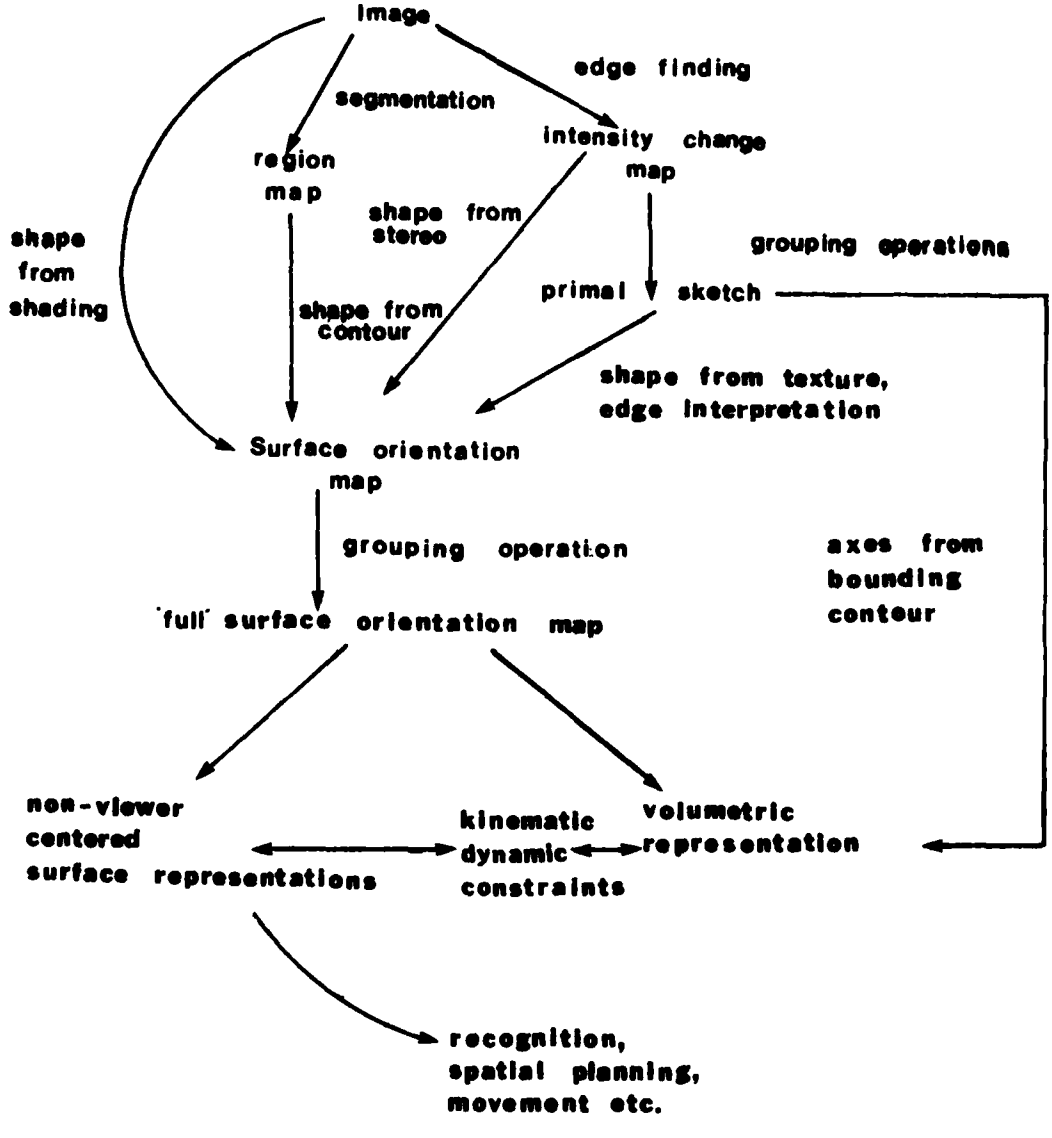
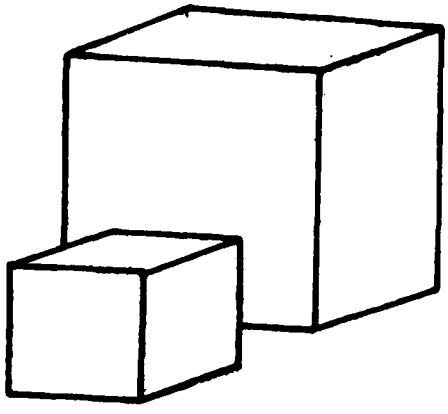
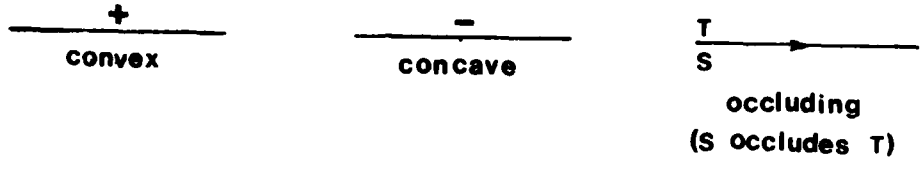


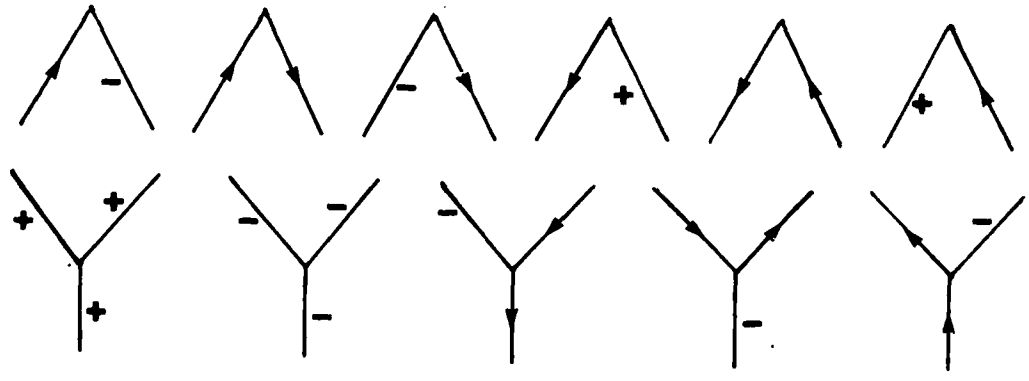
image
edge finding
segmentation
intensity change map
region map
shape from stereo
shape from shading
grouping operations
primal sketch
shape from contour
shape from texture, edge interpretation
Surface orientation map
grouping operation
axes from bounding contour
'full' surface orientation map
non-viewer centered surface representations
kinematic dynamic constraints
volumetric representation
recognition, spatial planning, movement etc.

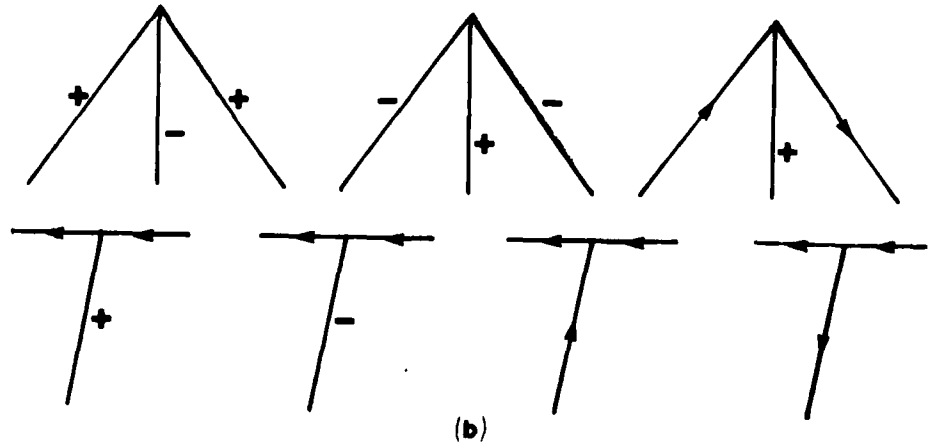
Figure 2





(a)





(b)

figure 4

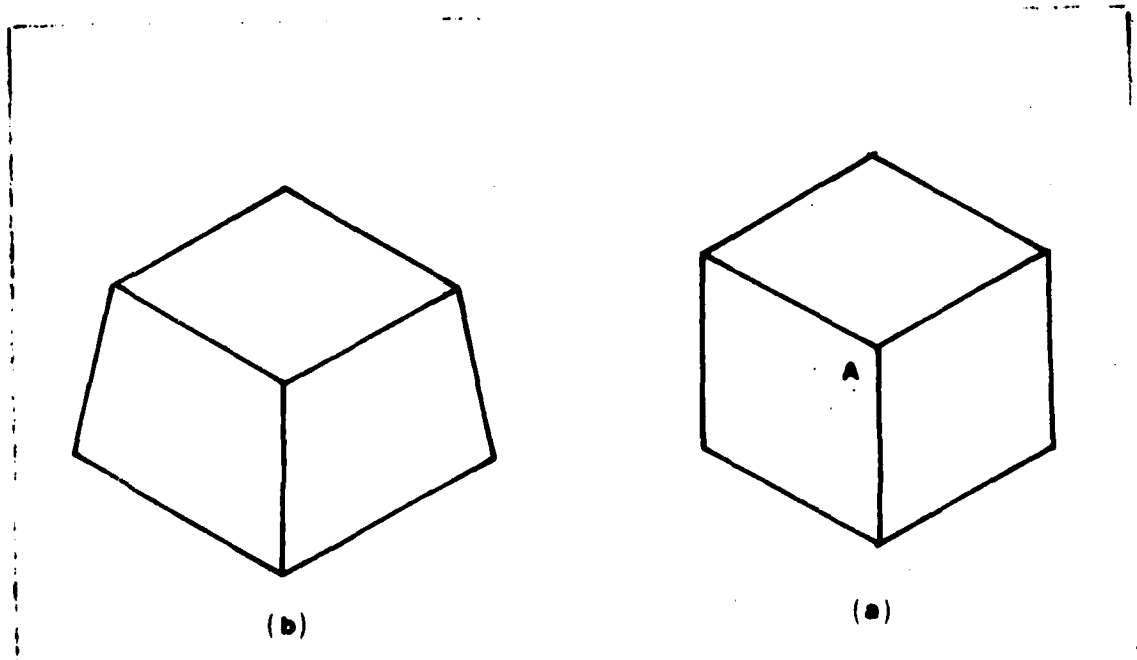


fig 5

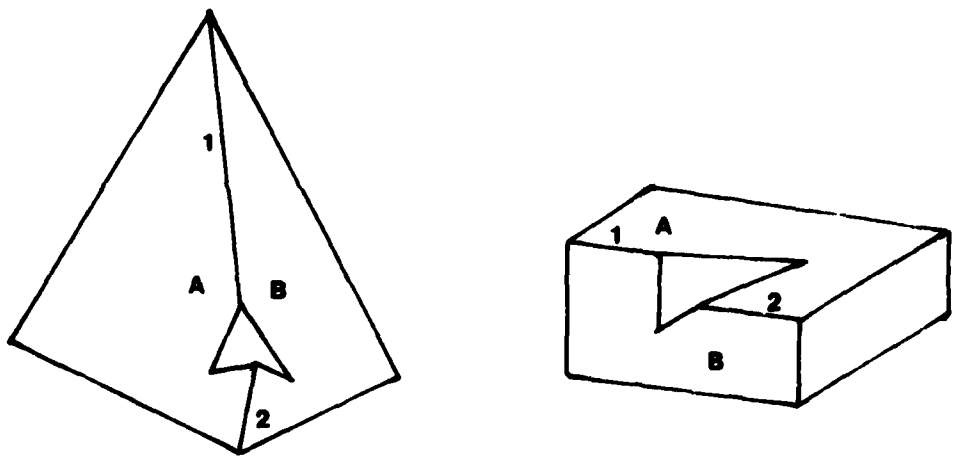
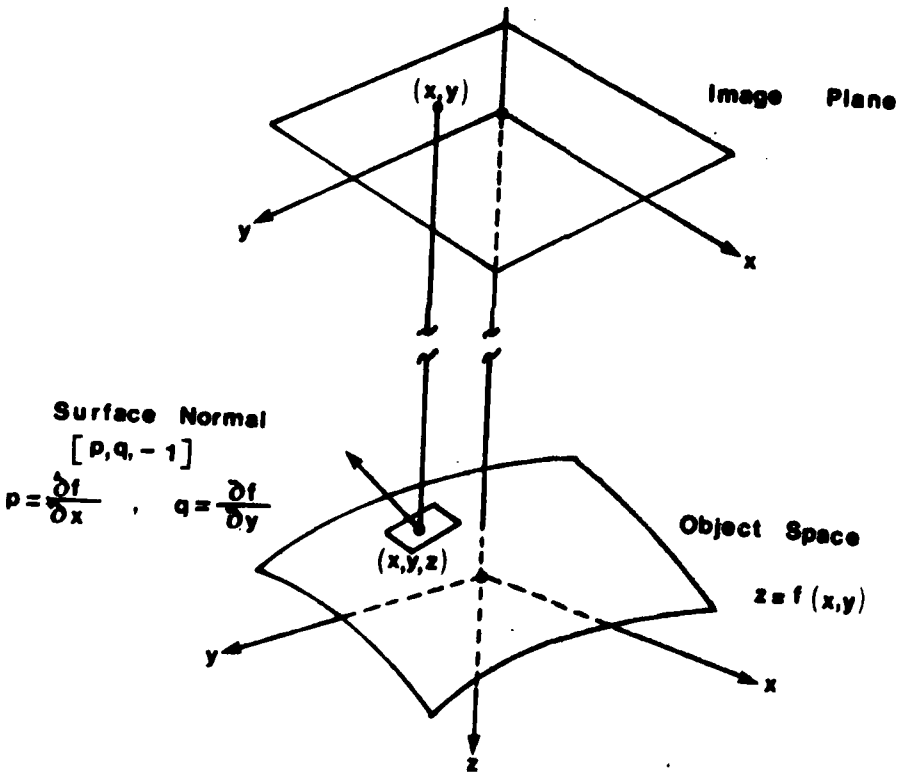


fig 6


(x,y)
Image Plane
y
x
Surface Normal
[p,q,-1]
$p = \frac{\partial f}{\partial x}$, $q = \frac{\partial f}{\partial y}$
(x,y,z)
Object Space
z = f (x,y)
y
x
z

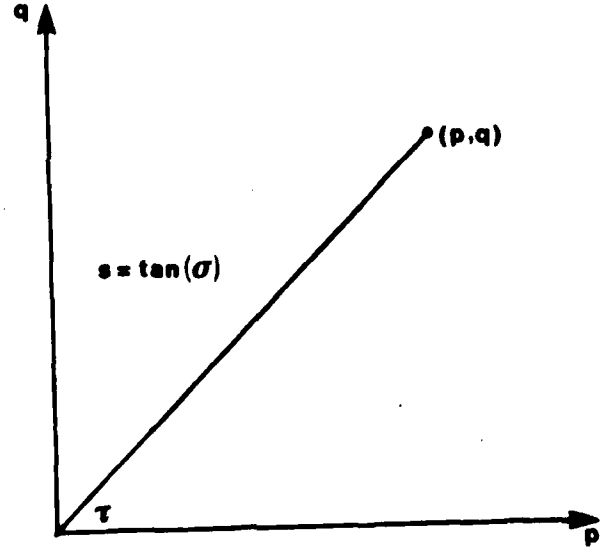

q
(p,q)
s = tan(σ)
τ
p

fig 8

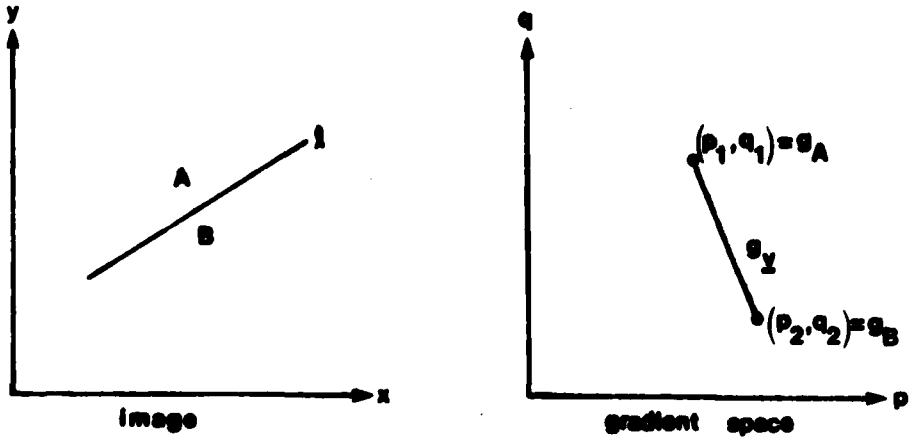
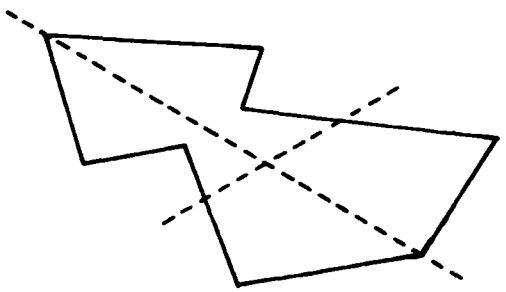
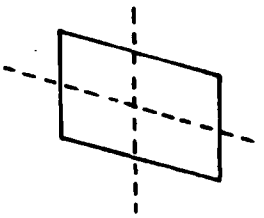


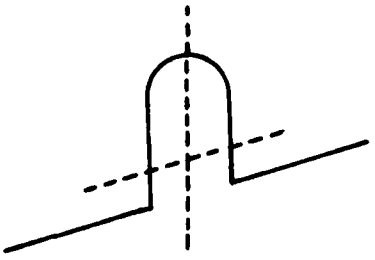
Fig 9



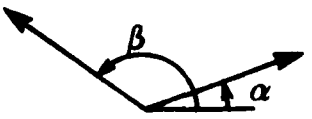
(a)



(b)



(c)



(d)

Fig. 6

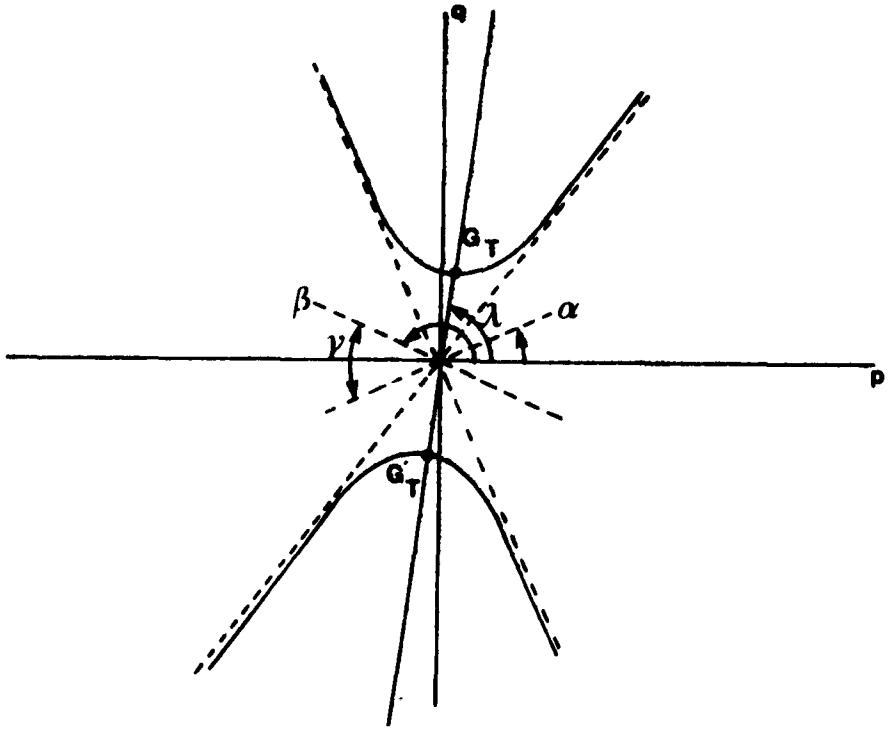


Fig. 11

-1	1

first difference

-1	0
0	1

Roberts $(\partial I / \partial (x+y))$

-1	0	1
-1	0	1
-1	0	1

Prewitt, O'Gorman-Clowes

-1	0	1
-2	0	2
-1	0	1

Sobel

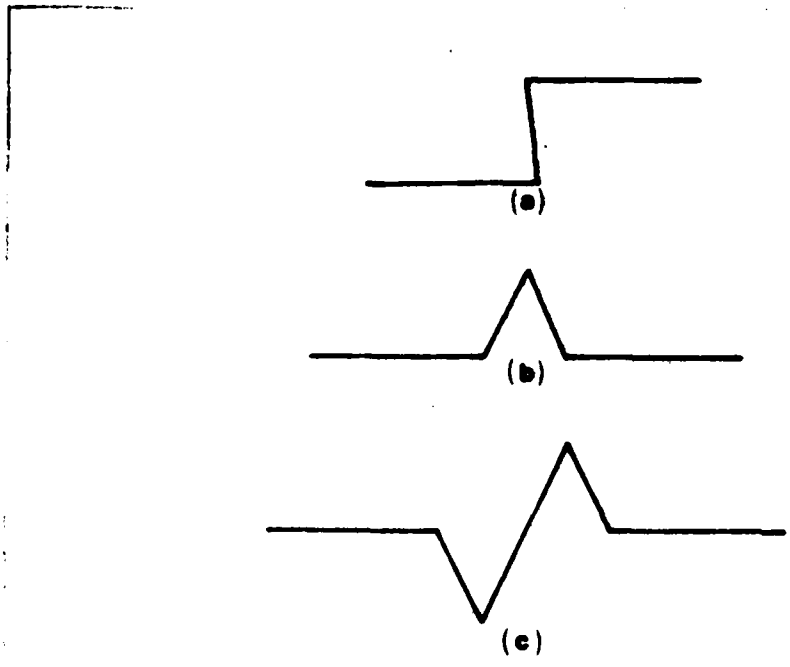
1	-2	1

second difference
bar mask

-1	2	-1
-1	2	-1
-1	2	-1
-1	2	-1

bar mask

fig 12



-100	-100	0	100	100
-100	-100	0	100	100
-100	-100	0	100	100
-100	-100	0	100	100
-100	-100	0	100	100

a) 0^{o}

-100	32	100	100	100
-100	-78	92	100	100
-100	-100	0	100	100
-100	-100	-92	78	100
-100	-100	-100	-32	100

b) 30^{o}

100	100	100	100	100
-32	78	100	100	100
-100	-92	0	92	100
-100	-100	-100	-78	32
-100	-100	-100	-100	-100

c) 60^{o}

100	100	100	100	100
100	100	100	100	100
0	0	0	0	0
-100	-100	-100	-100	-100
-100	-100	-100	-100	-100

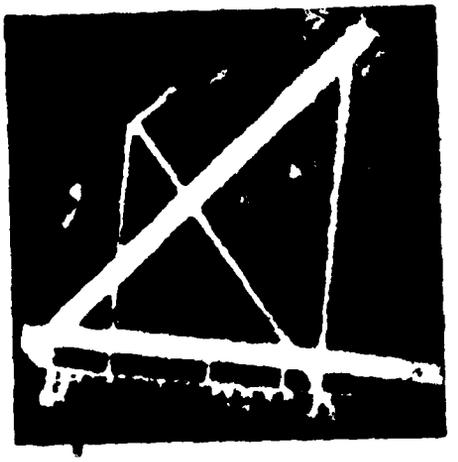
d) 90^{o}

100	100	100	100	100
100	100	100	78	-32
100	92	0	-92	-100
32	-78	-100	-100	-100
-100	-100	-100	-100	-100

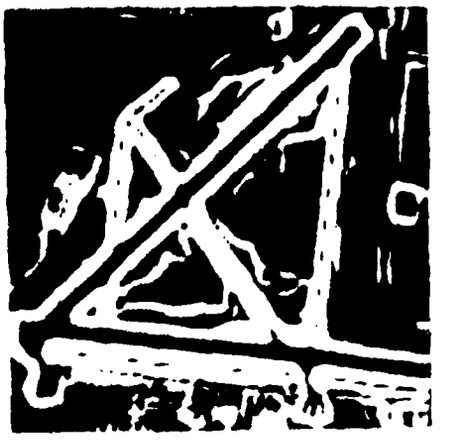
e) 120^{o}

100	100	100	32	-100
100	100	92	-78	-100
100	100	0	-100	-100
100	78	-92	-100	-100
100	32	-100	-100	-100

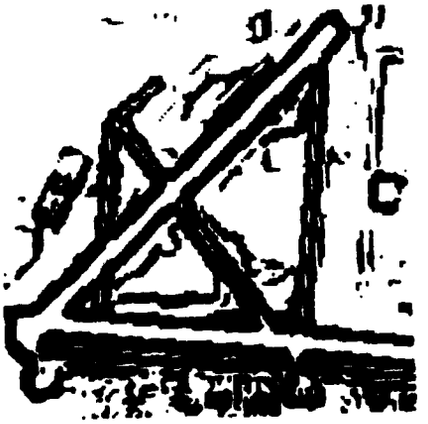
f) 150^{o}



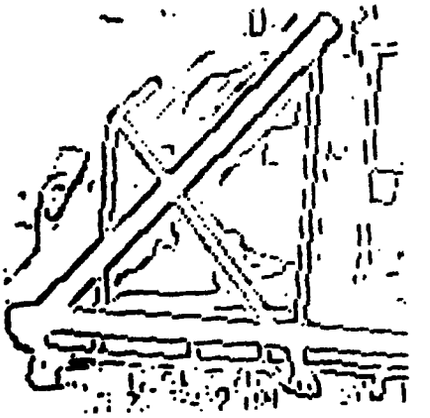
(a) Digital Image



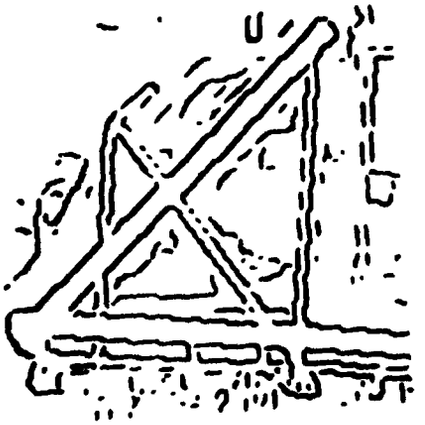
(b) Edge magnitude



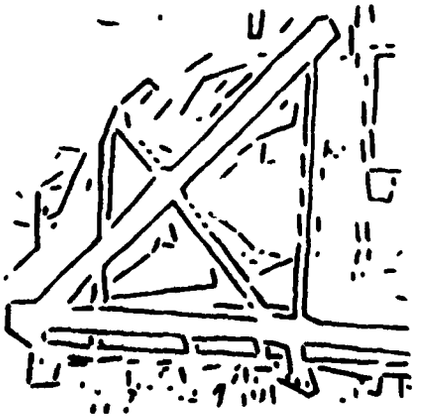
(c) Thresholded edges (not thinned)



(d) Thinned and Thresholded output)

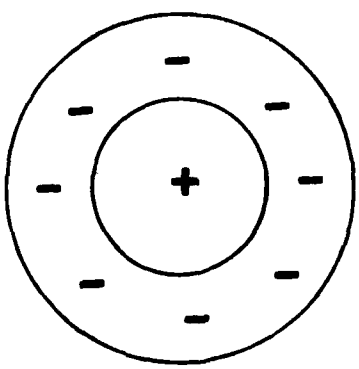


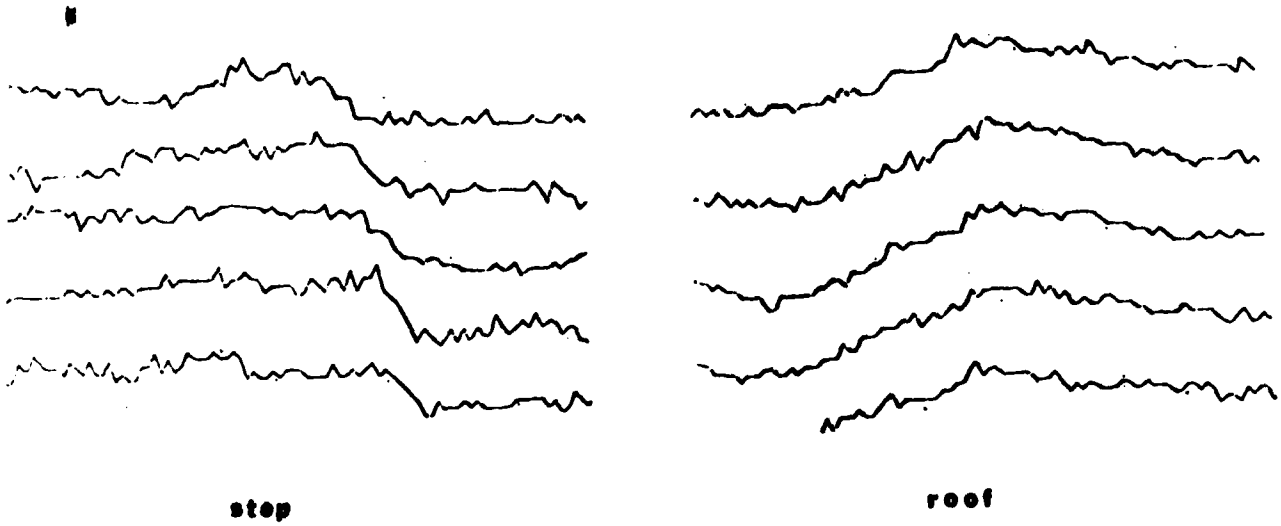
(e) Linked segments



(f) Linear segment approximation

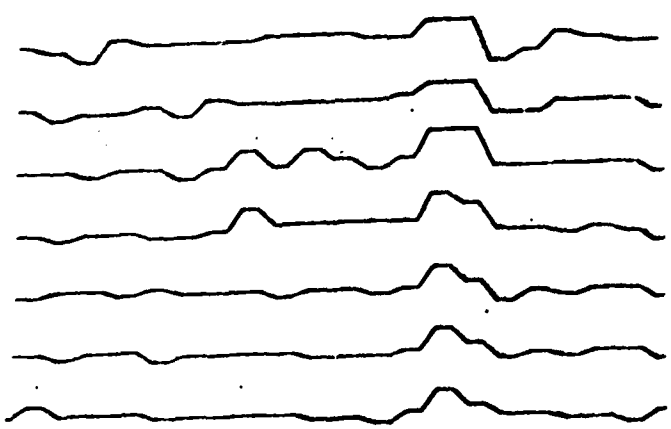
An airport image and various stages of processing





step

roof



edge

fig 1b

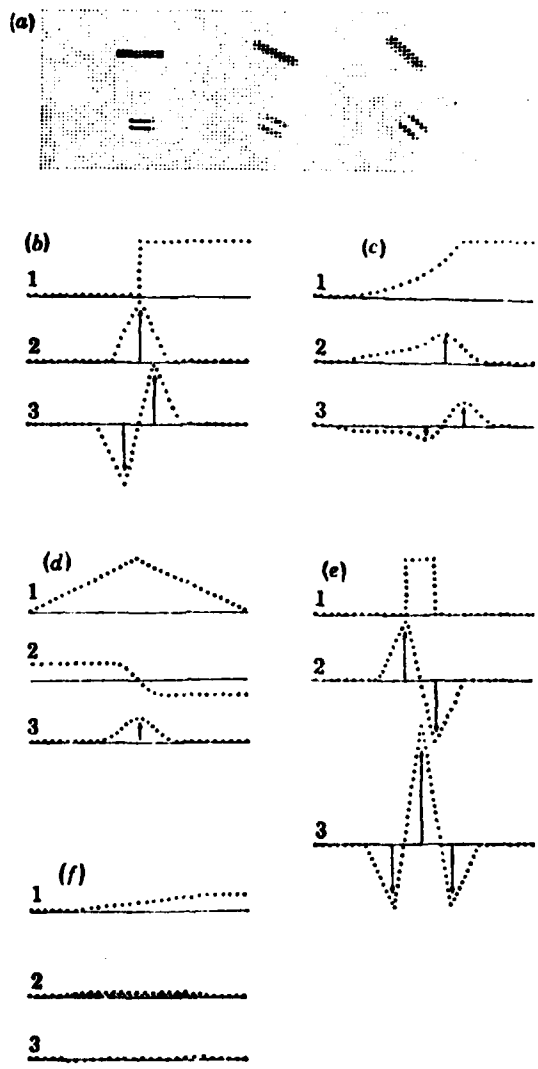
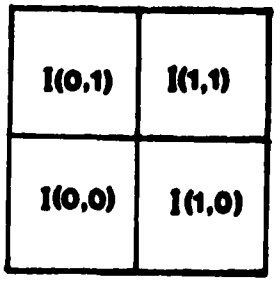
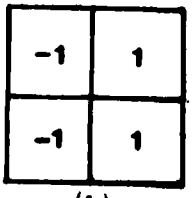

(a)
(b)
1
2
3
(c)
1
2
3
(d)
1
2
3
(e)
1
2
3
(f)
1
2
3

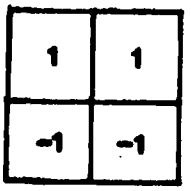
Fig 1



(a)

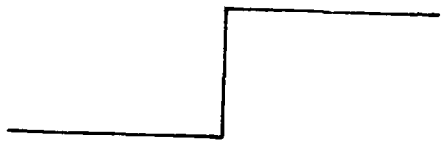


(b)

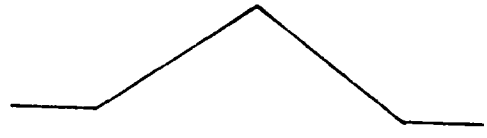


(c)

fig 19

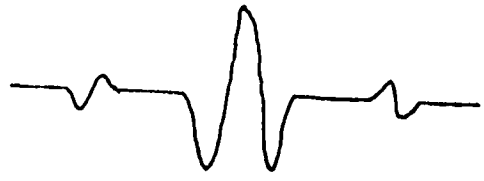


1.

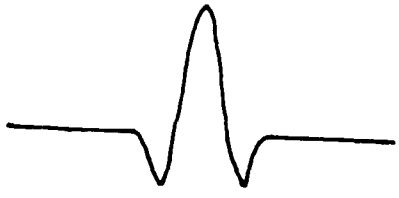


1.

2.



2.



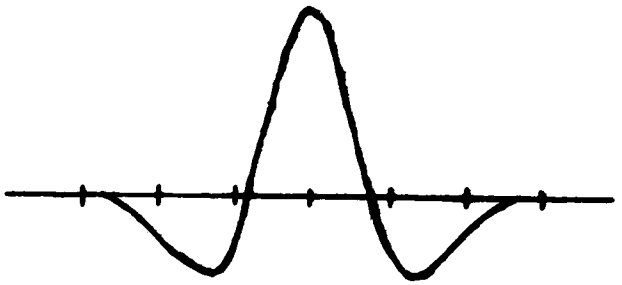
3.

(a)

3.

(b)

fig 20



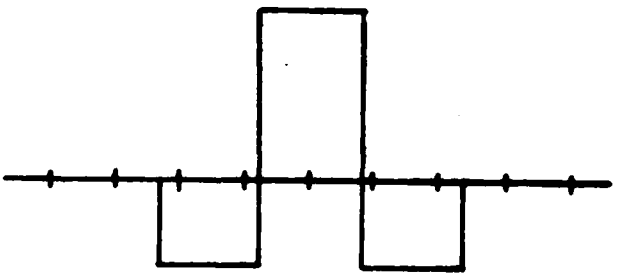
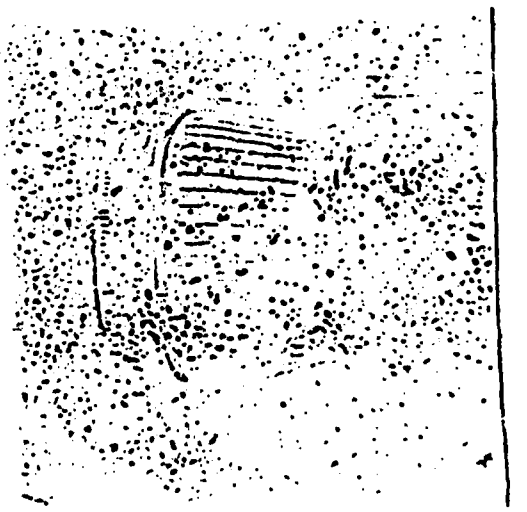
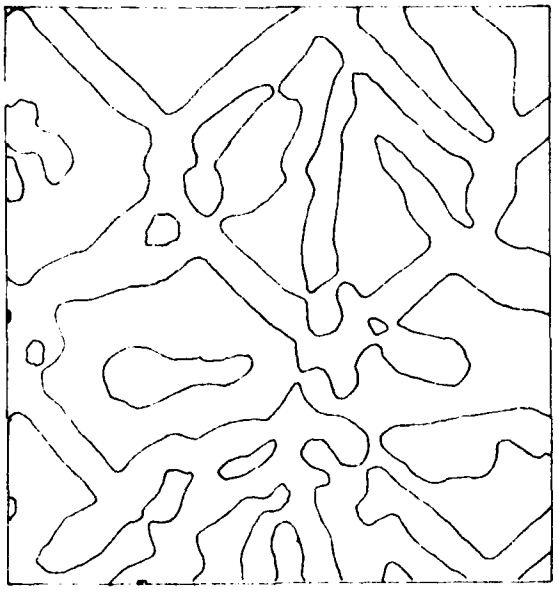
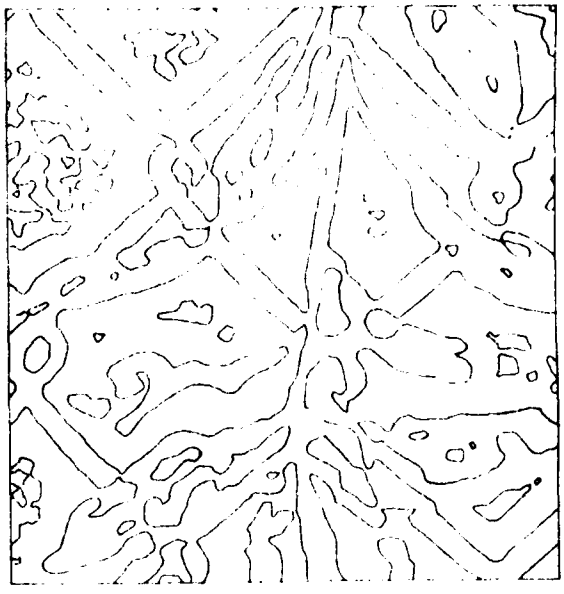


fig 22

Fig 22b, 23b







(a)

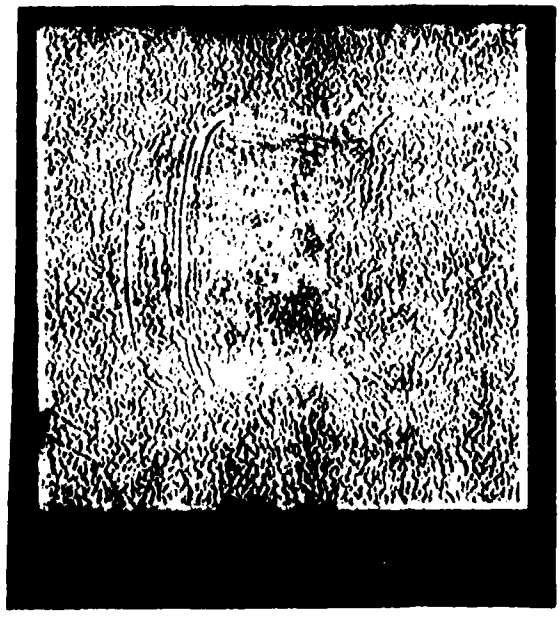
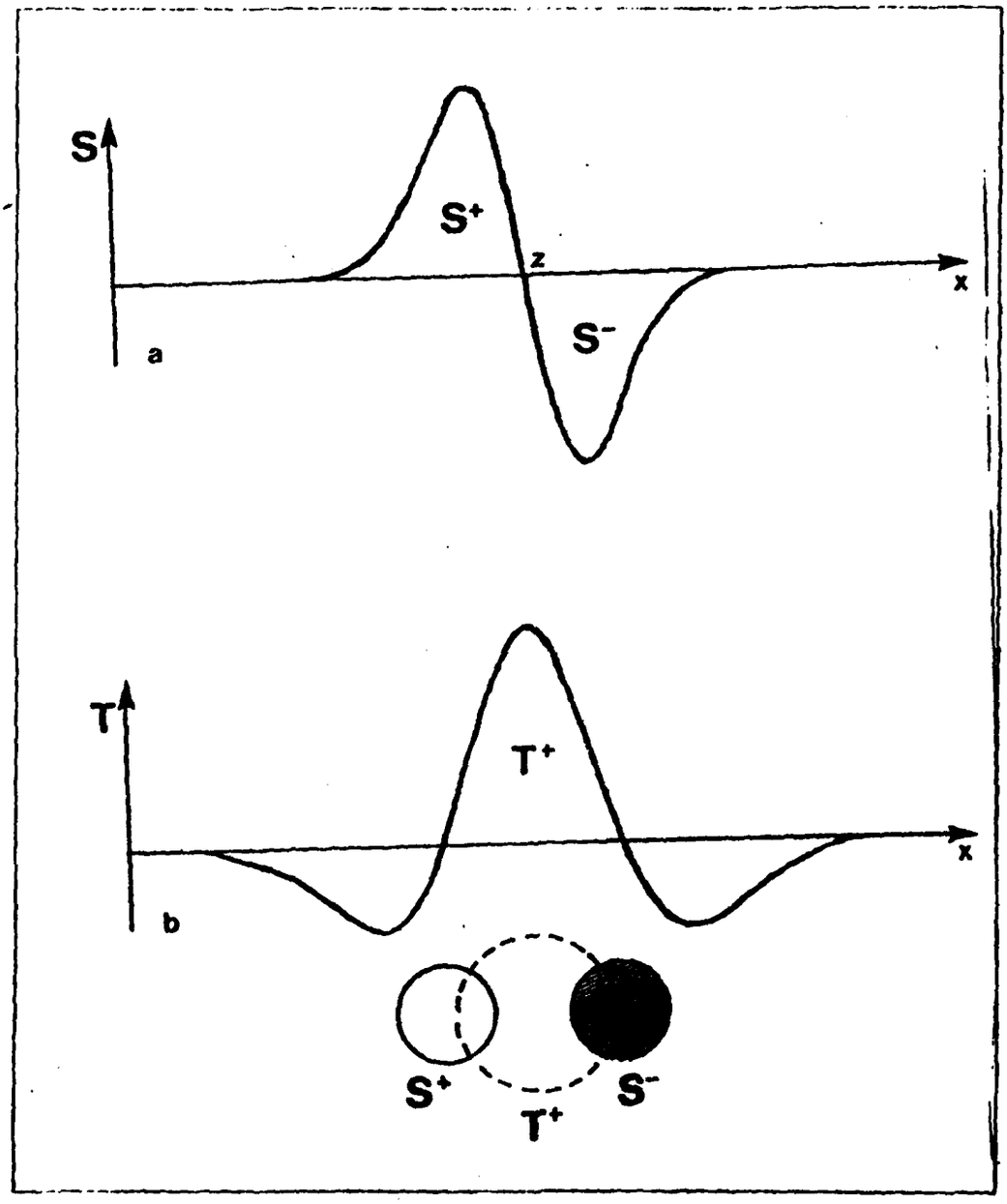
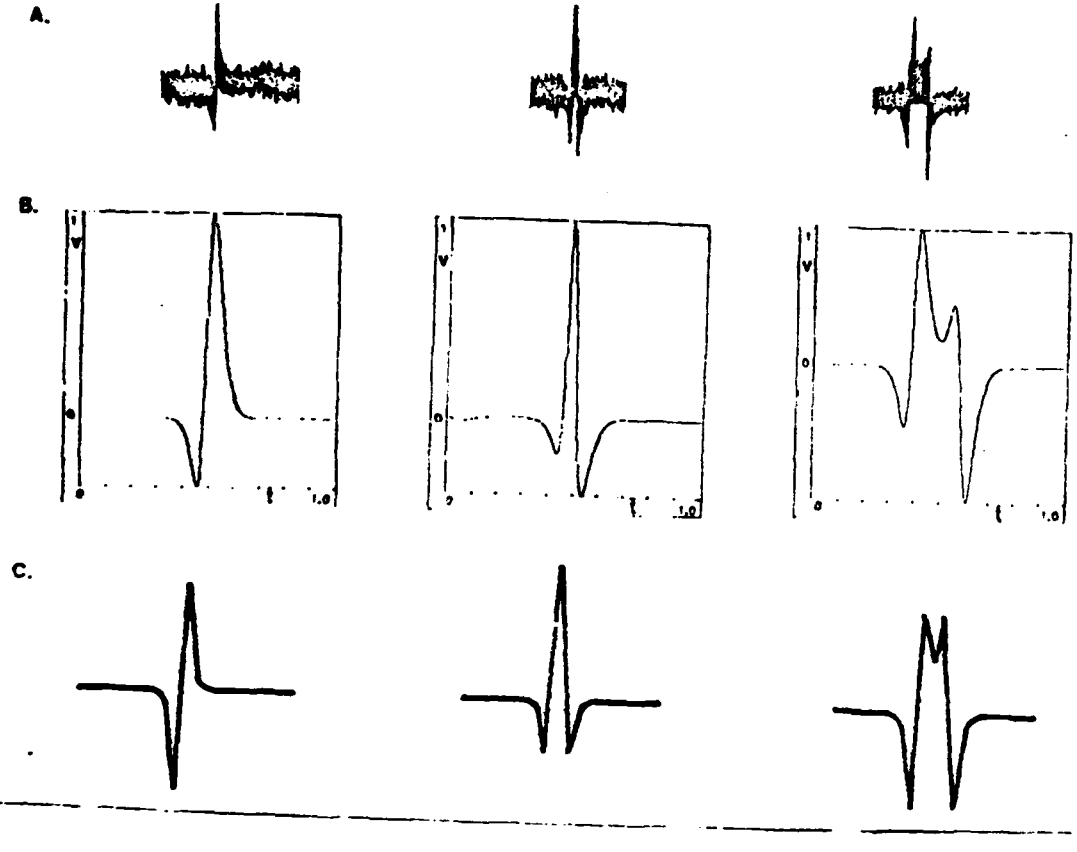
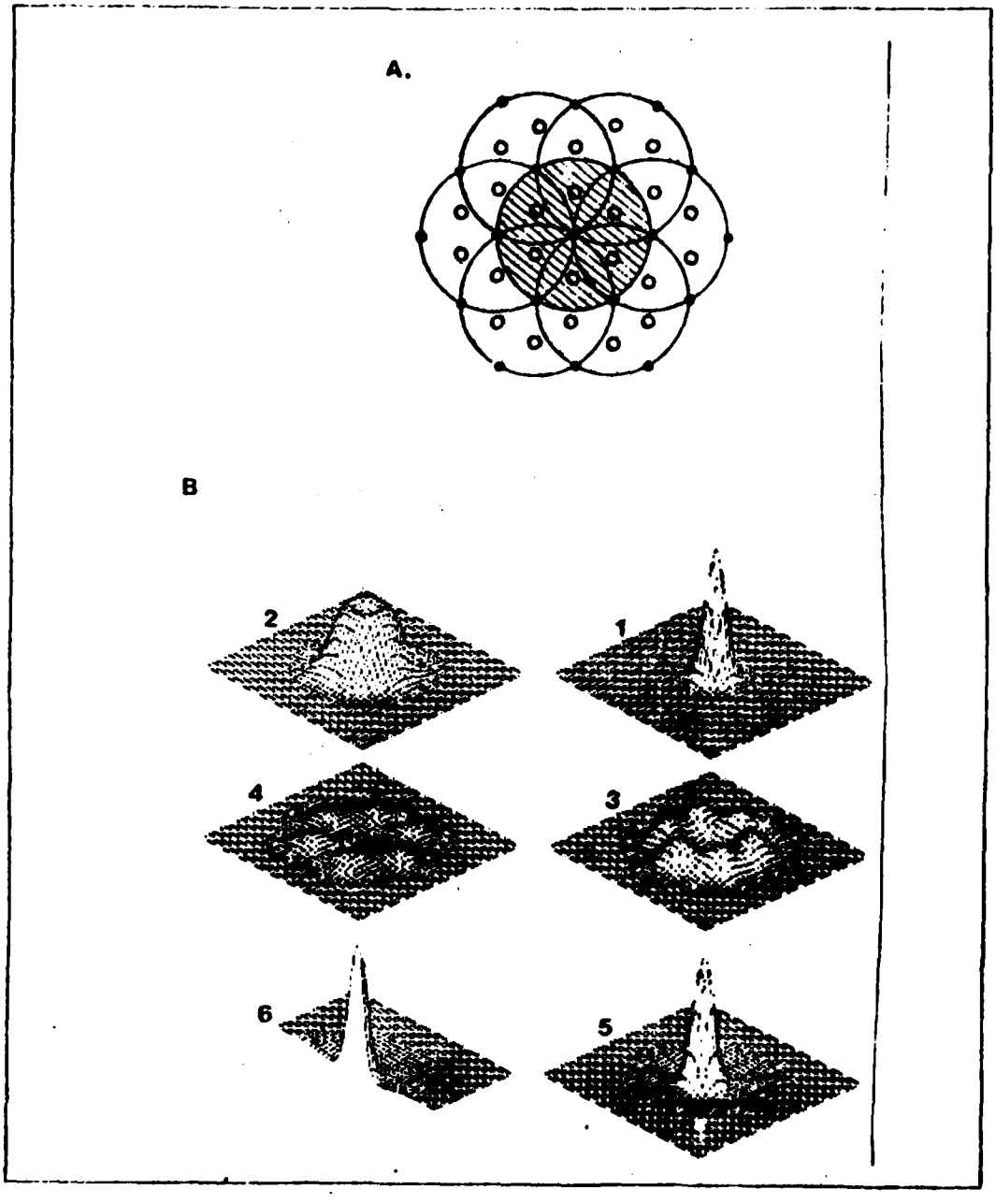
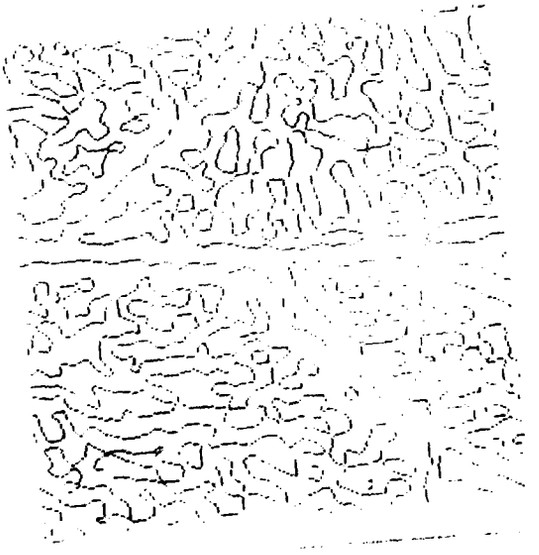


Fig 24


S
S+
z
x
S−
a
T
T+
x
b
S+
S−
T+

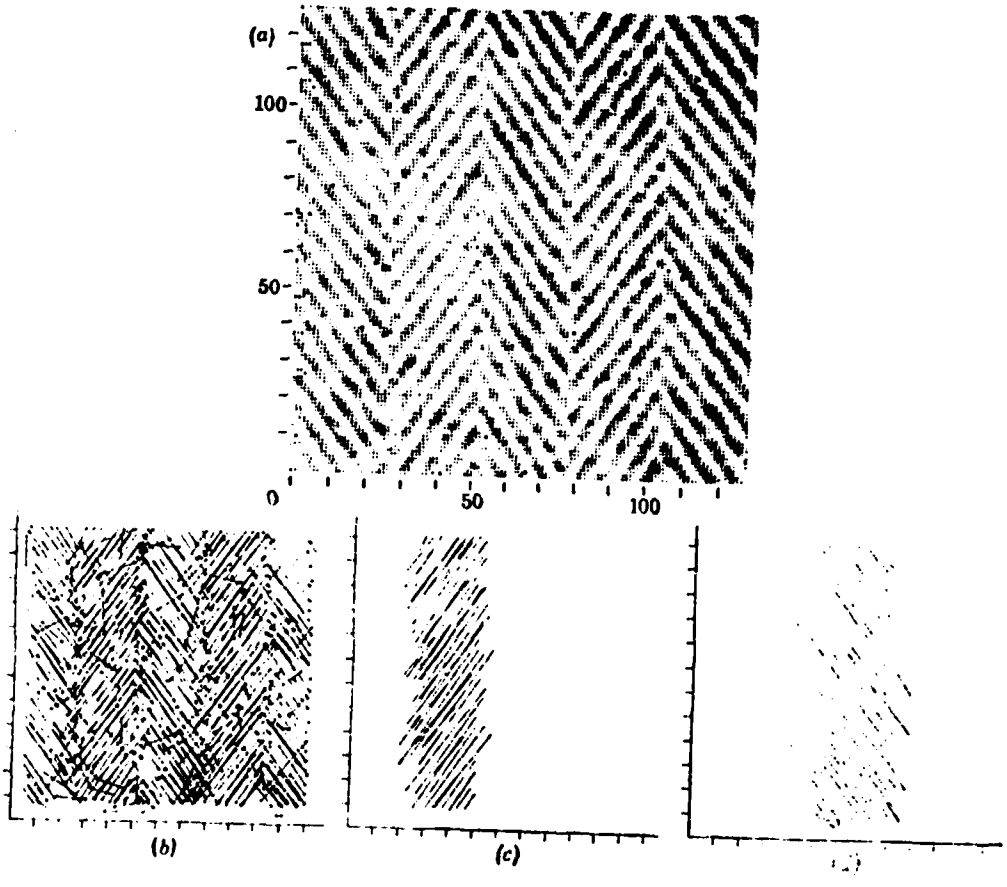

A.
B.
C.

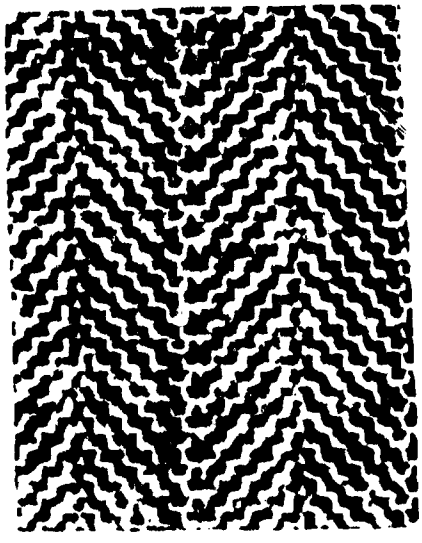

A.
B
2
1
4
3
6
5

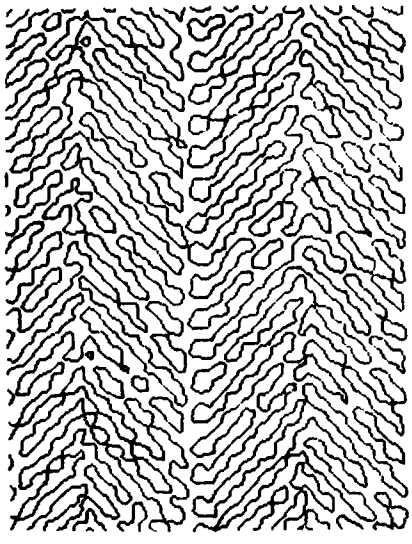


C-

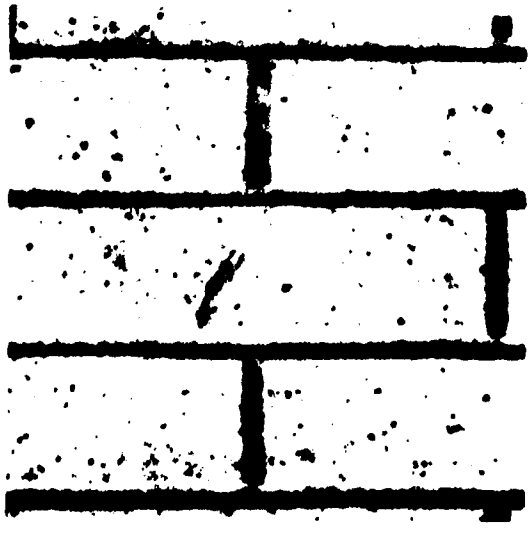
1.8 2.8

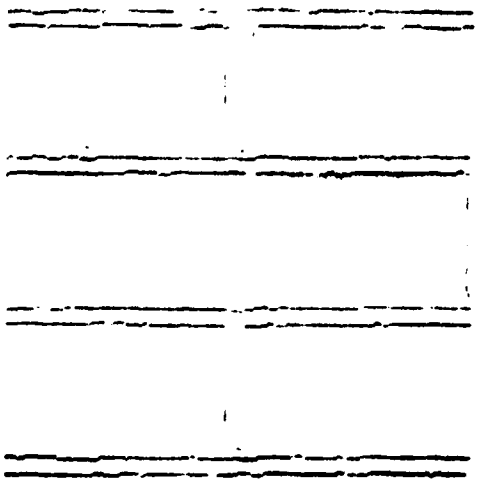

(a)
100
50
0
50
100
(b)
(c)

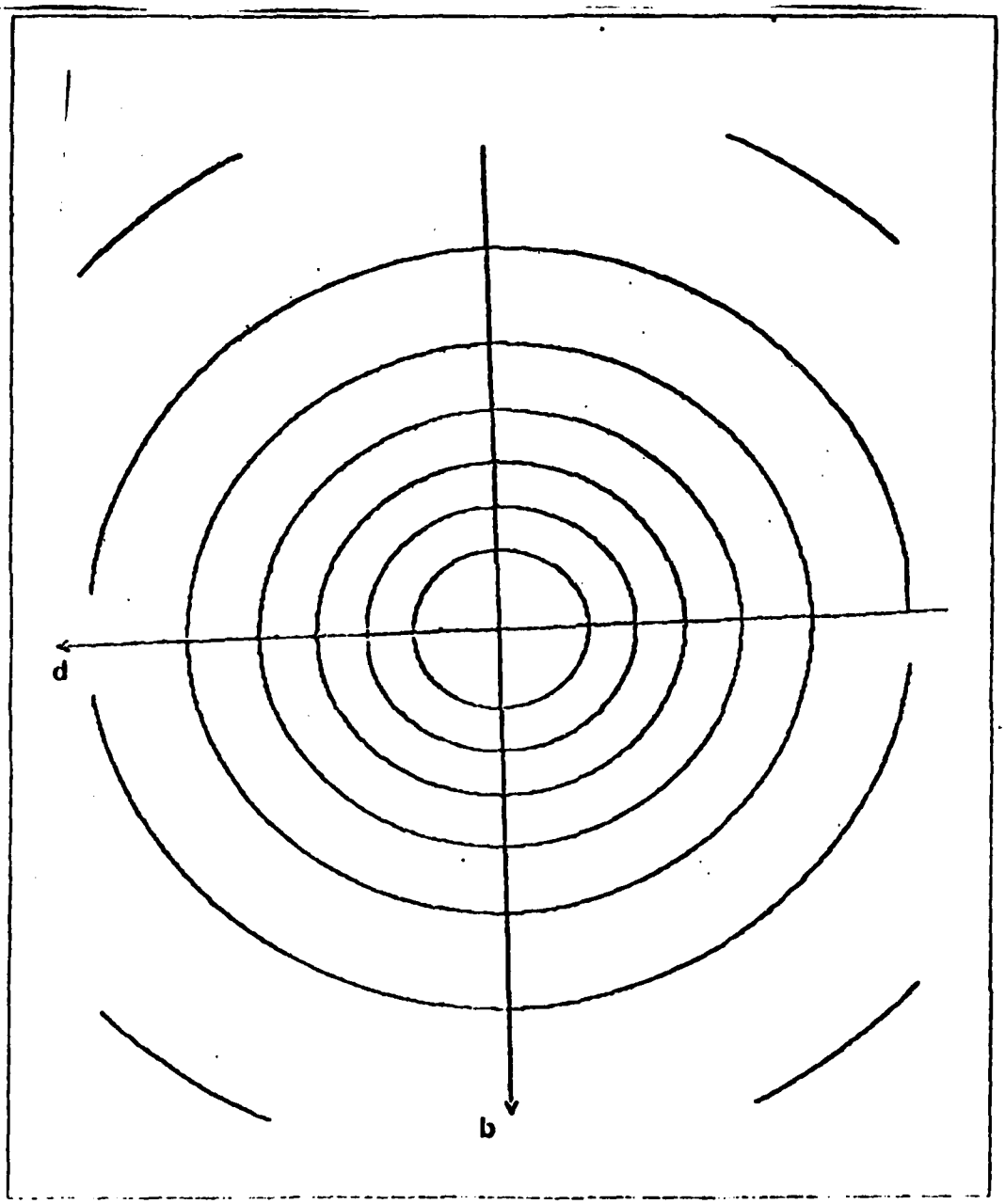


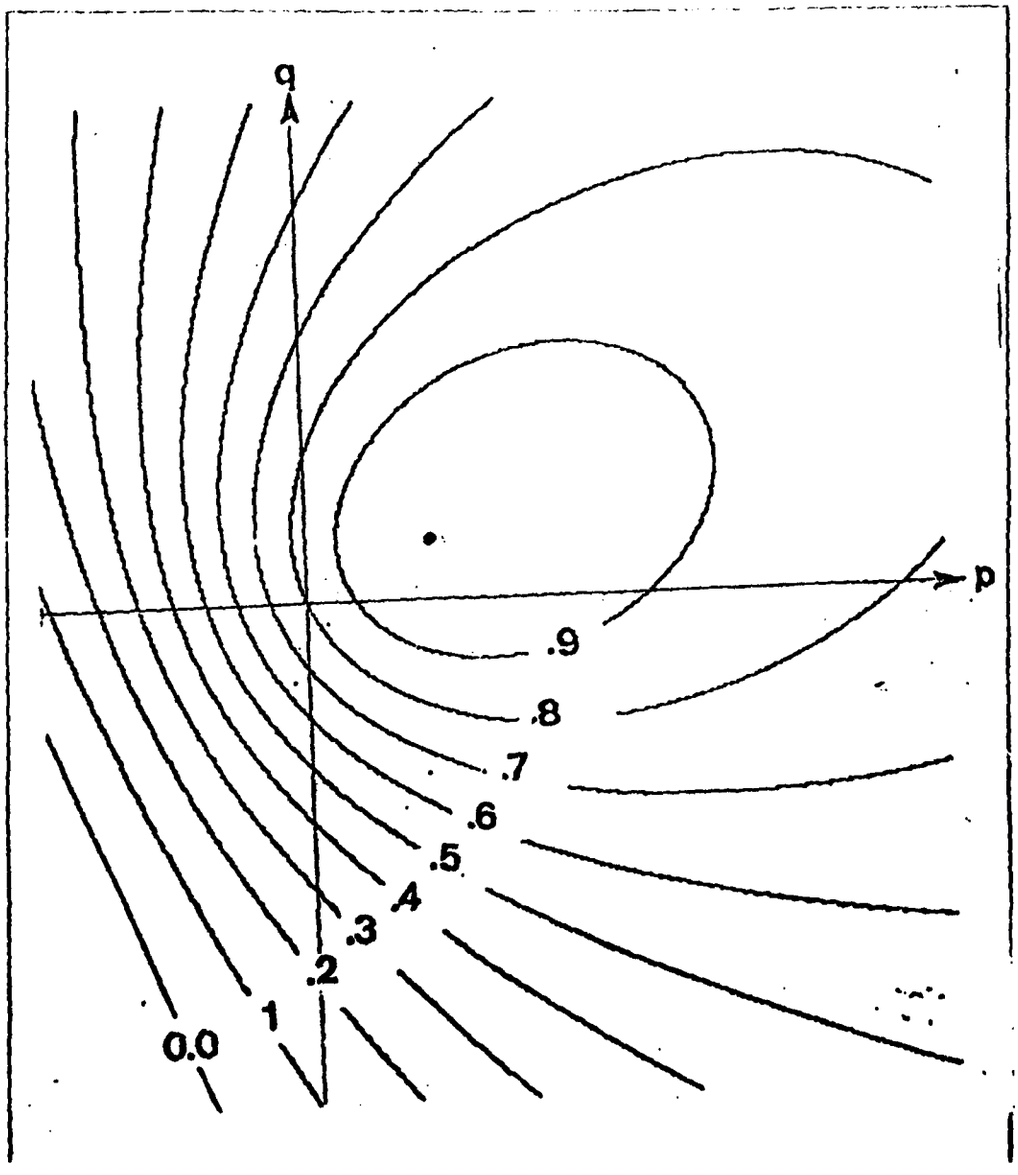


fg 30






d
b


q
p
.9
.8
.7
.6
.5
.4
.3
.2
1
0.0

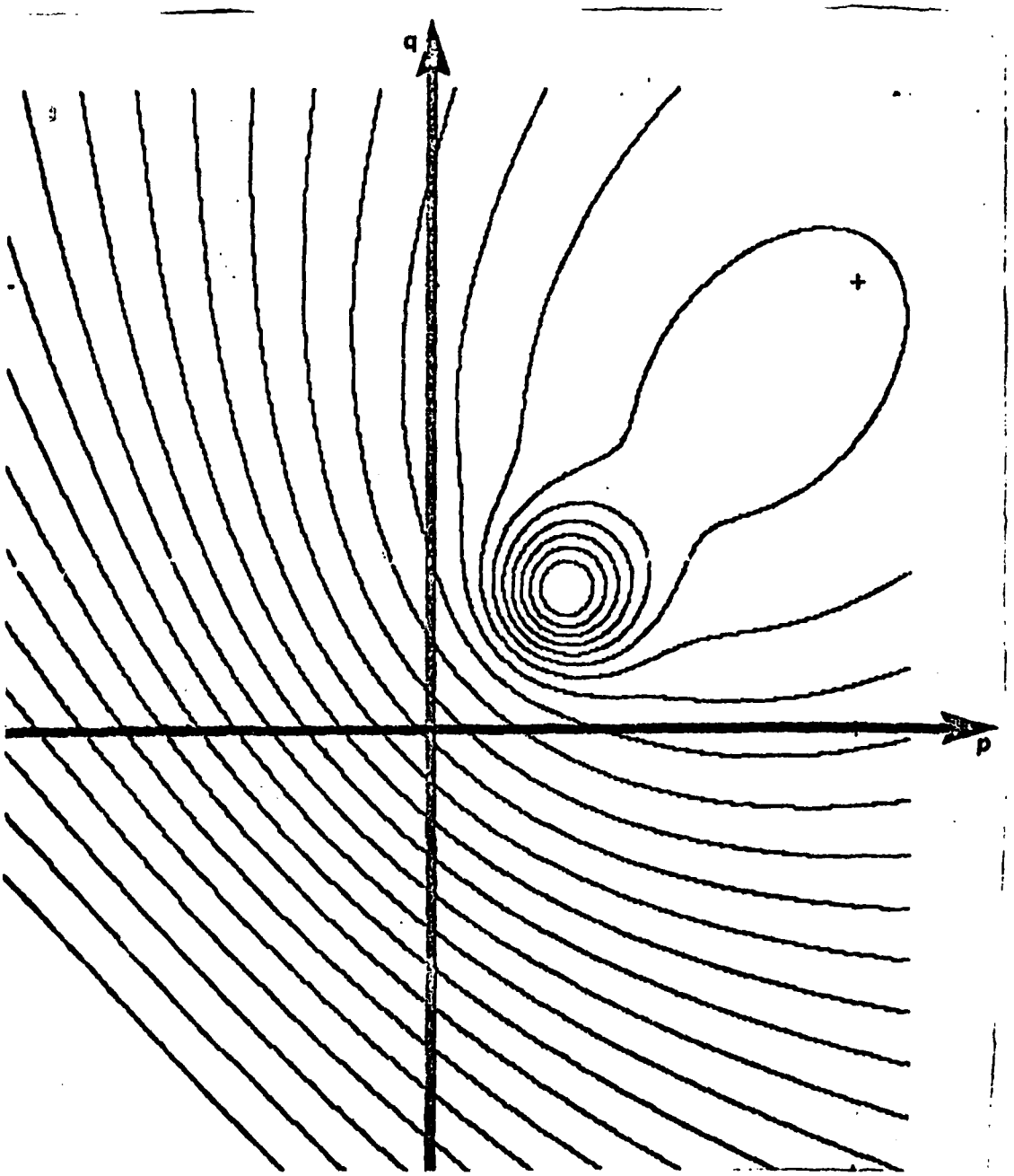
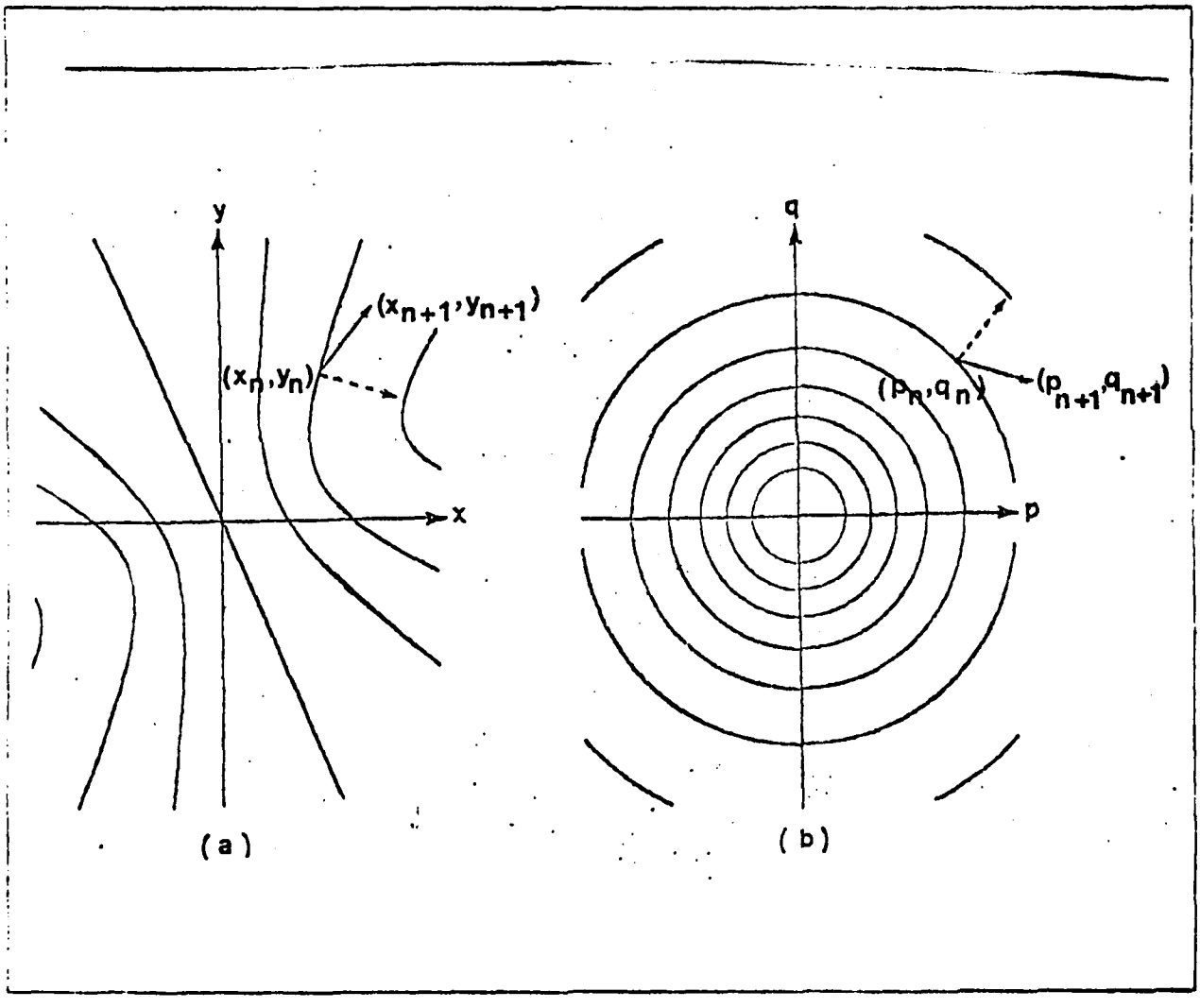

q
+
p

Fig 34



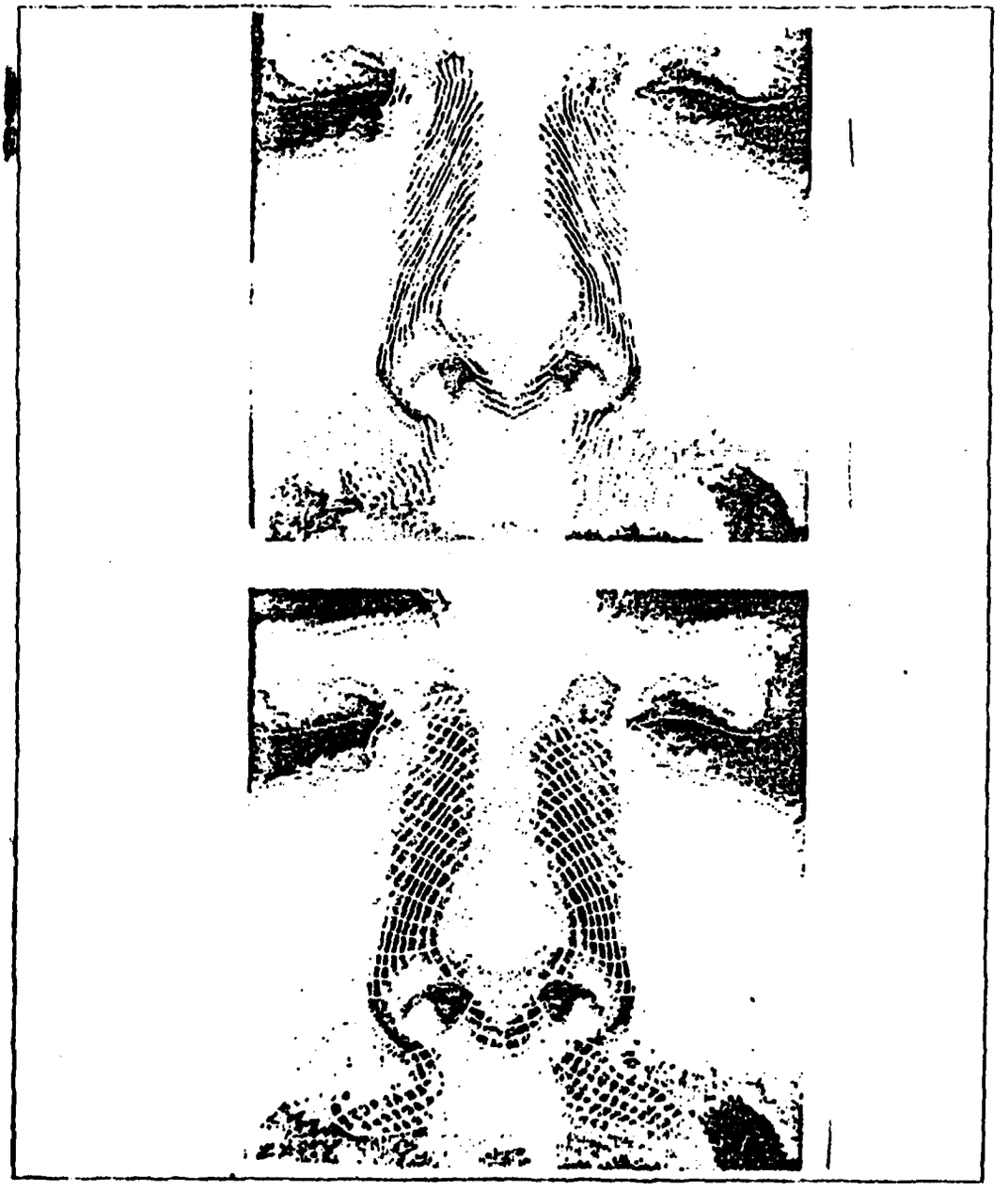
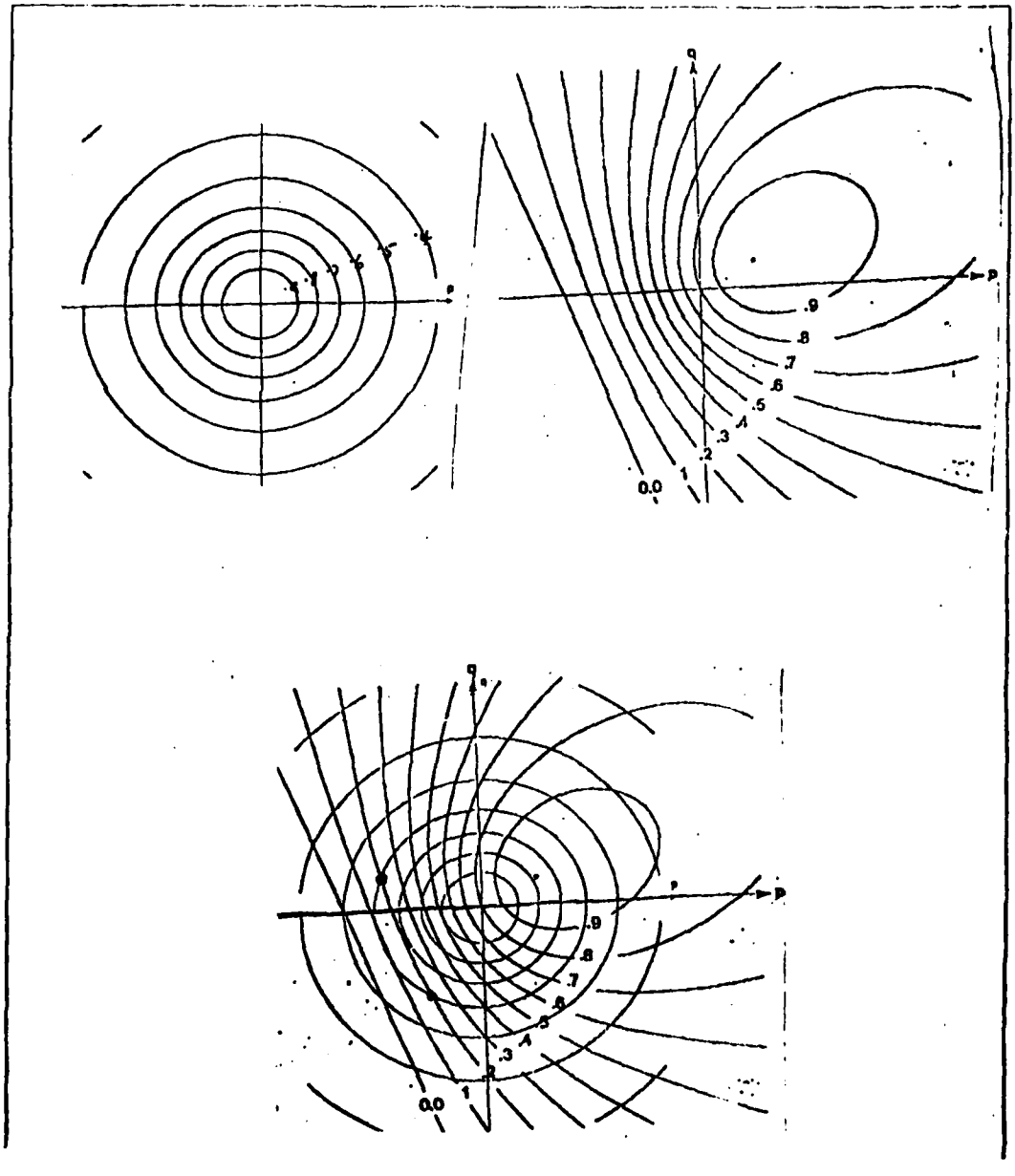
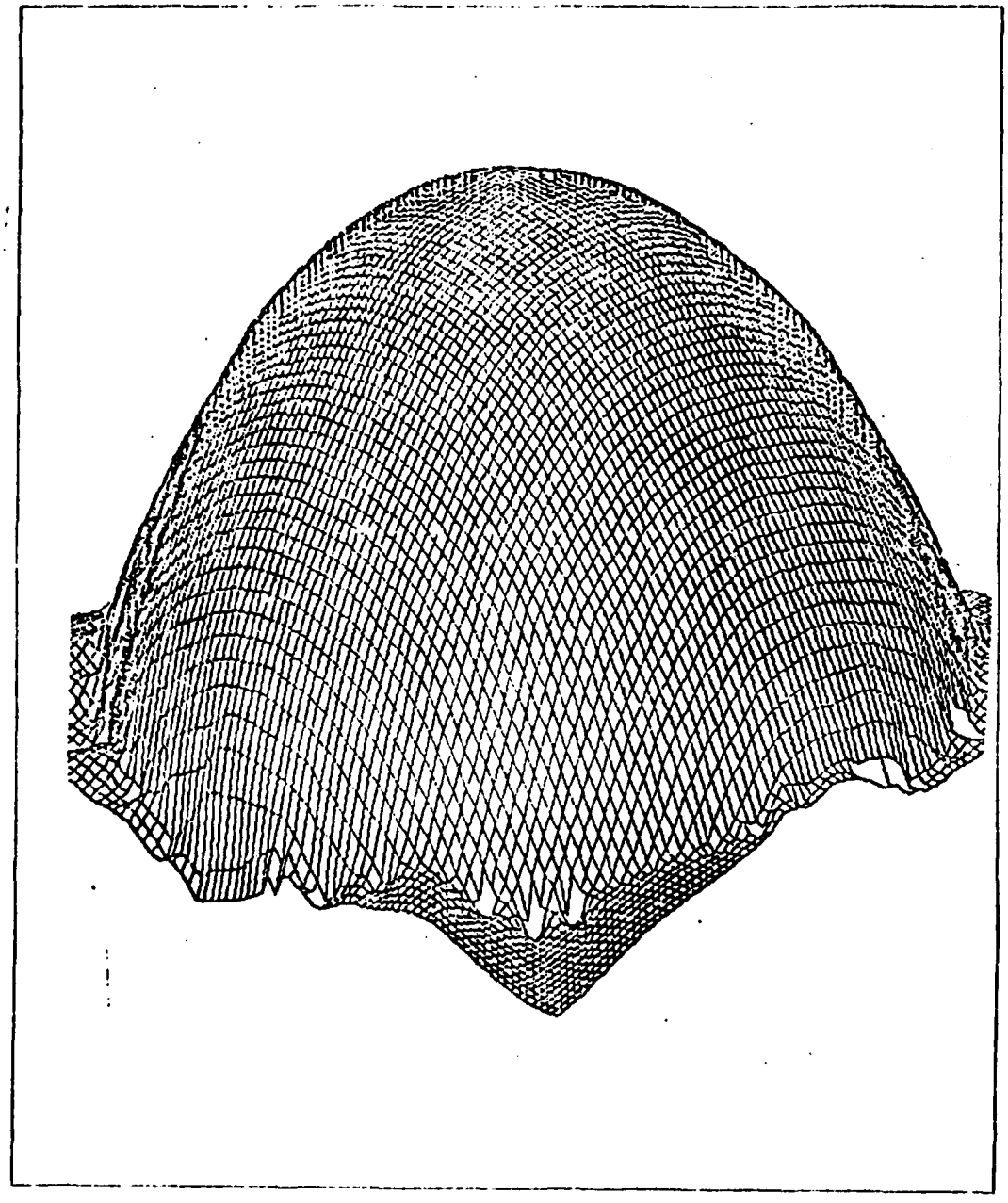
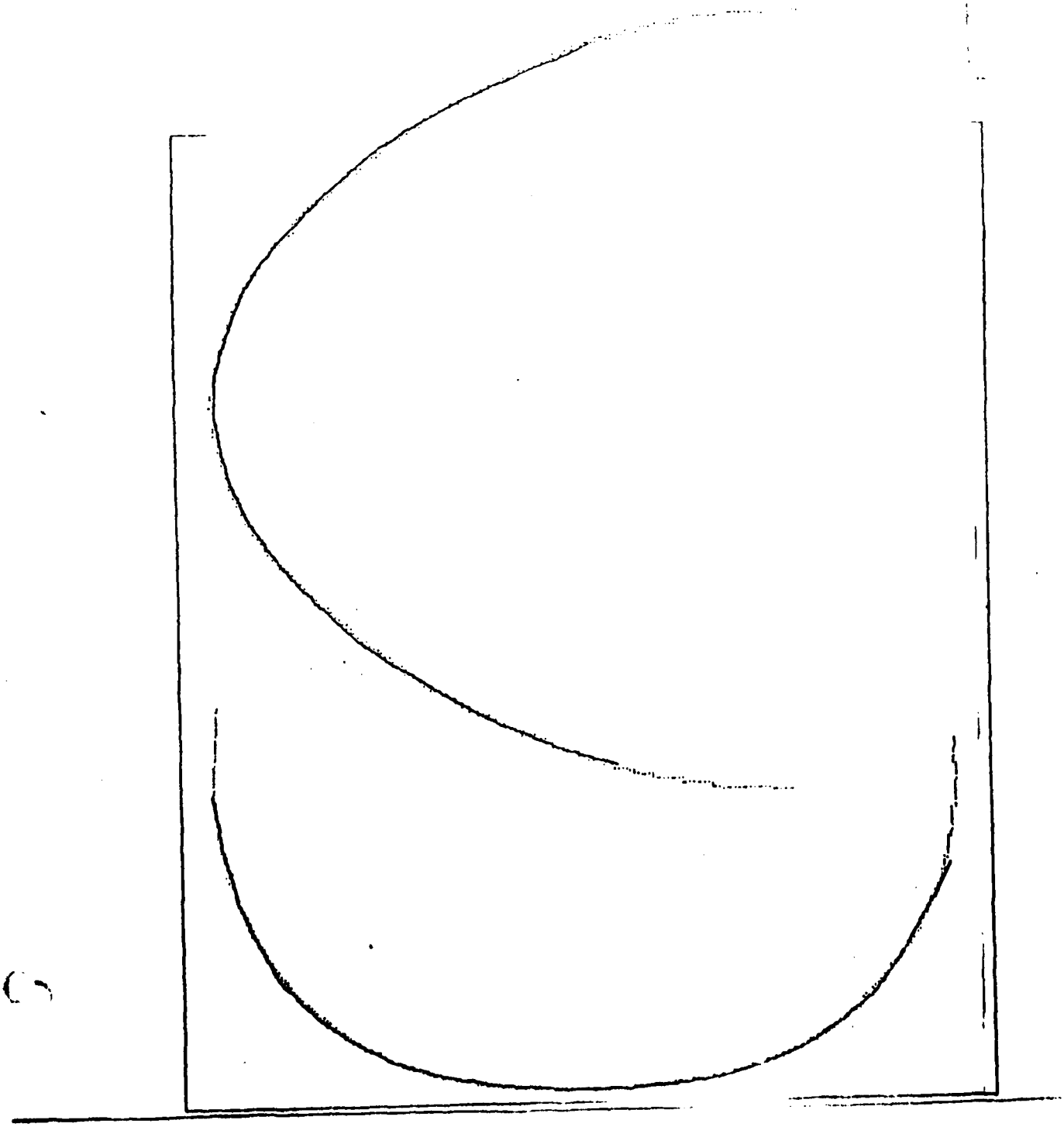
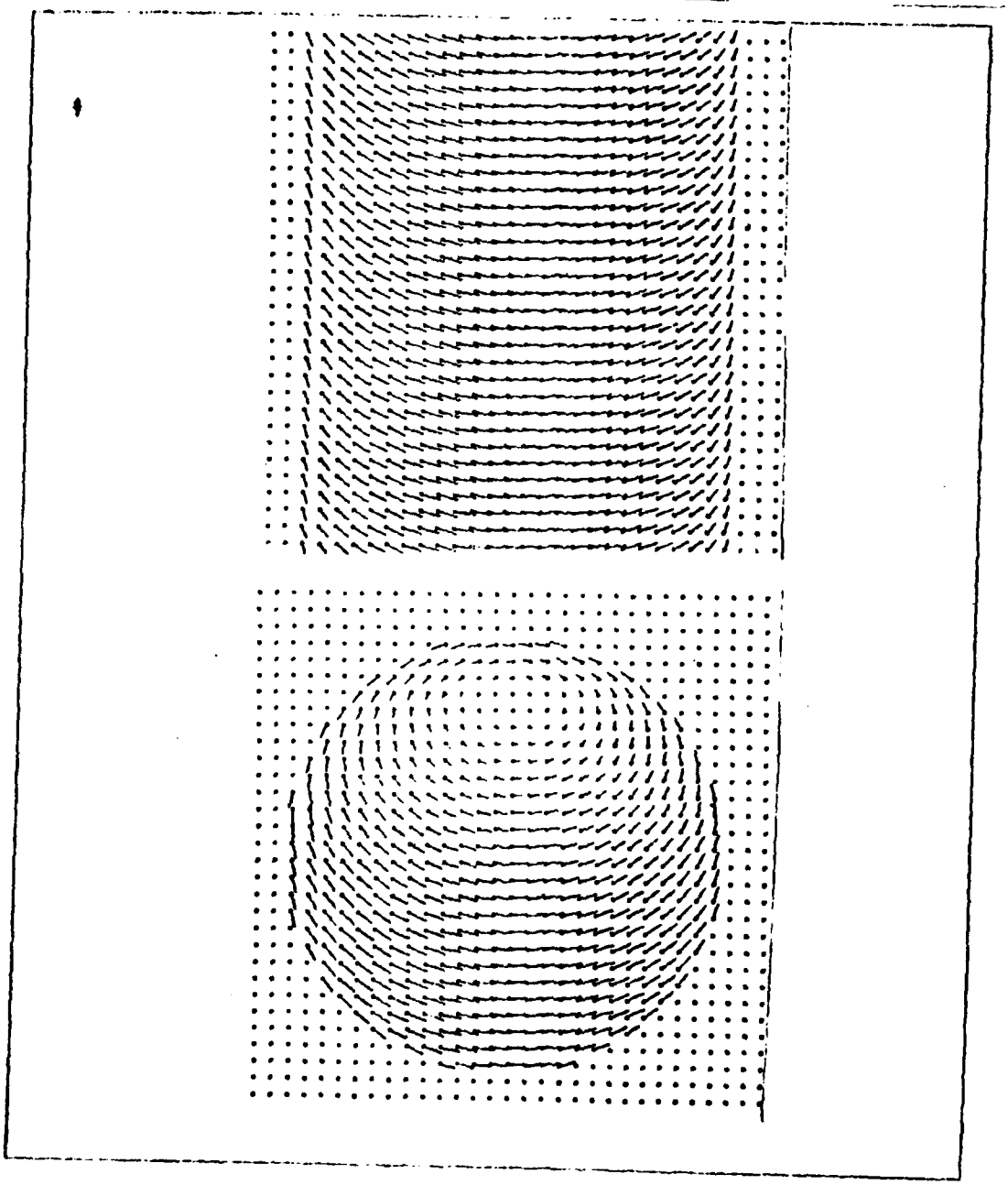


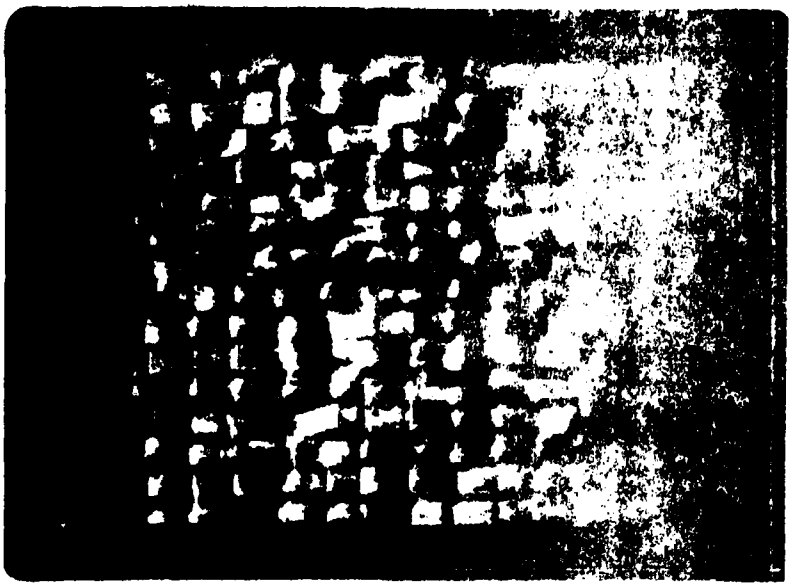
Fig 3c









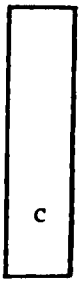


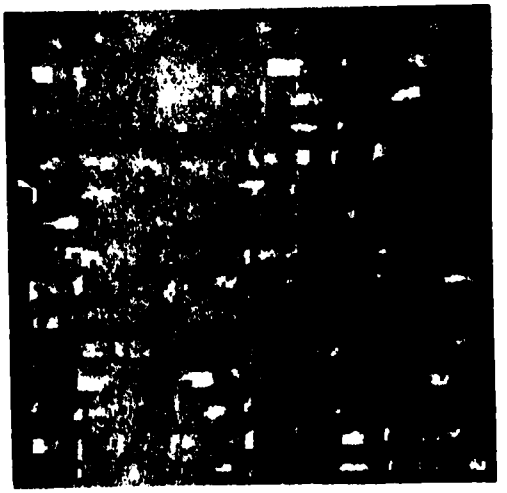
VERTICAL SCAN DIRECTION

STRONG EVIDENCE OF PERIODICITY (ELEMENT SPACING 8.00)

STRONG EVIDENCE OF PREDOMINANT ELEMENT SIZE (5.00) WITH MODERATE SUPPORT FOR ELEMENT SPACING (8.00)

RATIO OF SIZE TO PERIOD IS .63

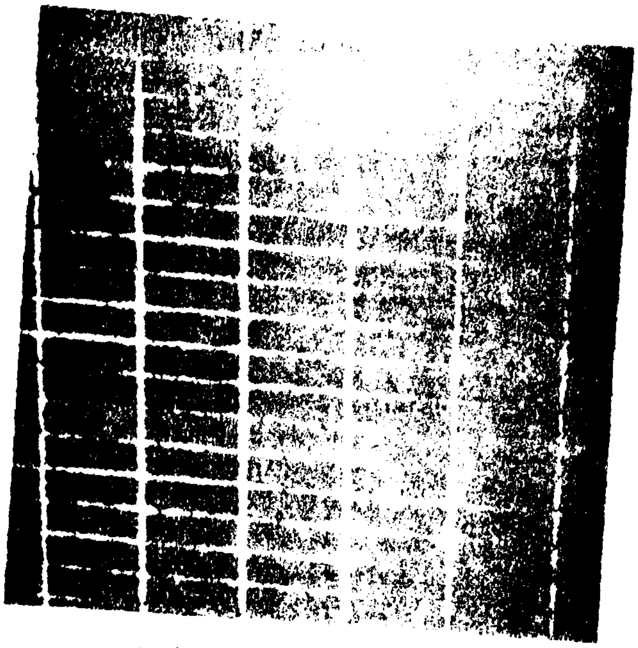




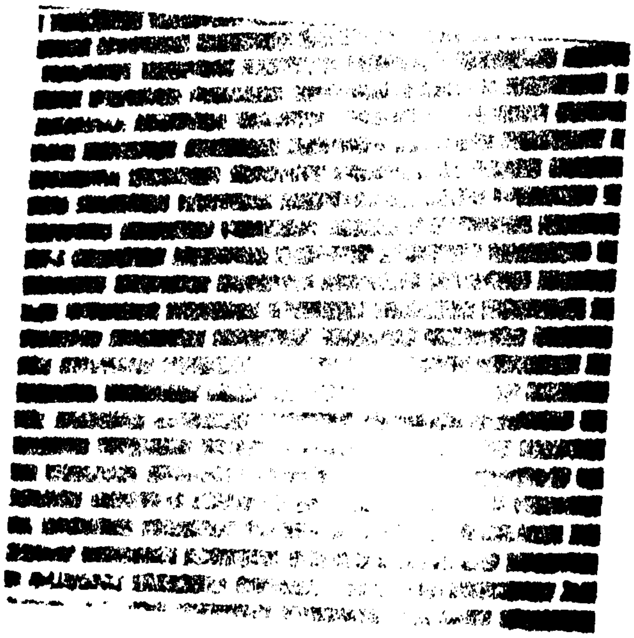
(a)

(b)

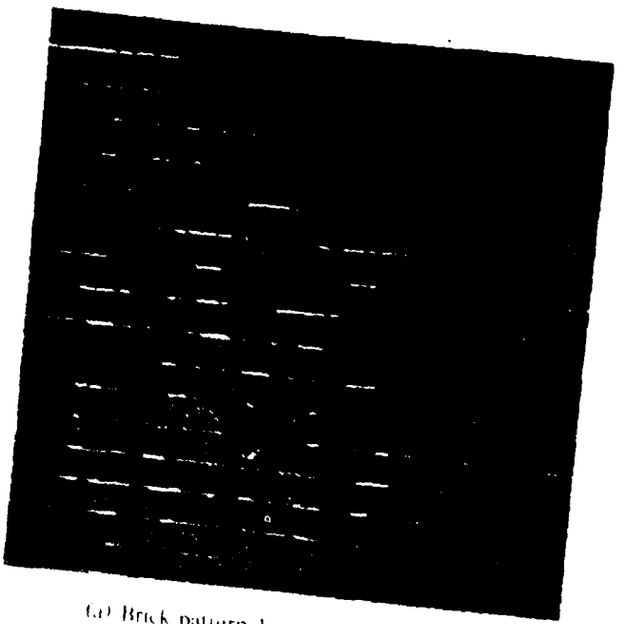
(c)



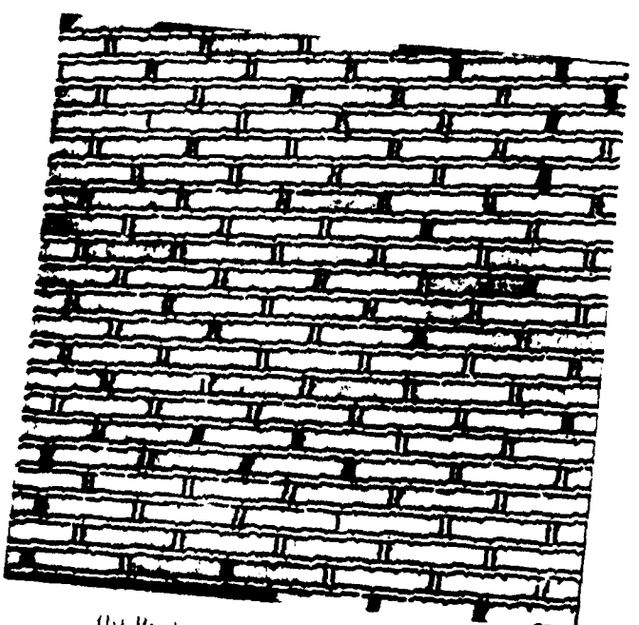
(c) Brick p[illegible]



(d) [illegible]

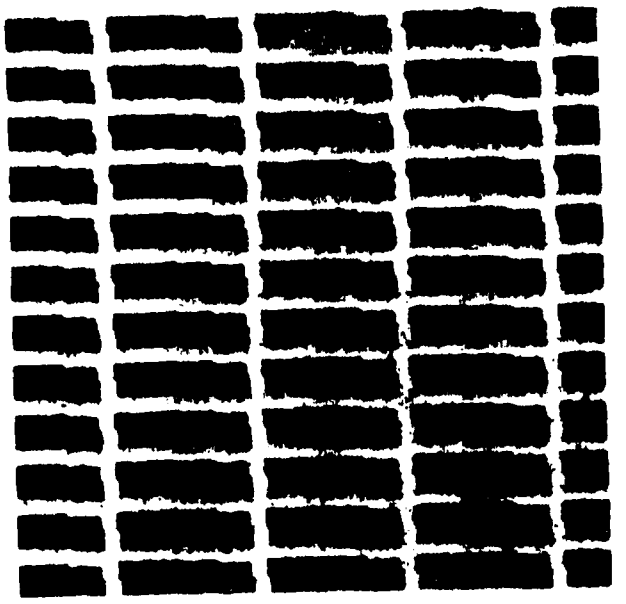


(a) Brick pattern 1 composite primitives

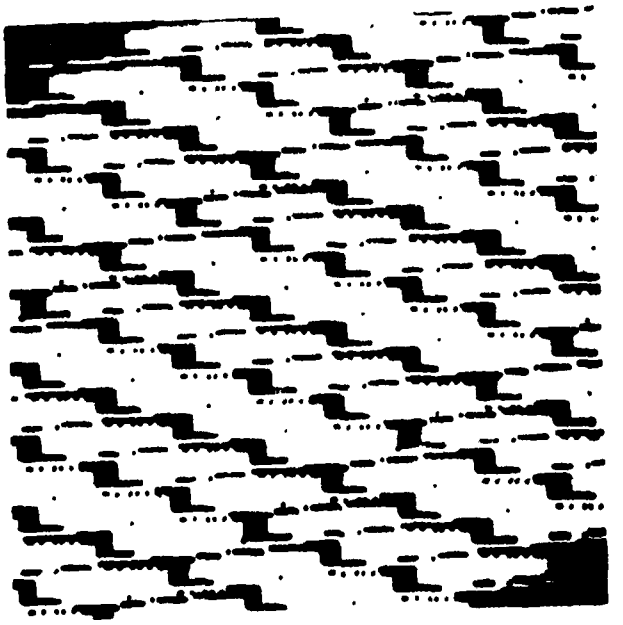


(b) Brick pattern 2 composite primitives

Figure 5

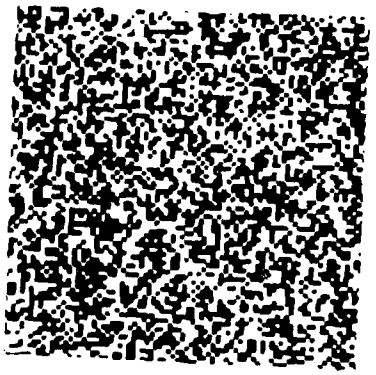


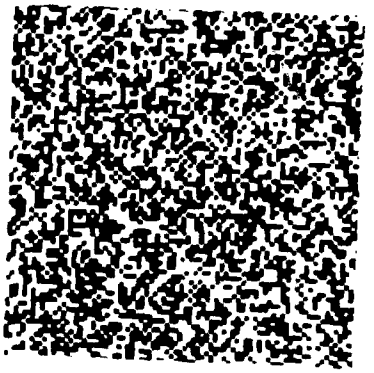
(a) Brick pattern 1 reconstruction

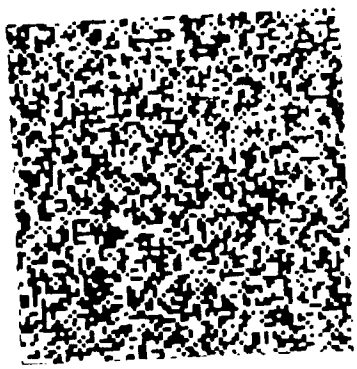


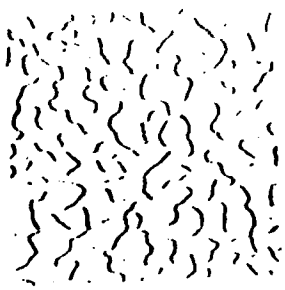
(b) Brick pattern 2 reconstruction

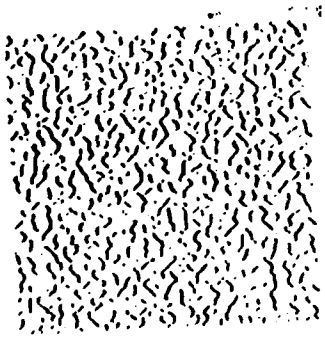
Figure 7

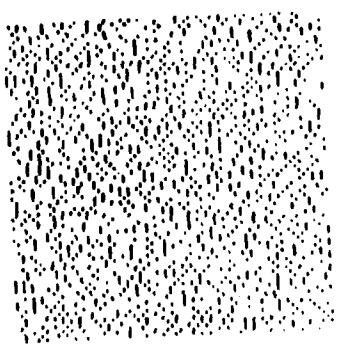












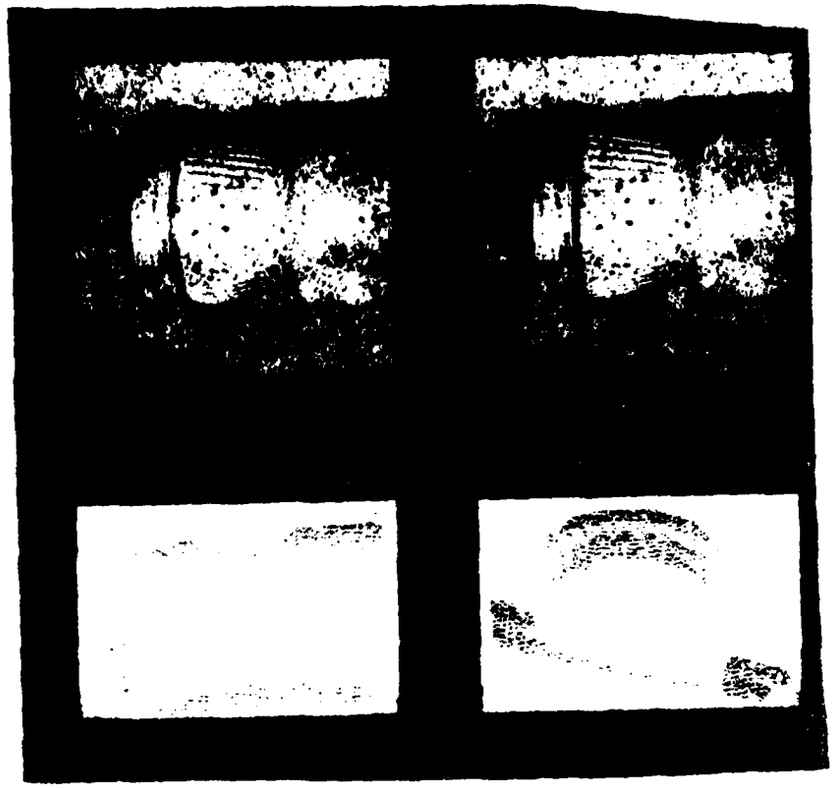


fig 4.4

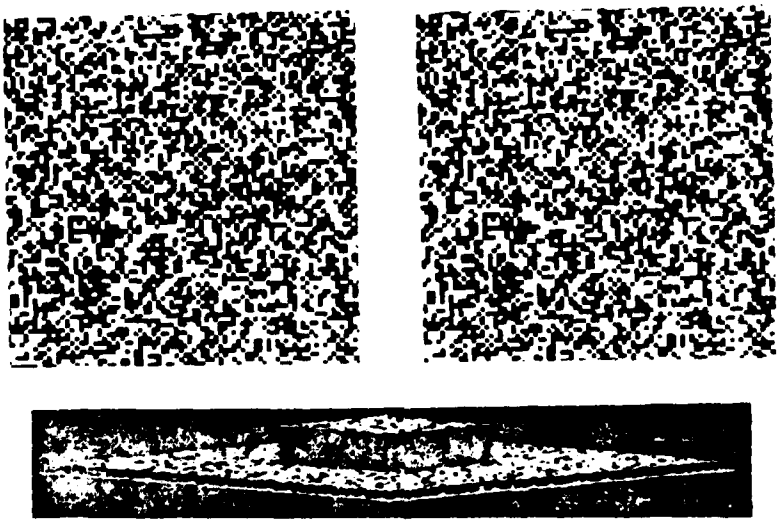


fig 45

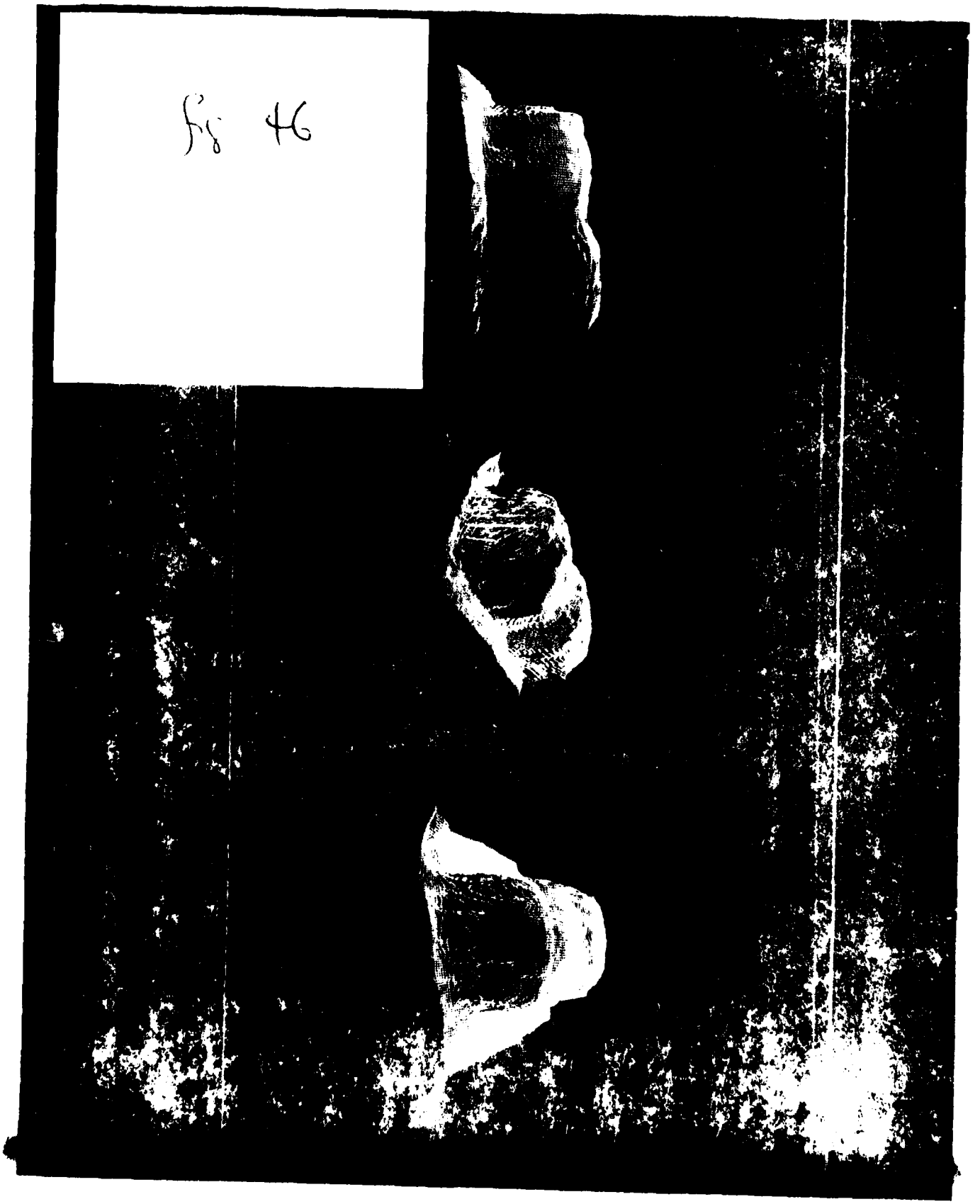
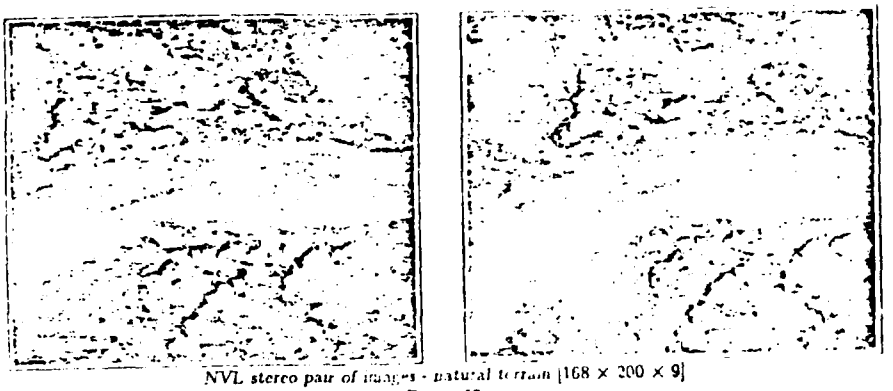
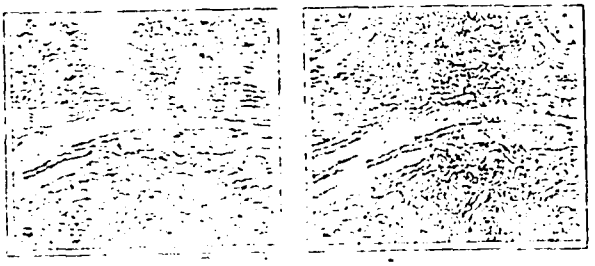


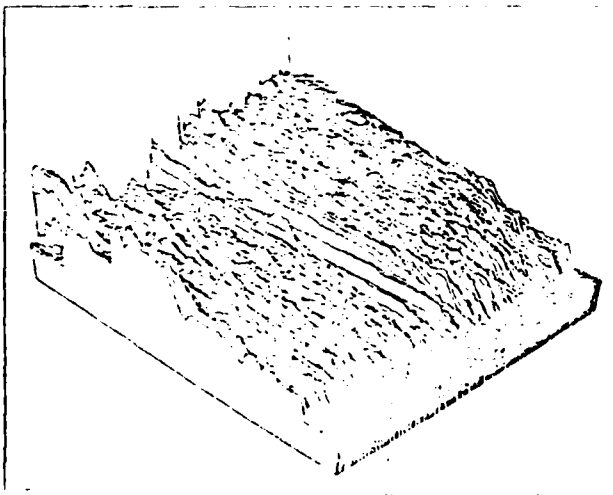
fig 46



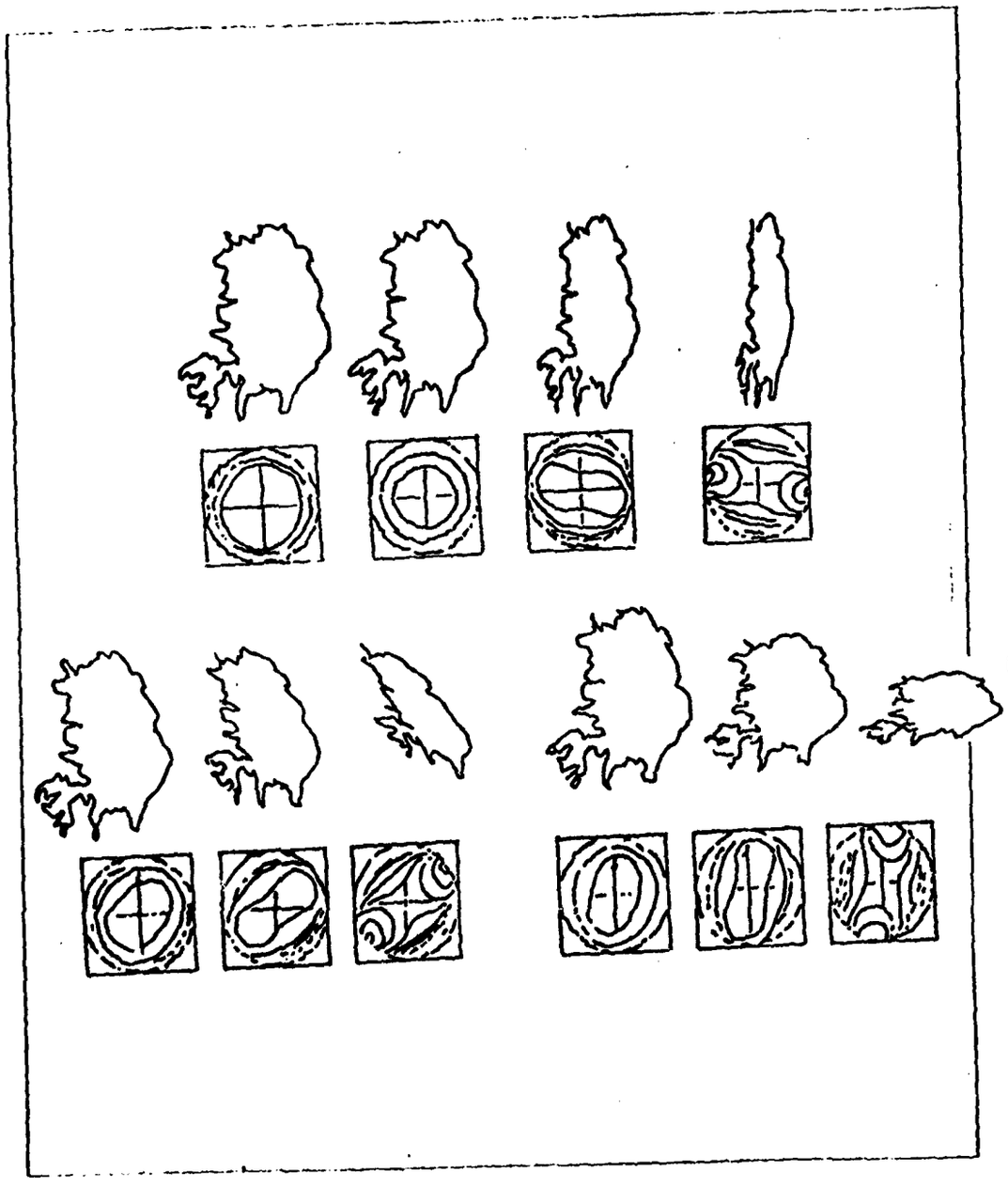
NVL stereo pair of images - natural terrain [168 × 200 × 9]

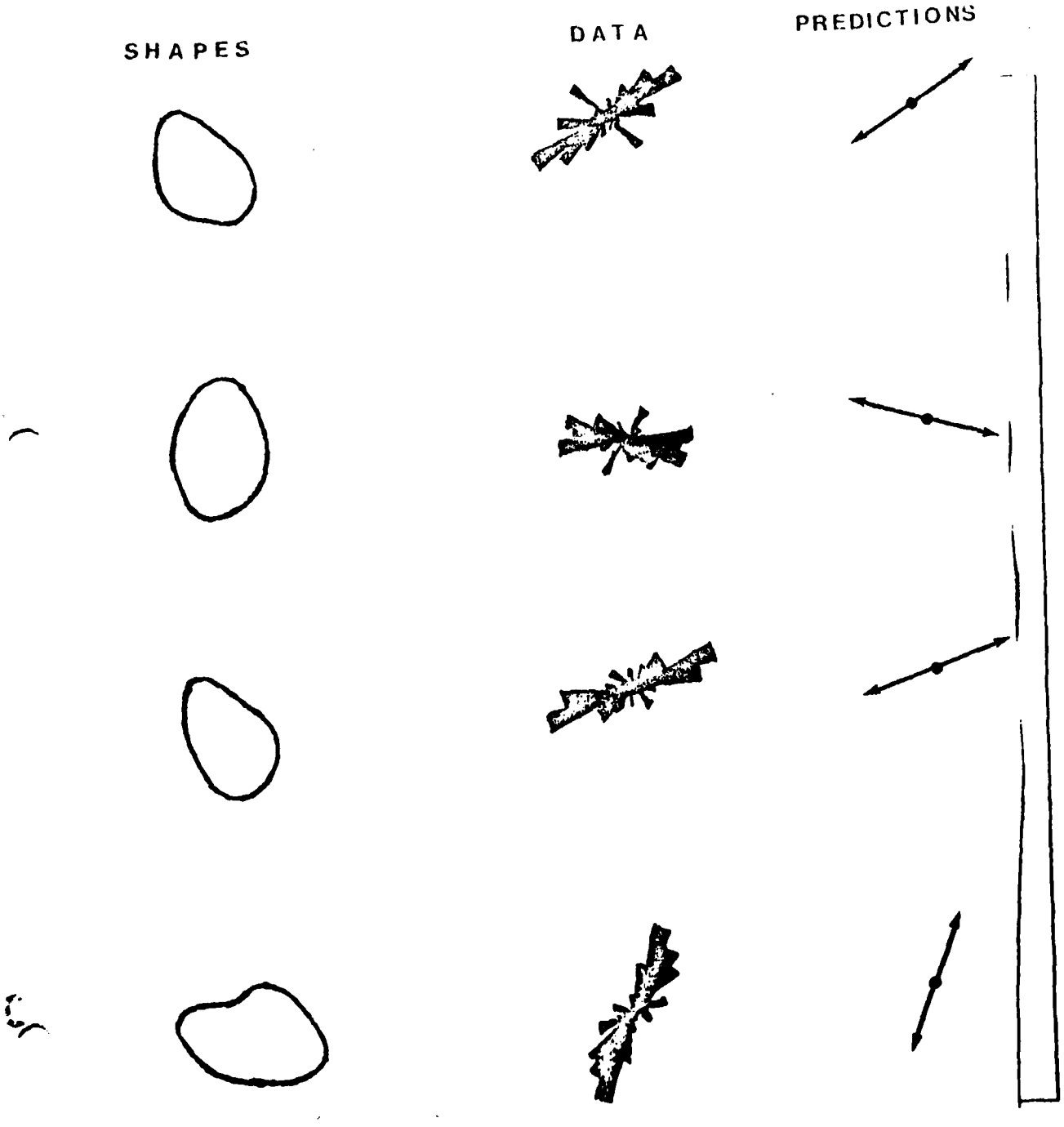
Figure 10



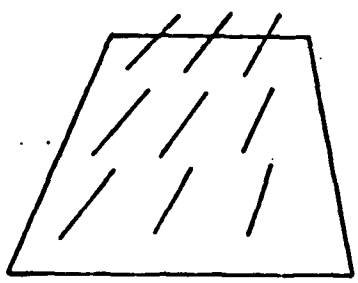


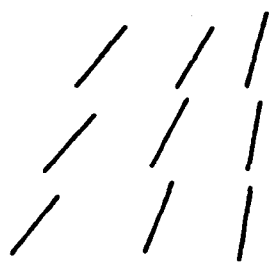
Perspective view of [illegible]




SHAPES
DATA
PREDICTIONS

Gendreau





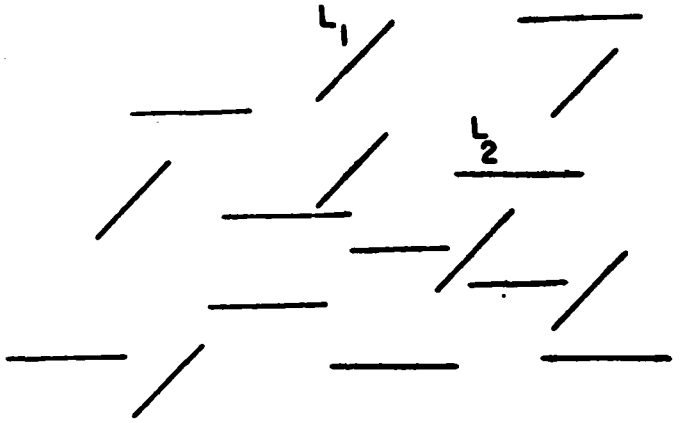
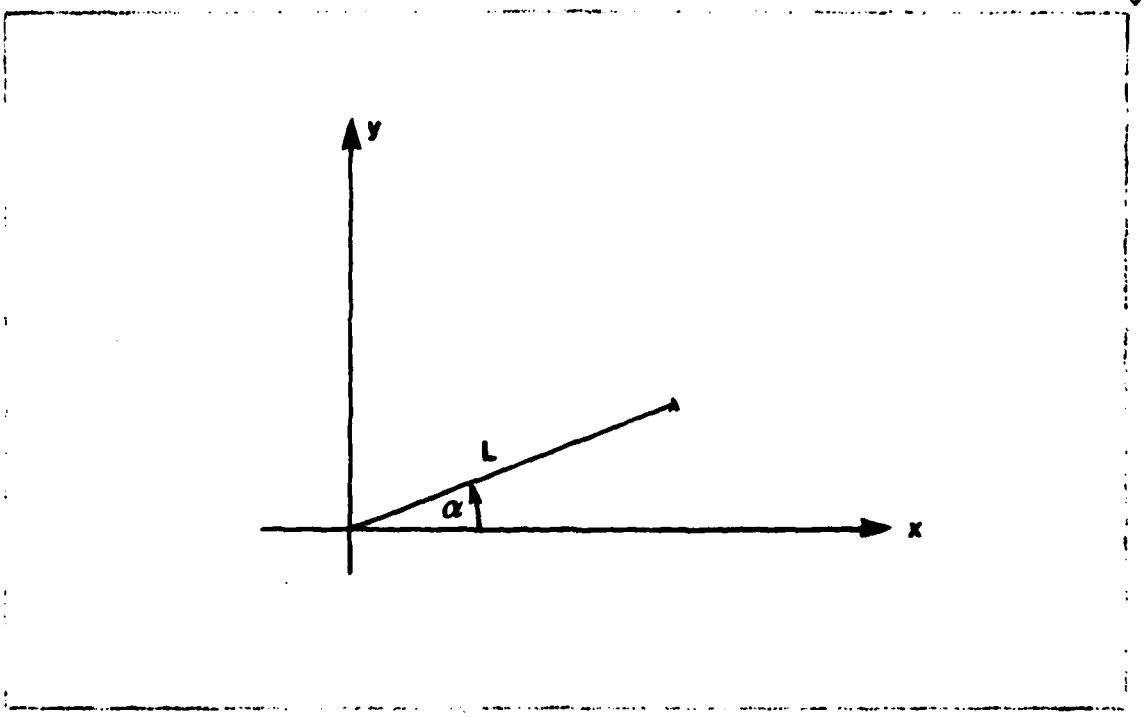
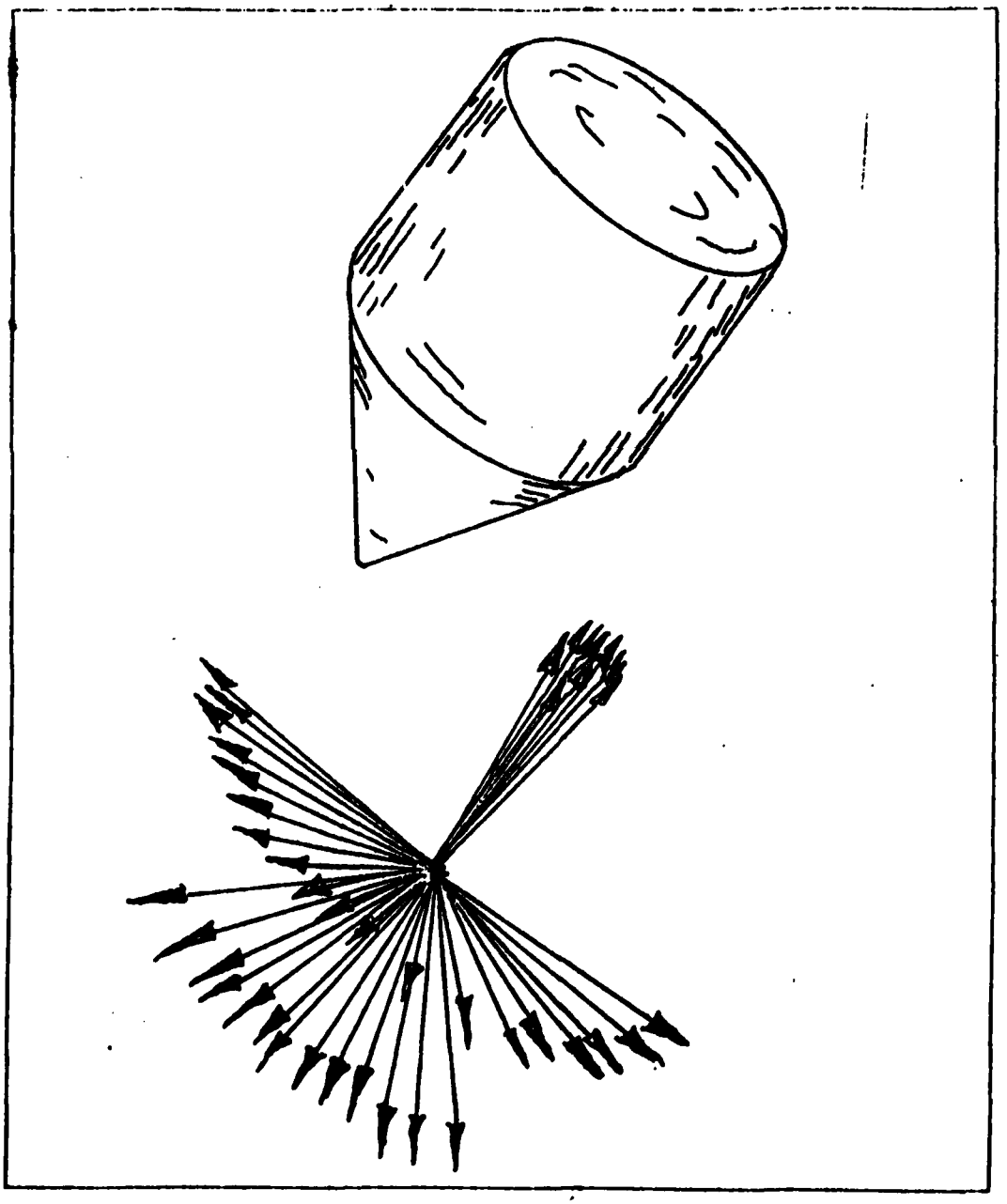
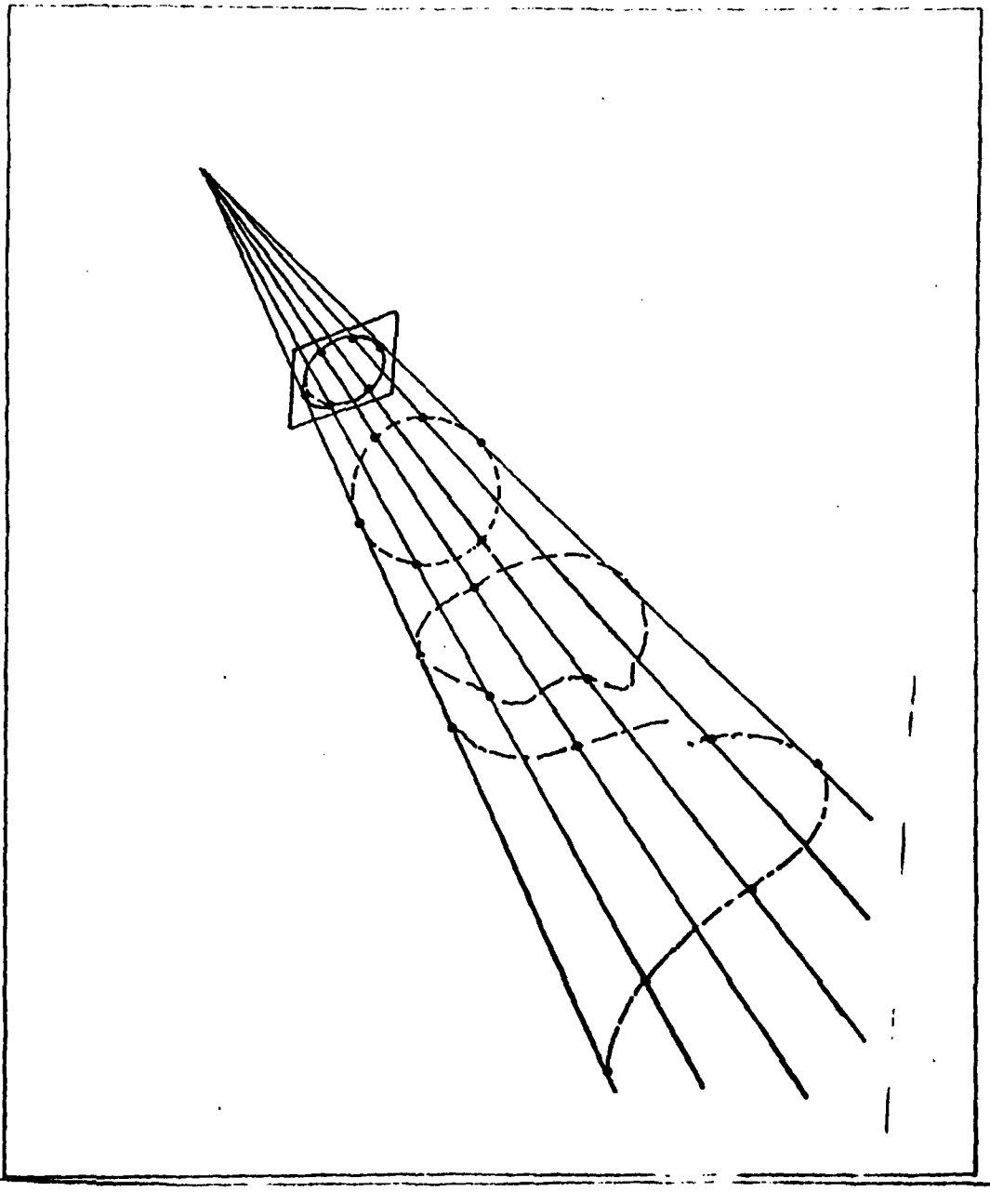
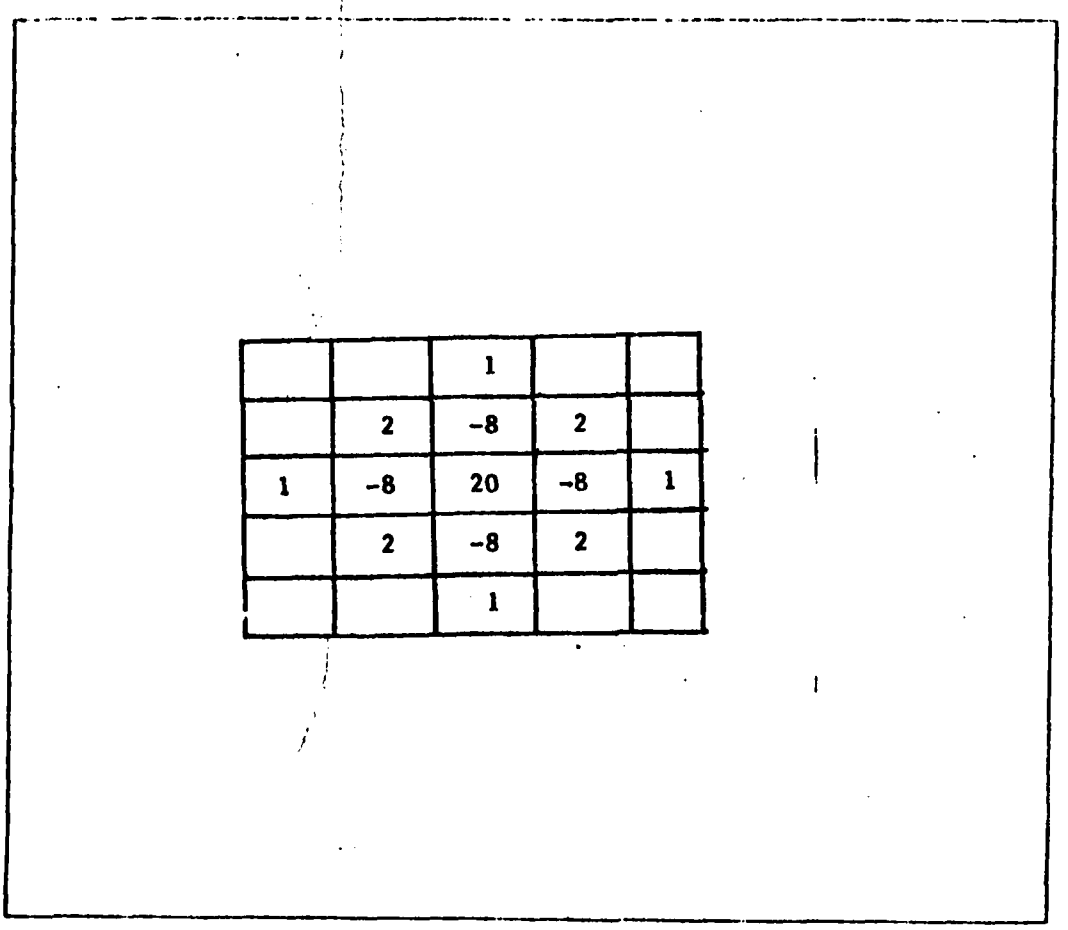

L_1
L_2

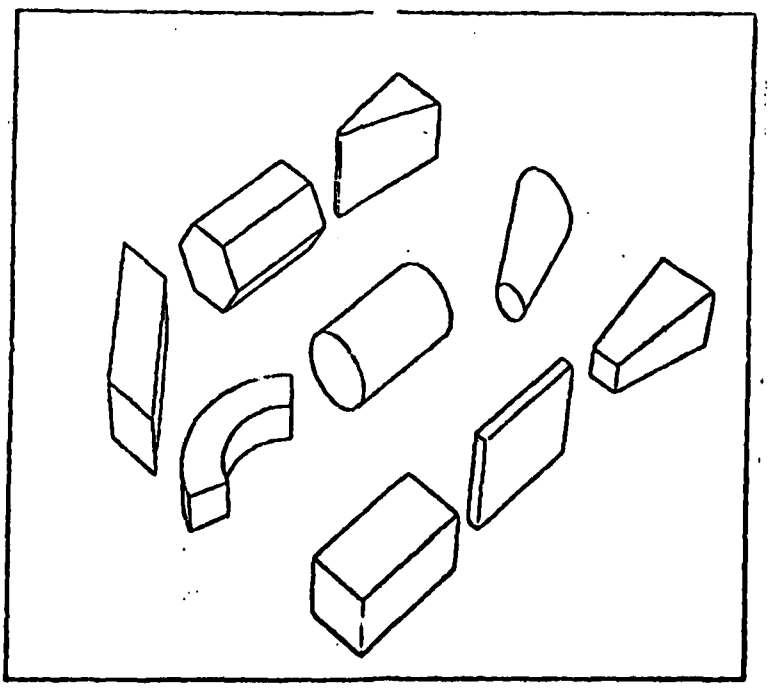
Fig 3








1
2 -8 2
1 -8 20 -8 1
2 -8 2
1



END

DATE
FILMED

1-82

DTIC