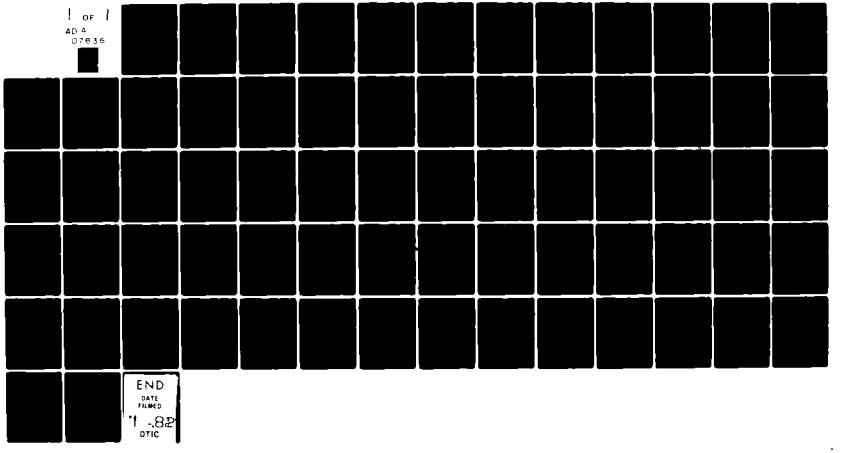
AD-A107 636 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 20/11
THE REPRESENTATION OF IMAGE TEXTURE.(U)
SEP 81 M D RILEY N00014-80-C-0505
UNCLASSIFIED AI-TR-649 NL



(12)

AI-TR-649

LEVEL I

THE REPRESENTATION OF IMAGE TEXTURE

Michael Dennis Riley

SEPTEMBER 1981

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AI-TR-649	2. GOVT ACCESSION NO. AD-A107636	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Representation of Image Texture		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael Dennis Riley		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0505 NSF 79-23110MCS
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 September 1981 13. NUMBER OF PAGES 69
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)

18. SUPPLEMENTARY NOTES

None

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

Visual texture
Texture boundaries
Texture edges
Visual representations
Texture discrimination

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

This thesis explores how to represent image texture in order to obtain information about the geometry and structure of surfaces, with particular emphasis on locating surface discontinuities. Theoretical and psychophysical results lead to the following conclusions for the representation of image texture: (1) A texture edge primitive is needed to identify texture change contours, which are formed by an abrupt change in the 2-D organization of similar items in an image. The texture edge can be used for locating discontinuities in surface structure and surface geometry and

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)

20. ~~and~~ for establishing motion correspondence. (2) Abrupt changes in attributes that vary with changing surface geometry -- orientation, density, length, and width -- should be used to identify discontinuties in surface geometry and surface structure. (3) Texture tokens are needed to separate the effects of different physical processes operating on a surface. They represent the local structure of the image texture. Their spatial variation can be used in the detection of texture discontinuities and texture gradients, and their temporal variation may be used for establishing motion corresondence. What precisely constitutes the texture tokens is unknown; it appears, however, that the intensity changes alone will not suffice, but local groupings of them may. (4) The above primitives need to be assigned rapidly over a large range in an image.

The Representation of Image Texture

by

Michael Dennis Riley

Massachusetts Institute of Technology

September 1981

Revised version of a thesis submitted to the Department of Electrical Engineering and Computer Science on February 4, 1980 in partial fullfillment of the requirements for the degree of Master of Science.

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 and in part by National Science Foundation Grant NSF 79-23110MCS.

Abstract

This thesis explores how to represent image texture in order to obtain information about the geometry and structure of surfaces, with particular emphasis on locating surface discontinuities. Theoretical and psychophysical results lead to the following conclusions for the representation of image texture:

(1) A ***texture edge*** primitive is needed to identify texture change contours, which are formed by an abrupt change in the 2-D organization of similar items in an image. The texture edge can be used for locating discontinuities in surface structure and surface geometry and for establishing motion correspondence.

(2) Abrupt changes in attributes that vary with changing surface geometry -- orientation, density, length, and width -- should be used to identify discontinuties in surface geometry and surface structure.

(3) ***Texture tokens*** are needed to separate the effects of different physical processes operating on a surface. They represent the local structure of the image texture. Their spatial variation can be used in the detection of texture discontinuities and texture gradients, and their temporal variation may be used for establishing motion correspondence. What precisely constitutes the texture tokens is unknown; it appears, however, that the intensity changes alone will not suffice, but local groupings of them may.

(4) The above primitives need to be assigned rapidly over a large range in an image.

Acknowledgements

David Marr provided insight, direction, enthusiasm, encouragement, and support. Any worthwhile ideas in the first part of this paper probably originated in discussions with him.

I thank Shimon Ullman for carefully reading the drafts and providing many valuable criticisms and suggestions that lead to the current organization and emphasis of this paper.

I also thank Whitman Richards, Mike Brady, and especially Kent Stevens for reading the drafts and for providing many useful comments.

Finally, I thank Patrick Winston for his encouragement and support.

Contents

1. Introduction

This paper explores how to represent image texture in order to extract information about the physical surfaces. Recent work by Marr [1977] suggests that the description of viewed surfaces plays a fundamental role in early visual processing and that determining the form of the descriptions given to the image and to the viewed surfaces should be one of the first steps taken toward understanding early visual processing. This paper analyzes texture in terms of these surface considerations and this representational viewpoint, investigating what aspects of texture should be made explicit in an image to obtain information of the geometry and structure of surfaces, with particular emphasis on locating surface discontinuities. This sets apart this study of texture from many others, which emphasize texture discrimination, a task that probably serves different goals.

In this introduction, we shall first expand on the aforementioned role of surfaces and representations in early visual processing, and on the use of texture to obtain surface information. Some methodological issues will then be discussed that reflect on the current level of understanding about the representation of texture.

The role of surfaces in visual processing

The visual world is composed mostly of surfaces. An image can thus be attributed to four physical factors: the *surface geometry* (how the surfaces lie in space), the *surface reflectance*, the *illumination*, and the *viewpoint* [Horn 1977]. For a sequence of images separated in time an additional attributing factor is needed: the *surface correspondence* between successive images (which will be non-trivial if the surfaces are in motion relative to the viewer). It would be of great value if these factors could be determined from an image or sequence of images since this would provide information directly of the physical world that is present only indirectly in their combination in an image. The human visual

processor's facility at finding the shape and arrangement of visual surfaces, their lightness and color, the location of discontinuities in surface orientation, depth, and reflectance indicates that this information can indeed be determined to a considerable degree. But how is it done?

Using image texture to infer surface information

The major sources of information about visual surfaces in an image include shading, stereo, motion, texture gradients and edges. The first several make direct use of the intensity changes present in an image. Shading obviously does so. Marr & Poggio [1978] have shown that the intensity changes present at several scales (the *zero-crossings*) are effective correspondence tokens for stereo matching. These intensity changes can also be used to obtain directionally sensitive motion information [Marr & Ullman 1981]. The intensity changes in an image thus seem to provide sufficient constraint to exploit these sources, and an understanding of the intensity change description was evidently crucial to the success so far [Marr & Poggio 1978, Marr & Hildreth 1980].

A precise understanding of how to distinguish among discontinuities in surface orientation, depth, reflectance, and illumination, of how to find motion correspondence over a large range in an image, and of how to obtain surface orientation and depth from texture gradients has proved more elusive. In part, this may be because the intensity changes in an image alone do not provide sufficient constraint to solve these problems easily, but that other aspects of the 2-D information in an image such as texture must also be made explicit and used. Let us briefly examine, in turn, each of these latter sources of surface information.

The location of a discontinuity in surface orientation, depth, reflectance, or illumination in an image often coincides with an intensity edge. But can the physical type of discontinuity (e.g. depth change, orientation change, illumination change) be determined from the intensities directly? By looking at the intensity gradient at an edge, Ullman's light source detection operator can, in principle, distinguish

a pure reflectance change from other discontinuity types (e.g. illumination change) [Ullman 1976]. By examining the edge profiles, other edge parsings may be possible [Horn 1977]. It is not presently known how well edges can be parsed into their physical correlates directly from intensity information in real images. A discontinuity in image texture originates at a discontinuity in surface structure or in surface geometry, and can therefore be used to locate these two kinds of physical discontinuity. The location of surface discontinuities provides information that is useful, for instance, to processes that must decide where smooth surface assumptions are no longer valid, as in the interpolation of a surface across points derived from stereo matching. Considerable emphasis will be given to locating surface discontinuities in this paper.

Motion correspondence across several degrees of visual angle in successive images (at which human's are quite adept -- the well-known apparent motion effect) is considerably more difficult problem than stereo correspondence since it involves increased range, unknown direction of motion, and the possibility of surface transformation over time. Given the profusion of intensity changes present in a real image, motion correspondence driven solely on the intensity changes results in many candidate matches for each motion token (e.g. edge fragment). Ullman [1979] approached this problem by assigning a likelihood to each possible match between images assuming nearby matches were more likely, and computing the maximum likelihood solution for that pair of images. An alternate approach would be to use larger scale tokens such as texture discontinuities and collinear groupings, which should have fewer candidate matches over a given range than the raw intensity changes, to bring the longer range motions into correspondence. Ullman noted that tokens that were more abstract than the raw intensity changes could be used to establish motion correspondence in humans, and called them *group tokens*.

Determining surface depth from texture gradients requires extracting a measure that shows no foreshortening in an image; this is necessary to factor out the effects of changing surface orientation from

those due to perspective [Stevens 1981a]. In Figure 1.1, surface depth cannot be obtained from the height of the ellipses since this measure is parallel to the texture gradient and will vary both with surface surface and depth. Thus, this distribution of heights could be due to either a cylinder (changing height due mostly to changing surface orientation) or a receding plane (changing height due entirely to changing depth). However, if the width of the ellipses is used and provided that the ellipses are congruent across the surface, then surface depth can be obtained, since this measure is perpendicular to the texture gradient and will not show foreshortening. Thus, the variation in ellipse widths will be due entirely to changing depth. Steven's method for finding this measure with no (or least) foreshortening essentially assumes that a description of image texture is available. In particular, such information as the position and dimensions of small blobs in an image would be useful, while the location of the intensity changes alone is probably too primitive a description of an image from which to extract an unforeshortened measure directly.

In summary, distinguishing among discontinuities in surface orientation, depth, reflectance, and illumination, finding long-range motion correspondence, and obtaining surface orientation and depth from texture gradients may prove difficult if only the intensity changes are examined directly, while if the information in image texture is used, these problems may prove tractable. This makes it imperative to understand what aspects of image texture should be identified in an image. Without knowing what relevant data will be available, it is impossible to precisely define, say, a motion correspondence process or a depth from texture process, with the best that can be determined are these processes' abstract computational needs. Thus, we could say that a motion correspondence process requires image tokens that remain in correspondence with the same physical feature in successive views and for which there are typically a small number of possible matches over the desired range. For depth from texture gradients, an unforeshortened measure in the image is needed. But to be much more specific requires knowing the

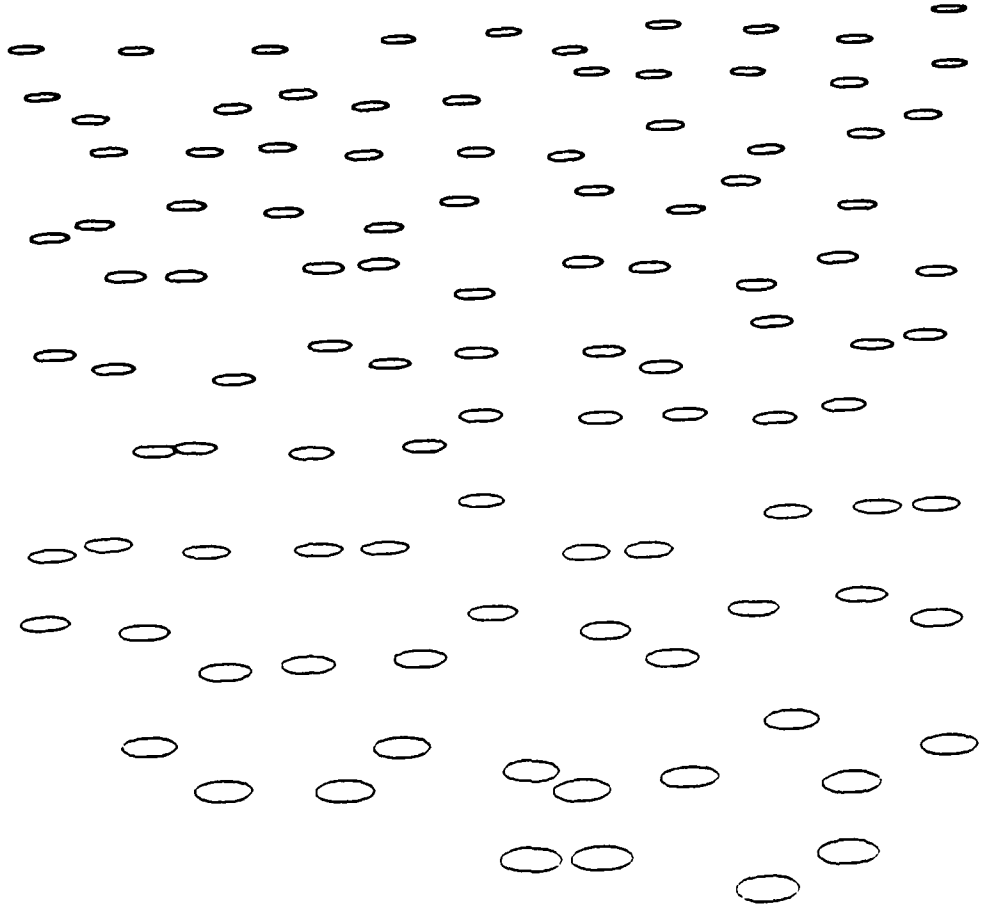


Figure 1.1 Surface depth cannot be obtained from the height of the ellipses, since this measure is parallel to the texture gradient and will vary both with surface orientation and depth. Surface depth can be obtained from the width of the ellipses, however, since this measure is perpendicular to the texture gradient and will not show foreshortening. Provided the ellipses are congruent across the surface, their width will be inversely proportional to their distance from the viewer. [Figure courtesy of K. Stevens]

form of the input data, in particular, knowing what aspects of image texture to detect in an image and how they should be represented in the visual system.

Representational Emphasis

We seek to determine the early visual representation of image texture, since the form of the description of image texture must be specified before its computation can be specified. If the broad goals of the computation are not well understood, but instead some image computation is defined prematurely, the results are likely to be of little value in the long term to the theory of vision. This representation's primitives -- the basic assertions that can be made about image texture -- need to be specified, in particular. Other important representational issues to be determined include the range and resolution over which these primitives can be assigned in an image, and the referencing system for retrieving these primitives (see Marr and Nishihara [1978] for a discussion of visual representations). Marr [1976] has called the early representation of the intensity changes and 2-D geometric structure in an image the *(full) primal sketch* (the *raw* primal sketch represents just the intensity changes).

The primal sketch is the first of several representations that Marr [1977] sees as having a central role in the computational theory of vision. The primal sketch is used to construct the *2½-D sketch*, a viewer-centered representation of the visible surfaces in a scene. It is in the 2½-D sketch that the various factors that produce an image are separated -- the surface geometry, surface reflectance, the illumination, and the viewpoint. Many processes that provide surface information from images, such as depth from texture, can be viewed as reading from the primal sketch and writing to the 2½-D sketch.

The term *early* texture representation is used to indicate that we are interested here in the description of texture that is produced early in the visual processing, and is used for extracting global surface information (the creation of the 2½-D sketch), and not a much richer description produced by local scrutiny that we might expect exists for the purposes of recognition, and is much more limited in

speed and image range than the early texture representation.

Informal definition of image texture must precede its precise computational definition

It is inevitable that the definition of image texture will be imprecise initially; we have to rely upon an intuitive definition. This has been the case with other aspects of visual processing. An intensity edge, for instance, is informally defined as a place in an image where the intensity changes abruptly, with a surface correlate of a discontinuity in surface orientation, depth, reflectance, or illumination. Recently, Marr & Hildreth [1980] have formally defined an edge in terms of the spatial coincidence of intensity changes at two nearby scales found by a convolution operation that will be described later. Their method defines a precise computation on an image for detecting edges. The informal definition, however, existed first, specifying roughly what is to be represented, and what significance it has with respect to physical surfaces. The formal definition then specifies how it is to be detected from an image. The idea of detecting abrupt intensity changes is very intuitive and was an important precursor to determining their precise computation. The aspects of image texture that should be detected is not as intuitively obvious. Thus, we must begin by understanding roughly what aspects of image texture should be represented in an image and what are their physical correlates. Once we have approximate definitions of what we want, we can then examine exactly how to compute them from an image. Such informal definitions can also be used to test for their psychophysical existence.

This paper is divided into two parts. Part I develops the theory of the representation of texture, and comprises Sections 2 through 5. In Section 2, physical constraints on surface structure are formulated. In Section 3 and 4, two kinds of image texture primitives, the *texture edge* and the *texture tokens* respectively, are introduced along with the rationale for their utility to the visual system. Section 5 summarizes Part I.

Part II of this paper is devoted to demonstrations of the human visual system's early representation of texture, serving as a check on the utility of these primitives to a successful visual processor. Section 6 describes demonstrations supporting the existence of a texture edge primitive in this representation, and Section 7 describes demonstrations that restrict the range of what constitutes the texture tokens in this representation. Section 8 summarizes Part II.

2. Physical Constraints on Surface Structure

An image is a two-dimensional projection of the three-dimensional world. An important goal of early visual processing is, in a sense, to invert this mapping. If the point in space corresponding to each image point could have arbitrary position and brightness, this task would be impossible. Our abilities to perceive the 3-D world visually indicate, of course, that this is not the case. The visual world must be otherwise constrained. These physical constraints on the visible world and on the projected image must be identified in order to understand how to infer backward from an image. Three physical constraints will be identified that are relevant to surface structure. These constraints in their original form are due to Marr [1981].

The predominance of surfaces

In the introduction, the visible world was considered composed mostly of surfaces that are smooth enough that their local surface orientation could be discussed. For instance, a leaf defines such a smooth surface. A hedge containing this leaf will itself define a smooth surface when viewed from sufficiently far away. Even at distances where its leaves can be resolved but the variation in the distance to them is small relative to their absolute distance from the viewer, the hedge can be considered an approximately smooth surface. Thus, only in a physical situation such as a snowstorm would suitable surfaces be hard to define.

A leaf's reflectance function would be fairly constant over its surface if it were uniformly pigmented. For a hedge, however, its composite structure and the effects of mutual illumination and occlusion would make the spatial variation of its reflectance function very complex. This illustrates our first constraint: *the visible world can be regarded as being composed of smooth surfaces having reflectance functions whose spatial variation may be complex.*

There are two consequences of this constraint in an image. First, image points typically originate from surface points. Second, it may be very difficult to determine analytically the geometry of a surface such as a hedge from the intensity values directly (i.e. by treating it as a shading problem) even if the location of the light sources is known, because of the complex nature of its reflectance function.

While an analytic statement of the spatial variation of the hedge's reflectance function may be complex, defining its spatial structure with respect to items that constitute it could be less so. The leaves that form the hedge's surface may be of uniform size and density. The leaves themselves may have markings with their own characteristic attributes. Explicit descriptions of each of these kinds of surface item present in the hedge will capture information that is otherwise buried in its analytic reflectance function. Two additional constraints formalize this notion.

Different processes form different kinds of surface items

A leaf and a leaf marking are different not only to our senses, but they are intrinsically different in terms of their physical nature and origin. In order to formalize this intuitively simple idea, we can think of leaves as being generated by some physical process operating on a surface at a given scale, while leaf markings are generated by some different processes operating at a smaller scale. This provides the second constraint: *physically different processes operate on a surface to form different kinds of items there.* One set of processes operating at a given scale, thus, determines the size and shape of the leaves in a hedge. Another forms the markings on those leaves. One set of processes determines the spatial arrangement of the hairs on an animal's coat. Others form the spots and markings on that coat. This constraint is important because it permits a physical distinction to be made between those aspects of surface structure that are essentially the same kinds of items (such as two leaves in a hedge), being due to the same physical processes, from those that are different kinds of items (such as a leaf and a leaf marking, or a leaf and a brick), being due to very different processes.

Items generated by the same processes are similar

The third constraint is: *surface items generated by the same physical processes tend to be more similar to one another in their size, shape, lightness, color, and spatial arrangement than to surface items generated by other processes.* This states that with respect to these attributes, a leaf is more likely similar to another leaf than, say, to a brick.

In an image, the projection of the surface items generated by the same processes will tend to be more similar to one another in size, shape, contrast, color, orientation, and spacing, than to the projection of other surface items that are generated by different processes. Note, however, that the similarity may be preserved only locally in an image. Changing surface geometry and perspective projection can destroy global similarity since size, contrast, orientation, and spacing can all vary with changing surface geometry.

3. The Texture Edge

As stated in the introduction, an important goal of early visual processing is determining the different physical factors that produce an image. In particular, this involves decoupling surface orientation, depth, and the location of discontinuities in these from surface reflectance and illumination. In this section, we shall focus on surface discontinuities. We shall see that one consequence of the previous section's constraints is that abrupt changes in texture in an image can be used to identify discontinuities in surface geometry and surface structure.

The location of surface discontinuities is not explicit in the intensity changes

The location of discontinuities in surface structure or surface geometry are not yet explicit in the intensity changes. There may be a myriad of contours present in the intensity changes, only a few of which coincide with a *discontinuity in surface geometry or surface structure*. Others will be due to the internal structure of a surface or to shadows and highlights. For example, in Figure 3.1 the bottom-most horizontal line, which coincides with the texture boundary, may indeed be present in the intensity changes but nothing there distinguishes it from the other horizontal lines, also present in the intensity changes, as the location of a texture change in the image, and thus the likely location of changing surface structure or surface geometry (e.g. a brick wall abutting a grass lawn). There may even be no significant intensity change coinciding with the image of a surface discontinuity, while contours defined by the *image structure may* still be present there. It is the image structure contours that hold the key to identifying discontinuities in surface geometry and surface structure.

Two types of image structure contours

Not every contour in an image is defined solely by intensity changes coincident with the contour. A contour can also be defined by image structure and in at least two different ways. One kind

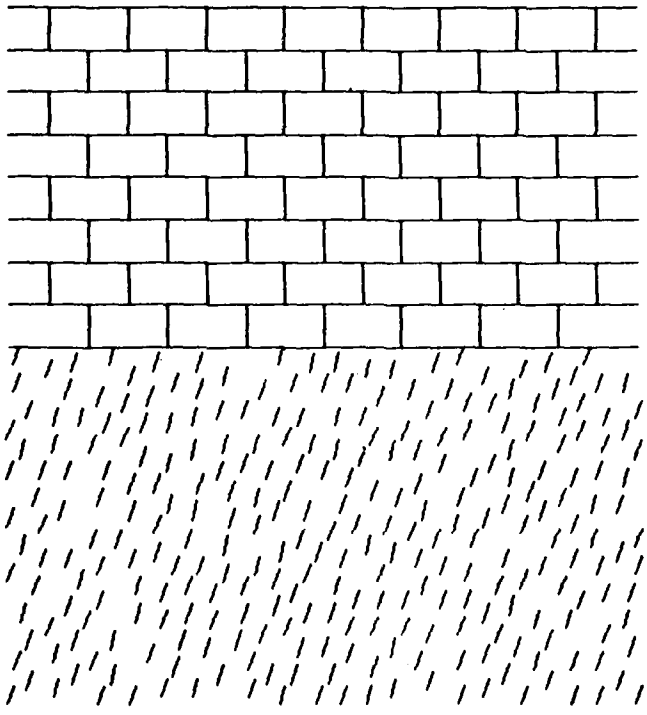


Figure 3.1 There are many contours in this figure that are explicit in the intensity changes; for instance, the bottom-most horizontal line at the texture boundary is present there. Nevertheless, this line has not yet been distinguished from the other horizontal lines, which are also present in the intensity changes, as the location of a texture discontinuity in the image. Locating such abrupt texture changes in an image is important, since they identify the likely location of discontinuities in surface structure or surface geometry.

can be created by an abrupt change in some 2-D organization in an image. In Figure 3.2, for example, the 45° change in the orientation of the line segments defines a contour that corresponds to the boundary between the two oriented regions. A sudden change in local density of the line segments in this figure also defines such a contour, which corresponds to the external boundary of the two regions, with the line segment density vanishing outside these regions. We shall refer to such contours as *texture change contours*. A second kind of contour can be defined by the local alignment of various image features. For example, the local alignment of the terminations of the lines in Figure 3.3 defines such a contour. We shall call these *alignment contours*.

We explore texture change contours and their use in identifying discontinuities in surface geometry and surface structure in this section. Alignment contours will, for the most part, not be treated in this paper. Let us examine next the relationship between texture change contours and surface discontinuities.

Discontinuities due solely to changing surface structure

First, consider a discontinuity in surface geometry where the surface reflectance function is constant across the discontinuity. Examples of this are two surface fragments that are adjacent in an image and have the same surface structure and coloration but have different surface orientation, depth, or rotation. For instance, Figure 3.2 could be the image of a creased surface as shown in Figure 3.4a or, instead, it could be the image of two surfaces, one rotated 45° with respect to the other as shown in Figure 3.4b. Figure 3.5 could be the image of two similarly textured surfaces differing in depth (one $\sqrt{2}$ farther away than the other), or again it could be a creased surface (with, say, one side parallel to the image plane and the other side at a 60° slant).

From the constraints of the previous section, the image of a local patch of a structured surface where the surface geometry does not change much will likely contain, at particular scales, items that are

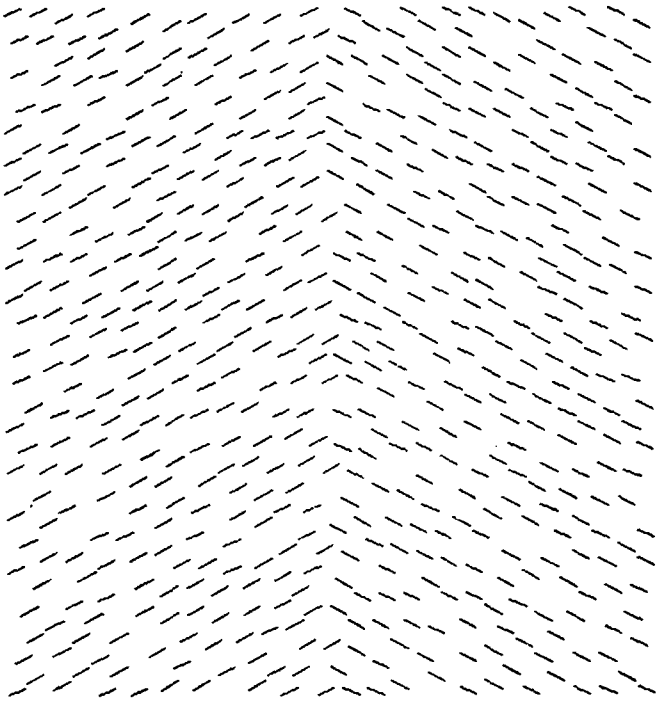


Figure 3.2 An image contour can be formed by a 45° change in the orientation of small line segments

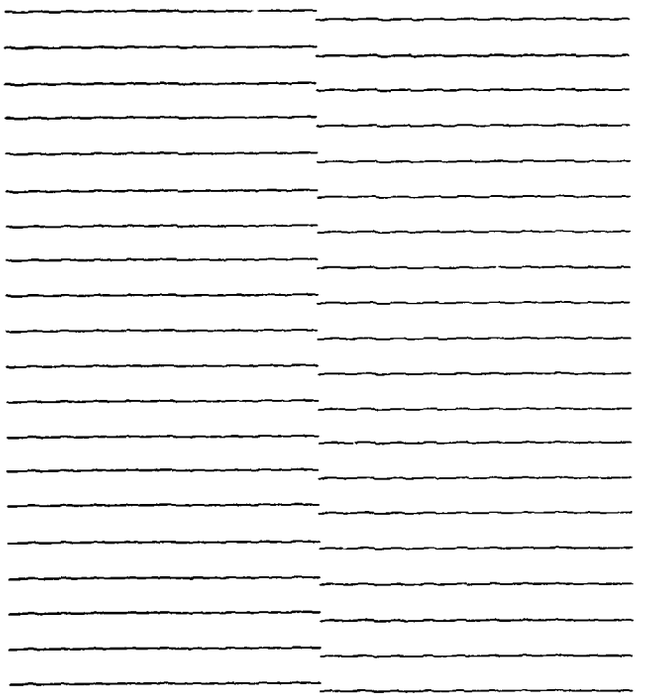
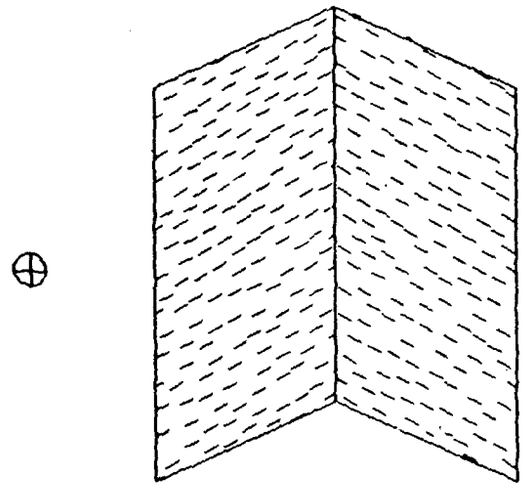
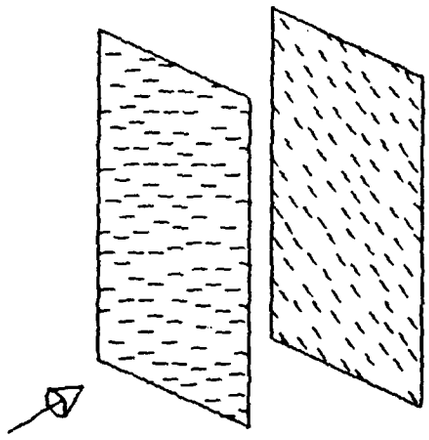


Figure 3.3 An image contour can also be formed by the alignment of line segment terminations.



(a)



(b)

Figure 3.4 Two of several possible physical origins for Figure 3.2: (a) a creased surface, and (b) a surface rotated relative to another surface with the same surface structure.

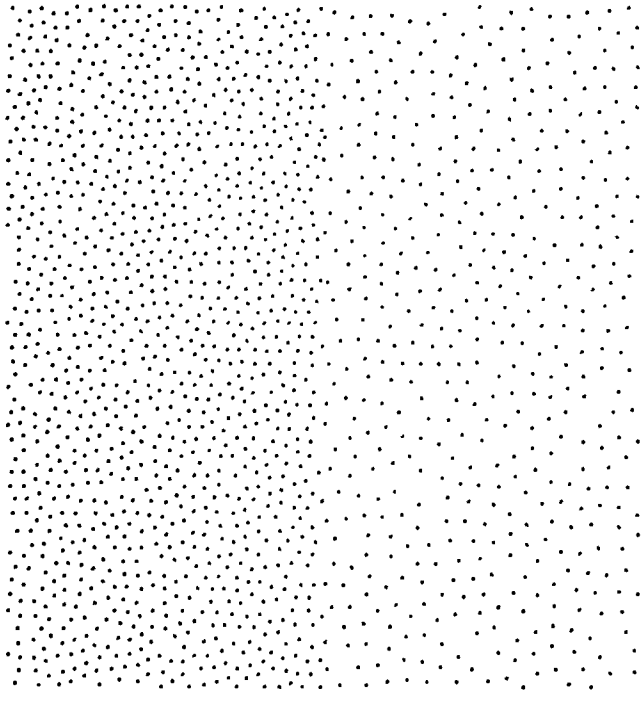


Figure 3.5 An image contour can be formed by a 2:1 density (number/area) change of small dots.

similar to one another in orientation, spacing, color, contrast, size, and shape. But where the surface geometry changes, geometric attributes such as orientation, density, and length of the image of the surface items will change. (Intensity, contrast, and color can also vary with changing surface geometry, although large contrast and color changes are unlikely since these would require perverse illumination or reflectance functions.) Thus, at a discontinuity due solely to changing surface geometry, there will often be an abrupt change in these geometric attributes of the image of similar surface items, forming a texture change contour.

Discontinuities due to changing surface structure

There is another physical source of texture change contours in an image, and this represents the other basic type of surface discontinuity -- one due to changing surface structure. For instance, Figure 3.5 could be the image of two adjacent surfaces lying in the same plane that have different dot densities. When surface structure changes, the similarity constraint of Section 2 indicates that items at given scales on one surface will likely be more similar to one another in orientation, color, contrast, size, and shape than to items on the other surface, resulting in abrupt changes in the items at each scale at the image location of the surface discontinuity, and giving rise to a texture change contour. In this case, however, any surface attribute can change, not just geometric attributes, the surface structure can change arbitrarily across this kind of surface discontinuity.

Texture change contours need to be made explicit

We have seen above that a texture change contour can be formed by a discontinuity in surface geometry or surface structure. A texture change contour can be due finally to some combination of these factors. Thus, a texture change contour identifies the likely location of a surface discontinuity of some form. This alone makes the representation of texture change contours valuable since, as we saw above,

the location of surface discontinuities may not be present explicitly in the intensity changes. This represents the first major implication for the early texture representation: *texture change contours should be made explicit in the image since they identify the likely location of discontinuities in surface geometry or surface structure, information that may not be explicit in the intensity changes alone.*

Separating the physical factors that produce texture change contours

Is it possible from an image to distinguish among those texture change contours due solely to changing surface geometry, those due solely to changing surface structure, and those due to some combination of these two factors? Unfortunately, the answer is that this cannot always be achieved from image texture information alone. When the surface structure changes completely, forming a texture change contour, there is no information in the image texture about whether the surface geometry changes there also. A structural change can also mimic a geometric change as, for example, when Figure 3.5 is due to a change in surface dot density, and not to a change in depth. However, it is possible to distinguish between those texture change contours that could be due solely to change in surface geometry, and those that must involve some surface structure change. The former contain only *geometric changes* in the image of the surface items across the texture change contour: it would be possible with suitable 3-D configurations of two surfaces having the same surface structure to project in the image as each of these texture changes. The latter contain *non-geometric changes*, as in Figure 3.6. No change in surface geometry can cause the squares in this figure to be transformed into dots having the same density as the squares. Instead, the surface structure must have changed. At the end of this section, we shall explore how to distinguish between geometric and non-geometric texture changes.

The texture edge primitive and its uses

The representation of an intensity change contour begins with intensity edge and bar primitives.

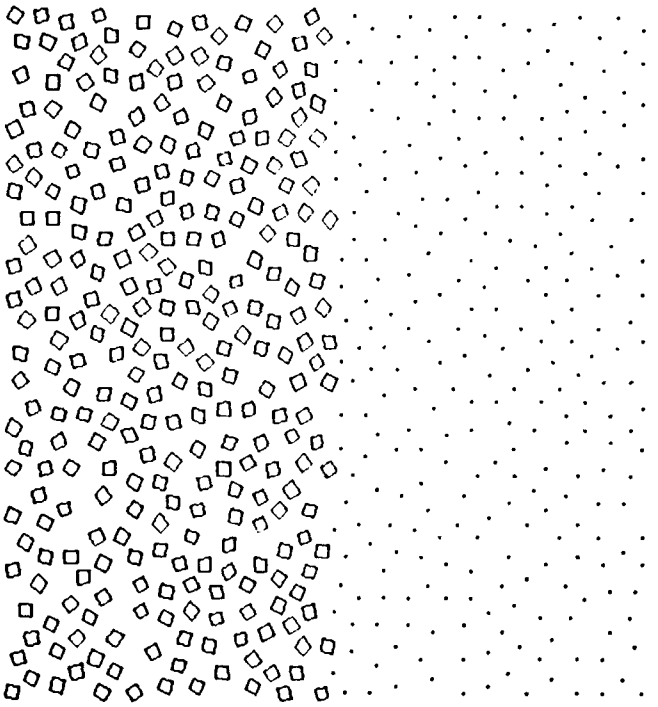


Figure 3.6 No 3-D configuration of two identically structured surfaces could produce this figure; no surface can appear composed of squares from one viewpoint, and of dots of the same density from a different viewpoint

which are local assertions assigned at many points along the contour that make explicit the position, local orientation, contrast, and width [Marr 1976, Marr & Hildreth 1980]. Analogous to this, points along a texture change contour in an image can be represented in our early texture representation by a *texture edge* primitive, which makes explicit local contour position and orientation at the very least.

We have already seen above that the representation of texture change contours is important for detecting surface discontinuities and can be used to distinguish between those discontinuities that possibly could be due solely to a change in surface geometry and those that cannot. In addition to this, the texture edge primitive could be useful for establishing motion correspondence. Given the many possible candidate matches of edge and bar fragments for motion correspondence over several degrees of visual angle, the larger scale and rarer texture edges give fewer possible matches over a given range.

Range of the representation

An issue of particular importance is the range in an image over which this texture edge primitive can be assigned, since this determines, in part, the computational burden of forming the early texture representation. One extreme of this range would be a representation that encompasses only a very small portion of an image (e.g. the fovea) at one time, or that allows only a very few primitives to be assigned at one time. At the other extreme would be a representation that encompasses the entire image and can allow as many primitive assignments as image resolution permits. While it is difficult at this point to say precisely where in this range our early representation of texture should lie, it can be said that it must lie closer to a full image range representation than to a very restricted but economical one that can represent only a small fraction of the texture edges found in an image. Very limited range or resolution may have be appropriate for some visual representations, but such limitations are undesirable for the early representation of image texture considering the uses to which this representation will be put.

As previously outlined, the full primal sketch, which represents both the intensity changes and

image structure, serves as the basic description of an image from which the 2½-D sketch, a viewer-centered representation of the viewed surfaces in space, is formed. In this framework, the early texture representation is considered a part of the full primal sketch. Further, the formation of the 2½-D sketch's description of the viewed surfaces -- their orientation, depth, reflectance, location of discontinuities -- is a fundamental goal of early visual processing. If, as has been argued above, the texture edge primitive makes explicit aspects of image structure that are useful for creating a representation of surfaces present throughout an image, then it follows that texture edges must be detected rapidly throughout the image. This is an expensive step, since it requires that considerable computational resources be brought to bear if an entire image is to be processed in a fraction of a second. Next, as texture edges are detected throughout an image, they need to be stored away somewhere, and the most direct way to do this is in a representational memory encompassing the entire image. This is particularly important for establishing large range motion correspondence using texture edges, since there is a wide image range over which a particular token could move. This approach may seem computationally expensive compared to the use of a scrutinizing processor for local analysis of surface structure that is directed more leisurely across the image. But such a local scrutinizing processor would be inherently too slow to rapidly cover large portions of an image and feed as input to the 2½-D sketch.

Detecting texture edges

Conceptually, the detection of texture edges can be divided into two major steps. First, the basic structural elements that will be used to represent the image texture locally must be made explicit. We shall call these primitive elements the *texture tokens*. Second, the spatial variation of these tokens are used to locate texture edges. It is not presently known what constitutes the texture tokens; this could conceivably range from grey-level values to primitives that represent individual texture elements and their attributes such as orientation, length, width, contrast, shape, and color (e.g. each line segment in

Figure 3.2). In Section 4, we shall see that the range in which the texture tokens lie can be restricted, but their precise form has yet to be resolved. Until it is, it will be difficult to say much about methods for detecting texture edges.

One issue that can be discussed at this time, however, is the desirable dimensions for the texture token attributes. We saw above that at a discontinuity due solely to changing surface geometry (constant surface structure across the discontinuity), it will be geometric dimensions such as orientation, length, and width that will vary with the changing surface geometry. It would therefore be desirable to have texture tokens that have attributes that change when the surface geometry changes, if discontinuities due solely to changing surface geometry are to be detected.

Discontinuities in surface structure can be detected in two ways. One way utilizes geometric attributes. When the surface structure changes, everything is likely to change including the geometric attributes given above. For example, the change in size of the items in Figure 3.6 could be used to identify the boundary between the two regions. A second way to detect discontinuities in surface structure would use changes in structural attributes. For example, the number of corners per item in Figure 3.6 could be used to identify the texture boundary between the two regions, since in the left-hand region there are four corners per item (square), while in the right-hand region there are zero per item (dot). This second method would be useful when all geometric attributes happen to match across the texture boundary causing the first method to fail. Whether this is likely to occur in natural images is uncertain however; a point that we shall return to in Section 6.

We have not yet discussed how to distinguish between discontinuties due solely to changing surface geometry from those that contain structural changes, but only how to detect either kind when present. For instance, we saw above that the changing size of the image of surface items could be used in some cases to detect either kind of discontinuity, but it would not distinguish between them. Let us turn

to this issue next.

Distinguishing geometric and non-geometric texture change contours

How can texture changes contours that possibly are due solely to a change in surface geometry be distinguished from those that must involve some non-geometric, structural change? When the surface geometry changes but surface structure does not at a texture change contour, many image properties usually remain invariant: the number of different scales at which surface items occur on a surface, the approximate contrast, color, and packing factor (how tightly packed) of the items at each scale, and whether or not they are oriented. When surface structure changes at a texture change contour, everything is likely to change including the above geometric invariants. A procedure that utilizes such geometric invariants would thus seldom err in distinguishing geometric from non-geometric contours.

4. The Texture Tokens

Using image texture to infer surface information involves two broad stages. In the first stage, the basic elements that are to represent the local structure of the texture, which we shall call the *texture tokens*, are made explicit. In the second stage, the spatial variation of these tokens can be used to infer local surface orientation, surface depth, and the location of surface discontinuities, and their temporal variation may be used to infer motion correspondence. It is not presently known what constitutes the texture tokens of the first stage; this could conceivably range from grey-level values to intensity changes to primitives that represent individual texture elements and their attributes such as small blobs of a particular orientation, contrast, and size. This section explores the nature of the texture tokens and attempts to restrict this range.

Separating the effects of different surface processes

A major function the texture tokens must serve is separating the effects of different surface processes in an image. As Section 2 stated, surface structure is often due to different physical processes operating on a surface, each at it own scale. Items generated by a given process on that surface will often be similar to one another in attributes such as size, shape, orientation, color, and contrast. The spatial variation of the projection of these items in an image can provide information about the structure and 3-D geometry of the surface on which the items reside; for instance, a discontinuity in the orientation of similar items in an image can signal a discontinuity in surface geometry or surface structure (see Section 3). To utilize this information, however, it is necessary to separate the effects of different processes, for otherwise any useful information carried by items generated by a given physical process will be obscured in an image by the effects of other processes also operating there. For example, if the common orientation of bricks in a wall is to be appreciated, then it is preferable that neither markings on those

bricks nor large spots encompassing several bricks interfere with the description of the organization of the bricks themselves.

The role of scale in separating the effects of different processes

Since different physical processes often operate at different scales on a surface, the particular scale at which an image of such a surface is examined should be a useful factor for separating the effects of the different processes operating there. For example, if Figure 4.1 is examined at very small scales, then neither a change in the distribution of grey-level values nor a change in the orientation distribution of the intensity changes can identify the boundary between the two regions that are composed of **w**'s of differing orientation, since the amount of ink per unit area is the same on each side of this boundary, and the orientation distribution of the component line segments is the also same on each side of the boundary -- 50% are horizontal and 50% are vertical. The orientation information needed to identify the boundary between the two regions is carried at a larger scale in the orientation of each **w** as a whole, and not at a smaller scale in the orientation distribution of its component line segments.

The intensity changes at a particular scale can be made explicit using a method developed by Marr & Hildreth [1980]. In their theory of edge detection, they propose that an intensity change in an image I(x,y) at a particular scale can be found by (in effect) first smoothing the image with a Gaussian filter G of the desired bandwidth, and then applying the Laplacian operator ∇^2 to the smoothed image. The loci of zero-crossings in $\nabla^2(G * I) = \nabla^2 G * I$ define the location of intensity changes at that scale. Figure 4.2 shows the zero-crossings in the convolution of Figure 4.1 with a $\nabla^2 G$ operator having an excitatory region of width about the same as the width of the **w**'s. Note that at this scale, the approximate boundaries defined by the individual **w**'s comprise the zero-crossings. Thus, the predominant local orientation of the zero-crossings is the same as the local orientation of the **w**'s, and the significant change in their orientation at the boundary between the two regions in Figure 4.1 could be used to make that

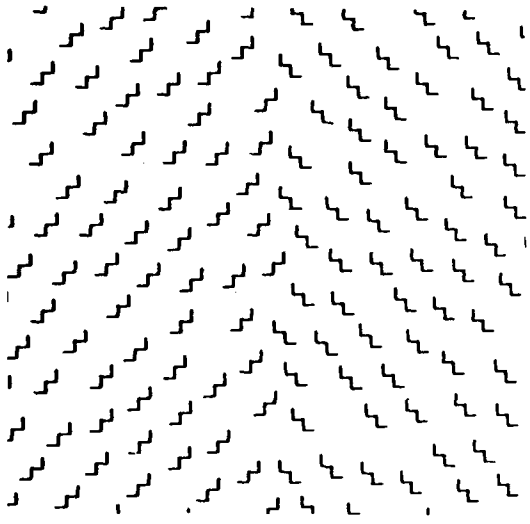


Figure 4.1 The orientation distribution of the component line segments is the same in both the left and right regions of this figure -- 50% of the line segments are horizontal and 50% are vertical. It is the changing orientation of the individual w's and not their component line segments that defines the texture boundary.

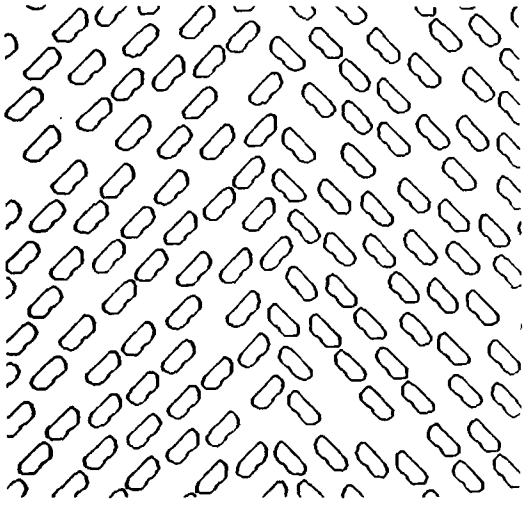


Figure 4.2 The zero-crossings of Figure 4.1 when convolved with a $\nabla^2 G$ operator having an excitatory region with width about the same as the width of the w's. Since the zero-crossings at this scale make explicit the rough boundary defined by each w, the local predominant orientation of the zero-crossings will match the orientation defined by the individual w's, and will change significantly at the texture boundary.

boundary explicit. Thus, we see that if this image is examined at the appropriate scale, the effects of the process that determines the orientation of each w can be separated from those smaller scale processes that determine its component line segment structure, and in this case the intensity changes at that larger scale are sufficient to separate the approximate boundaries of the w's from their internal structure.

The ∇^2G operator can also be used in certain cases to find intensity changes that are coincident with the texture boundary itself. Figure 6.8, consisting of convolutions of a 90° change in orientation of small line segments shows, however, that there need not be any significant intensity changes present there. In fact, we should not expect any to be there unless the average intensity changes between the textured regions on each side of the texture boundary.

The raw intensity changes are not always sufficient for separating the effects of different processes

In view of Figure 4.2, it would be tempting to think that the ∇^2G zero-crossings at various scales may be sufficient as the set of texture tokens. There are, however, physical reasons that we should not expect this to be so. The intensity changes at a given scale will not solely correspond to structural items at a particular scale, but will be affected to some degree by items at all scales and their affect will vary with the contrast of these items. In the brick wall example, high contrast markings on the bricks could noticeably influence the zero-crossing description at the scale of the bricks themselves -- something that was earlier considered undesirable for the description produced by the texture tokens. To show that this affect indeed occurs, a technique devised by Stevens [1981b] was used to create Figure 4.3. This figure is composed many small 2x2 black and white checkerboards. Stevens reasoned that if such small checkerboards appeared on a background of grey that is the psychophysical average of the black and white, then the output of any smooth convolution operator that encompasses several of these checkerboards will not differ significantly from that operator's output when encompassing just the grey background. The idea behind this particular figure is that although each collinear triple defines an

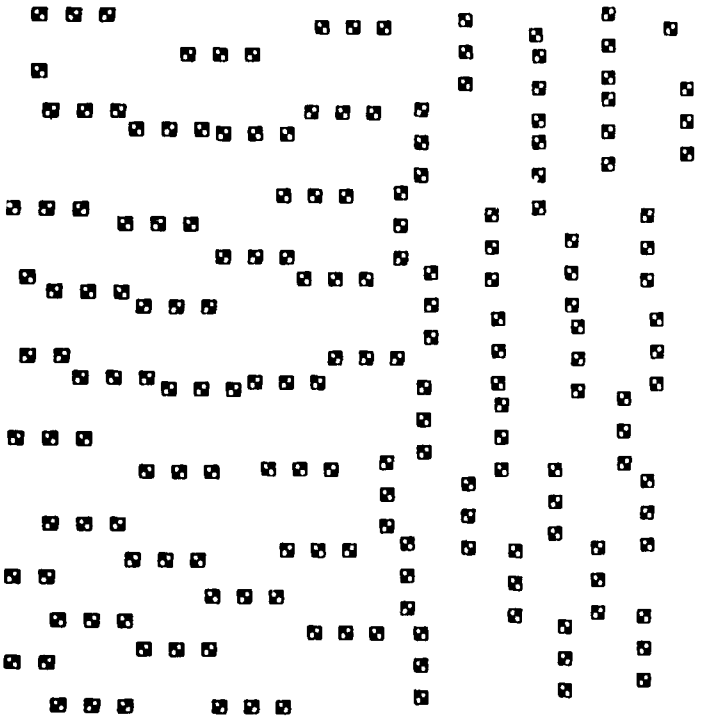


Figure 4.3 The texture elements in this figure consist of collinear triples of 2x2 checkerboards, which oriented horizontally in left region and vertically in the right region. When this figure is provided with the matching grey background, there is no scale at which a significant change occurs in the orientation distribution of the $\nabla^2 G$ zero-crossings at the texture boundary between the two regions, as there was for Figure 4.1.

oriented element and the 90° change in orientation of the triples defines a texture boundary, there will be no scale at which the distribution of the intensity changes can be used to identify this texture boundary, when this figure is provided with the matching grey background. Figure 4.4 gives the $\nabla^2 G$ zero-crossings for Figure 4.3 near the texture boundary. At the smallest scales of the $\nabla^2 G$ operator, the edges of the component squares of checkerboards are tracked by the zero-crossings. At the largest scales, as expected, the zero-crossings are of low amplitude (amplitude is not depicted in these figures) and seem to meander randomly. At intermediate scales, parts of the rough boundary defined by each collinear triple appear in the zero-crossings, but many zero-crossings corresponding to the each triple's internal structure also appear. But at no scale is the boundary of the triples made explicit and their internal structure filtered out as was possible for the w's above, making extraction of the triples' orientation and the texture boundary non-trivial. In Section 7, we shall see that the human observer can rapidly detect a boundary created by an orientation change of such checkerboard triples.

To reinforce this idea that the raw intensity changes cannot always separate the effects of different processes, a second example will be given. The previous example showed that the substructure of an item can influence the intensity changes at large enough scales to leave that item only implicit in the intensity changes. The second example again uses items at two different scales, but this time, the smaller items are not components of the larger items, but instead are independent of them. Figure 4.5 consists of line segments of two different lengths. The shorter line segments are oriented at 45° on the left-hand side of the figure and at -45° on the right-hand side, while the longer line segments are randomly oriented across the figure. Without the longer line segments, there would be a sharp orientation change in the zero-crossings at the scales that capture the smaller line segments. The randomly oriented, longer line segments, by adding noise to the local orientation distributions, weaken this sharp change in the zero-crossings. Thus, we again have an example where items from one process interfere with the intensity

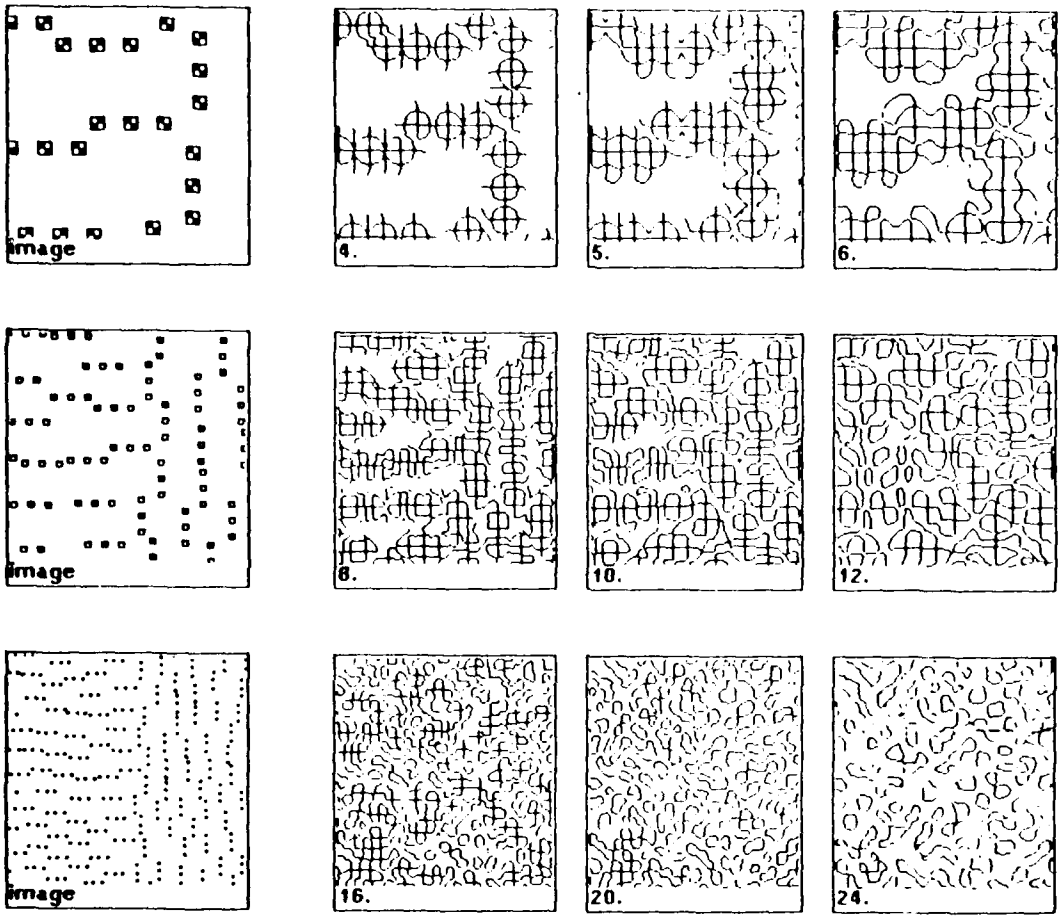


Figure 4.4 The zero-crossings of portions of Figure 4.3 (when given the matching grey background) near the texture boundary when convolved with $\nabla^2 G$ operators of various sizes. The left-most figure of each row depicts the area of Figure 4.3 used to produce the zero-crossings in that row. The number adjacent to each figure gives the diameter of the excitatory region of the $\nabla^2 G$ operator used to produce the zero-crossings in that figure, where each 2x2 checkerboard is 2x2 units in size. At no scale is the boundary of each checkerboard triple explicit in the zero-crossings and its internal structure filtered out. Further, there is no scale at which the boundary between the two regions of different checkerboard triple orientation is demarked explicitly by a zero-crossing contour, nor is there a significant change in the local orientation distribution of the zero-crossings at the texture boundary.

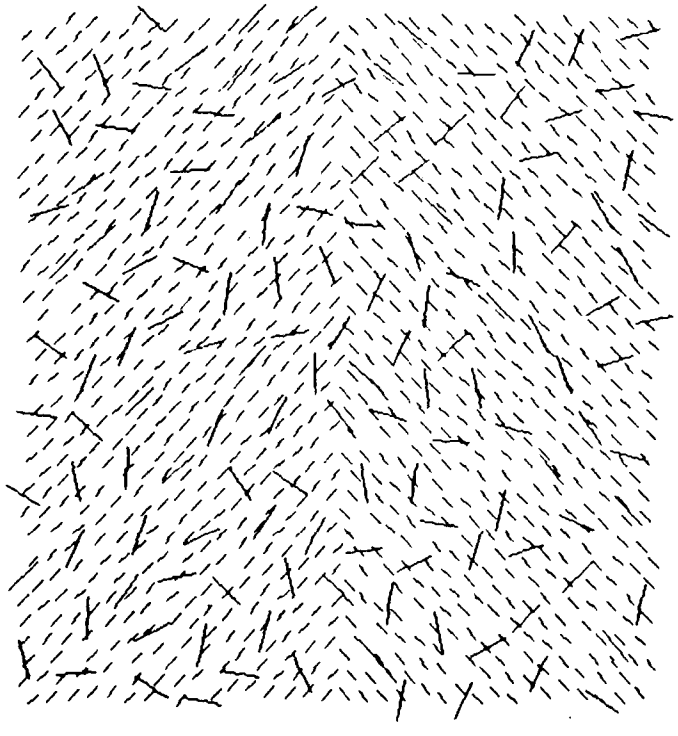


Figure 4.5 The shorter line segments in this figure are oriented at 45° on the left-hand side and at -45° on the right-hand side, while the longer line segments are randomly oriented across the figure. Without the longer line segments, there would be a sharp orientation change in the zero-crossings at the scales that capture the smaller line segments. The longer line segments weaken this change in the zero crossings.

changes that best capture those items from a different process. The information necessary to separate these two kinds of items is clearly present in this image, however; it is contained in the differing lengths of the individual line segments themselves.

What are the texture tokens?

We have seen above that the raw intensity changes appear to be too primitive a description of image texture to suffice as the sole texture tokens. In the above two examples, it is groupings, not individual points, of the intensity changes that correspond to the items that produce the texture boundary -- the oriented triples in the first example and the short line segments in the second example. This suggests that some form of local grouping of the intensity changes that results in tokens that roughly correspond to individual line segments, small blobs, local clusters and collinear groupings of these could provide a description of the local structure of image texture that better separates the items produced by different physical processes. Marr [1976] has proposed that much local image structure can be made explicit by assigning *place tokens* to such items as terminations, small blobs and line segments, which are presumably found from the intensity changes, and then by grouping these tokens to find collinear groupings and local clusters, which are then also assigned places tokens. These tokens would correspond to small markings, scratches, surface elements and local groupings of these on physical surfaces. It is not presently clear whether the early representation of texture requires tokens that faithfully and precisely represent these kinds of items everywhere in an image. Perhaps some computationally less expensive processing that roughly identifies a sizable fraction of such items would suffice at this stage, with a more precise description available with scrutiny if needed.

Exactly what the texture tokens are thus remains an open question. It solution is important not only for understanding how to detect texture boundaries, which has been emphasized here, but also for depth from texture and motion correspondence. The texture tokens could provide the unforeshortened

measure needed to obtain depth from texture, as discussed in the introduction. Further, the texture tokens, like the texture edge, would represent larger scale and rarer primitives for motion correspondence that have fewer candidate matches over a given range than the intensity changes. But being more precise about these processes must await the determination of the texture tokens, and not much can be said definitely about the form of the texture tokens at this point other than it appears that the intensity changes alone will not suffice.

5. Summary of the Theory

Three physical constraints on surface structure...

(1) The visible world can be regarded as being composed of smooth surfaces having reflectance functions whose spatial variation may be complex.

(2) Physically different processes operate on a surface to form different kinds of items there.

(3) Surface items generated by the same processes tend to be more similar to one another in their size, shape, lightness, color, and spatial arrangement than to surface items generated by other processes.

...combined with the goal of producing the *2½-D sketch*, a viewer-centered representation of the visible surfaces where the factors that produce an image -- surface geometry, surface reflectance, illumination, and viewpoint -- are separated, lead to the following conclusions for the representation of the image texture:

(1) A *texture edge* primitive is needed to identify texture change contours, which are formed by an abrupt change in the 2-D organization of similar items in an image. The texture edge can be used for locating discontinuities in surface structure and surface geometry and for establishing motion correspondence.

(2) Abrupt changes in attributes that vary with changing surface geometry -- orientation, density, length, and width -- should be used to identify discontinuties in surface geometry and surface structure.

(3) *Texture tokens* are needed to separate the effects of different physical processes operating on a surface. They represent the local structure of the image texture. Their spatial variation can be used in the detection of texture discontinuities and texture gradients, and their temporal variation may be used for establishing motion correspondence. What precisely constitutes the texture tokens is unknown; it appears, however, that the intensity changes alone will not suffice, but local groupings

of them may.

(4) The above primitives need to be assigned rapidly over a large range in an image.

6. Texture Edge Demonstrations

The primary purpose of this section is to present psychophysical evidence that texture edges are detected by the human visual system and that they are represented over a large range in an image. The secondary purpose is to characterize those types of texture changes that can give rise to perceived texture edges.

Texture discrimination and texture edges

Most previous psychophysical studies of visual texture have concentrated on their discrimination (e.g. Julesz [1973,1981] and Beck [1966]). For example, in Figure 6.1 we can immediately see without scrutiny that the lower left region of the textured pattern is different from the rest of the pattern; we can discriminate the regions. In Figure 6.2, the textured pattern looks homogeneous without scrutiny even though the upper right corner is composed of backward **R**'s, while the remainder of the pattern is composed of forward **R**'s [Julesz 1973]. In this case, we cannot discriminate the regions. Several theories have been advanced to explain why some textures are discriminable while others are not, with Julesz's second-order statistic conjecture probably the best known [Julesz 1973].

The problem with applying texture *discrimination* to the texture edge problem is that texture discrimination is an "anything goes" task; the viewer may use any means at his disposal to try to discriminate the textures within the allotted time. Suppose a viewer is asked which one of four quadrants of a texture pattern is different from the others (as in Figure 6.1) and suppose that he correctly identifies that quadrant. Did he find the correct quadrant by first finding the texture boundary between the different regions, or did he instead sample four elements, one from each quadrant, and compare them? Because it is conceivable that texture discrimination can occur at

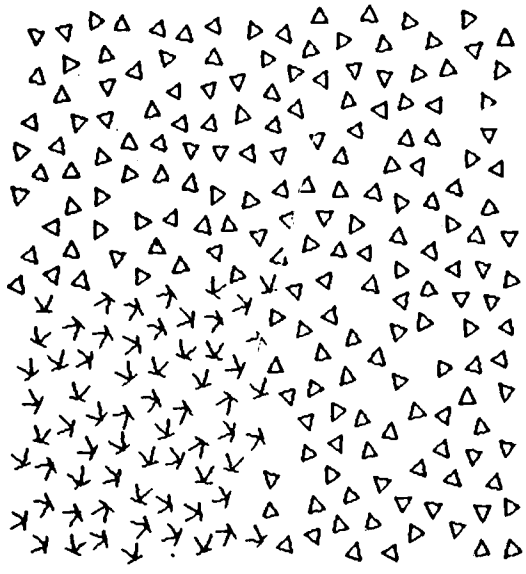


Figure 6.1 A discriminable texture. The lower left region can be seen immediately to have a different texture from the rest of the figure.

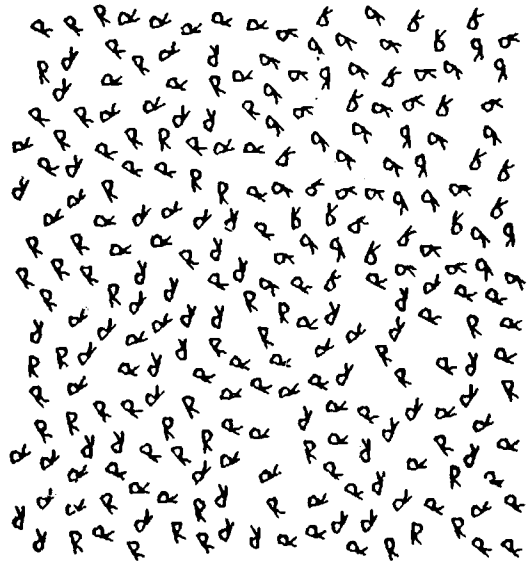


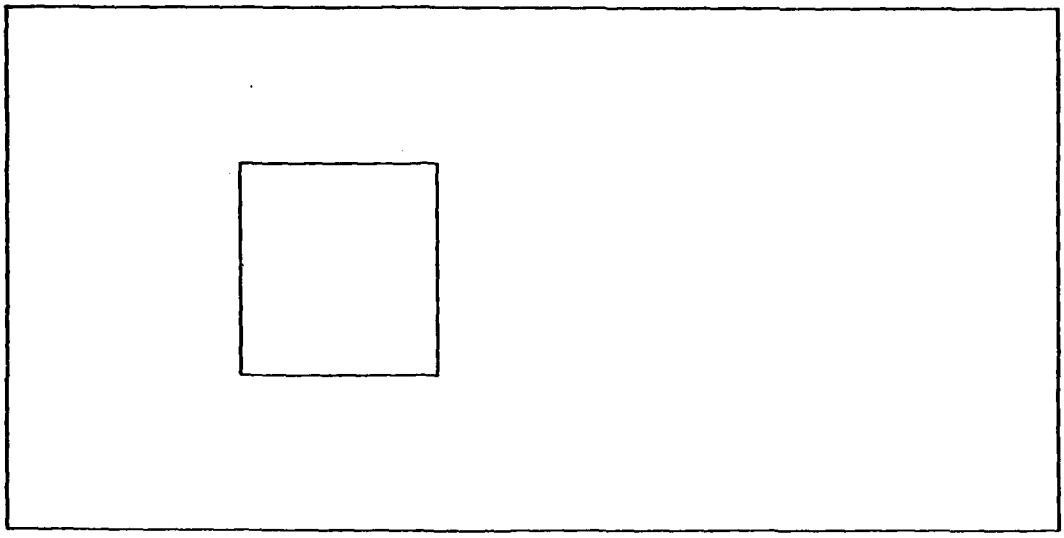
Figure 6.2 An indiscriminable texture. The figure initially appears homogeneous. Close inspection reveals that upper right region is composed of backward R's, while the remainder of figure is composed of forward R's [Julesz 1973].

least in some cases without the texture boundaries being explicitly represented, such texture discrimination studies cannot be used as evidence that texture edges are detected by the human visual system. For our purposes, these studies can only show that there are some texture differences (e.g. Figure 6.2) for which texture edges are not detected, since if they were detected, we could presumably discriminate them. But given the "anything goes" nature of the discrimination task, it can not be assumed that all discriminable textures have their boundaries explicitly represented. This means that different paradigms to study texture edges must be utilized.

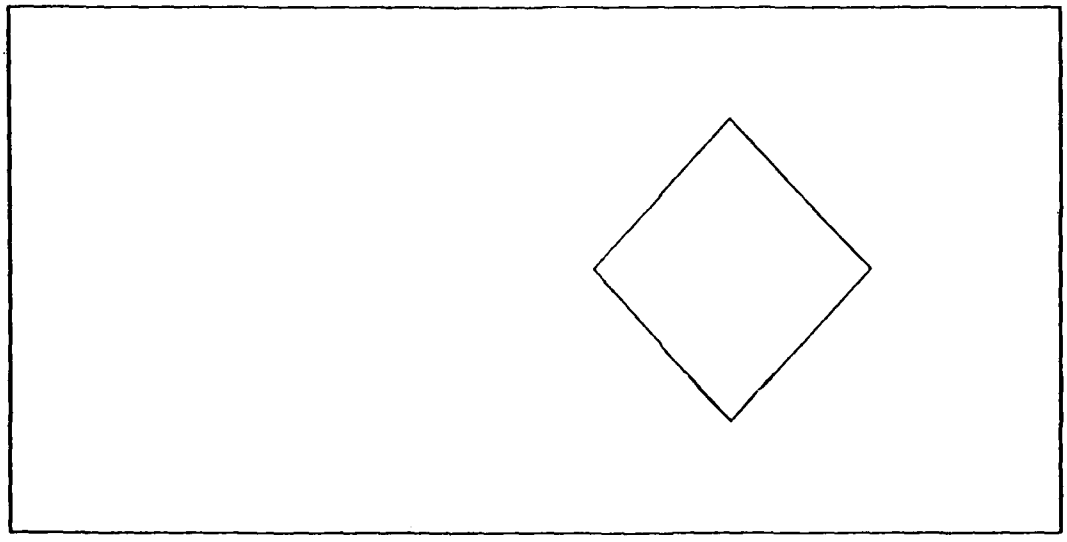
The apparent motion paradigm

It was suggested earlier that texture boundaries could be used to establish motion correspondence. We can test this hypothesis and test the human ability to perceive texture edges by using an apparent motion paradigm. It is well known that if a display sequence such as Figure 6.3a followed by Figure 6.3b is presented to a viewer with a short (say 30 msec) interstimulus interval (ISI), the viewer will perceive apparent motion -- in this case a single square will be seen to move to the right and rotate 45°. Interestingly, if the straight line sides of the square are replaced by texture edges, the correspondence can still be achieved. When the sequence in Figure 6.4 is presented, the whole pattern is seen to move to the right with the embedded square appearing to both move to the right *and* rotate 45°. Here the texture boundary is formed by a 90° orientation difference in the small line segments. Typically, an embedded square of about 5° visual angle and a presentation sequence of 300 msecs was used for each frame with an ISI of 30 msecs, but the correspondence can be achieved over a wide range of visual angle and does not depend critically on the ISI. It will be shown below that there is no intensity edge at any scale present at the boundary between the two textured regions so the correspondence must be established from the texture difference.

Ramachandran, et al [1973] have reported establishing apparent motion using a texture

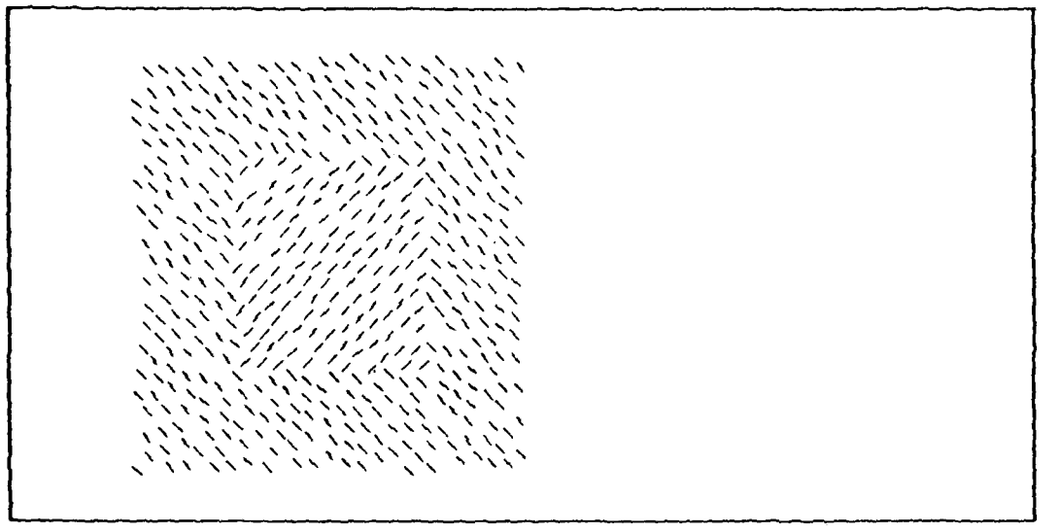


(a)

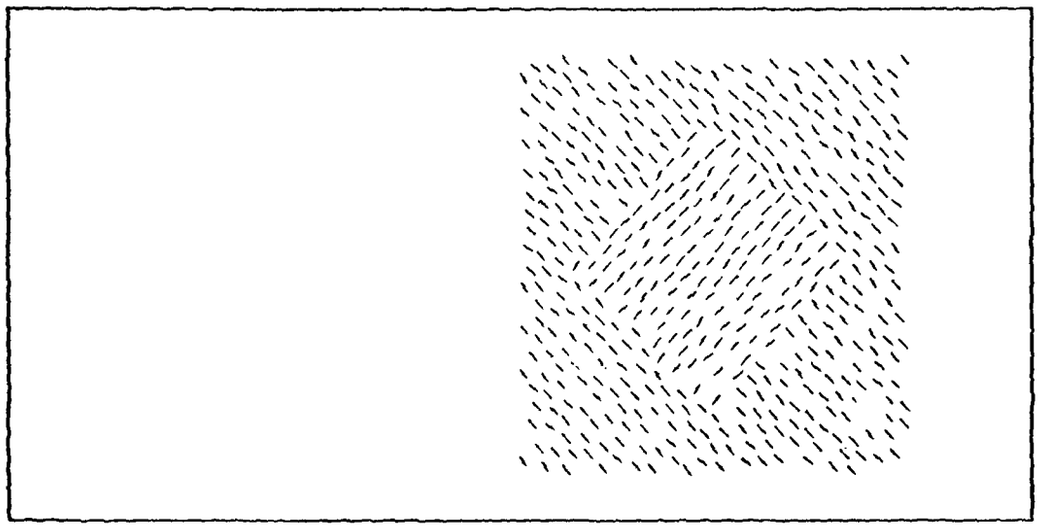


(b)

Figure 6.3 An apparent motion sequence. Display (a) is presented for 300 msec, a blank display follows for 30 msec, and then Display (b) is presented for 300msec. The viewer perceives a single square moving to the right and rotating.



(a)



(b)

Figure 6.4 Apparent motion that uses texture edges. As in Figure 6.3, Display (a) is presented for 300 msec, a blank display follows for 30 msec, and then Display (b) is presented for 300 msec. The viewer perceives the whole pattern moving to the right with the embedded square appearing both to move to the right and rotate. This apparent motion paradigm can be used to test for those texture changes that produce clearly perceived texture boundaries.

boundary with a second-order statistical difference (with equal first-order statistics). In their paradigm, an embedded square is translated but not rotated. This latter format has the disadvantage for our uses that the direction in which the embedded square of different texture is moved can be perceived even when its boundary is only weakly, if at all, perceived. By adding the rotational component to the embedded square's motion, only a clearly perceived boundary gives rise to a square that appears to both translate and rotate. The key point here is that unlike the texture discrimination tasks, it is difficult to imagine how a viewer successfully can complete this motion task without his visual system making explicit the boundary between the two regions of differing texture.

The static shape recognition paradigm

A second paradigm that involves static shape recognition can also provide evidence of human ability to perceive texture edges. If an embedded figure in a texture pattern is sufficiently complex in shape and can still be recognized without scrutiny, then it seems likely that that shape's boundary is detected by the visual system. In Figure 6.5, which uses the same texture change as in the motion example, there is little difficulty in recognizing which letter of the alphabet corresponds to the embedded shape. Thus, this gives evidence from two independent techniques -- the apparent motion paradigm and the static shape recognition paradigm -- that a particular kind of texture boundary (one formed by a 90° difference in small line segments) is detected by the visual system. Kidd, Frisby and Mayhew [1979] have found that texture boundaries can initiate vergence movements for stereopsis. This could serve as a basis for a third paradigm for studying texture boundaries, but this has not been investigated here.

An orientation difference of line segments is not the only sort of texture boundary that is successful in the apparent motion and shape recognition paradigms. Figure 6.6 shows a difference

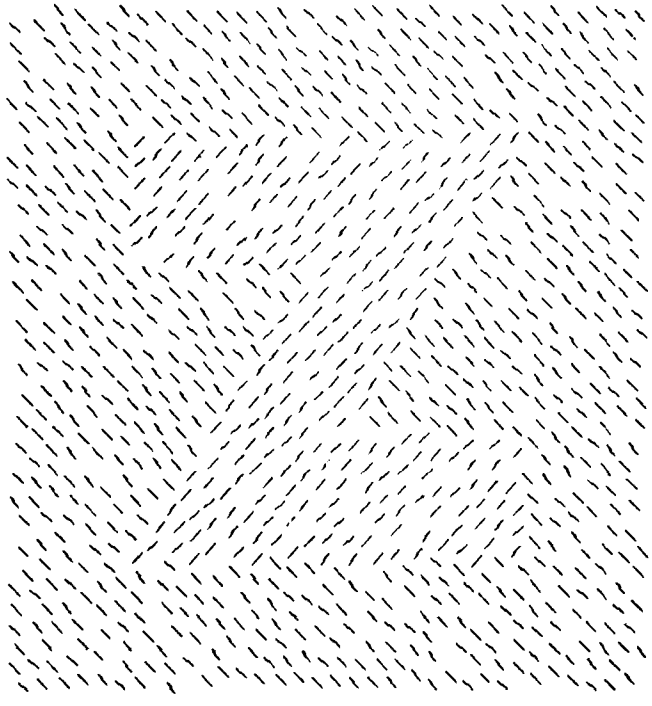


Figure 6.5 The shape of the embedded region with line segments of differing orientation can be recognized easily as the letter *Z*. This shape recognition paradigm provides a second test for those texture changes that produce clearly perceived texture boundaries.

in the dot density (4:1) that results in immediate shape recognition. In the *apparent motion* paradigm, the same texture change results in the embedded square being perceived as moving to the right and rotating. There are many sorts of texture changes that fail in both the shape recognition and motion paradigms. Figure 6.7 show several types of texture changes for which static shape recognition is difficult without scrutiny. These same texture changes do not result in the correspondence of the embedded square in the apparent motion paradigm; no embedded square is seen moving to the right *and* rotating. In particular, Figure 6.7c, which *fails* the tests for perceived texture edges, *passes* the Julesz-style test for texture discrimination (Figure 6.1). While some texture boundaries result in motion correspondence and shape recognition and others do not, in all the texture boundaries that have been tried, motion correspondence is established if and only if shape recognition is immediate. This strengthens the hypothesis that texture edges are explicitly represented by the visual system.

Texture edges are not always explicitly present in the zero-crossings

It was claimed that in Figure 6.5 there is no average intensity change at the texture boundary at any scale, and thus this boundary is not explicit in the intensity changes. This claim can be substantiated by convolving the figure with several sizes of the $\nabla^2 G$ mask of Marr and Hildreth [1980], and examining the zero-crossings in the output. As described earlier in Section 4, the zero-crossings of a $\nabla^2 G$ operator, which is the composition of a Gaussian and the Laplacian, identify the locations of the intensity changes at the scale determined by the bandwidth of the Gaussian. Figure 6.8 shows the zero-crossings in the convolutions of a portion of the texture boundary in Figure 6.5 with $\nabla^2 G$ masks of various sizes. Note at the smallest scale, the individual line segments are captured, and at the largest scale the the external boundary is captured, but at no scale is the boundary between the two regions present in the zero-crossings.

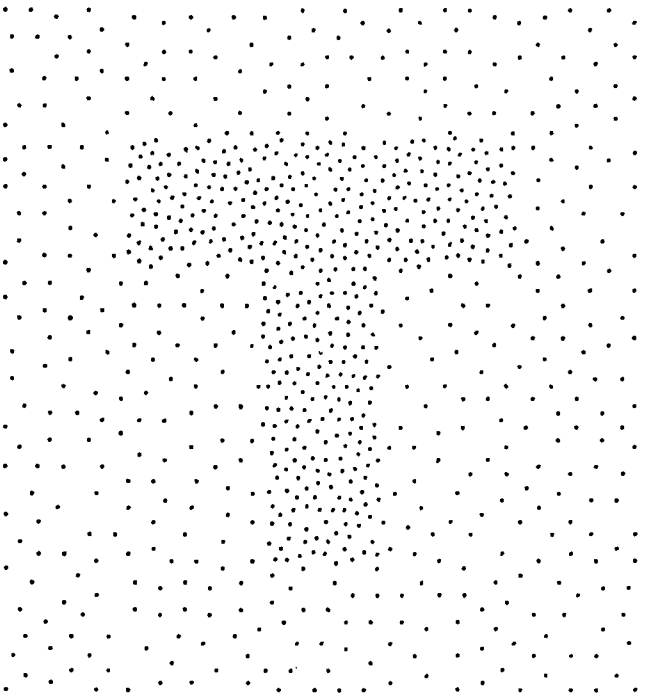


Figure 6.6 A 4:1 dot density difference can also give rise to shape recognition. The viewer can recognize immediately the shape of the embedded region of greater density as the letter T. In the density case, however, it is difficult to separate experimentally the relative influences of large scale intensity changes and changes in token density at the perceived boundary.

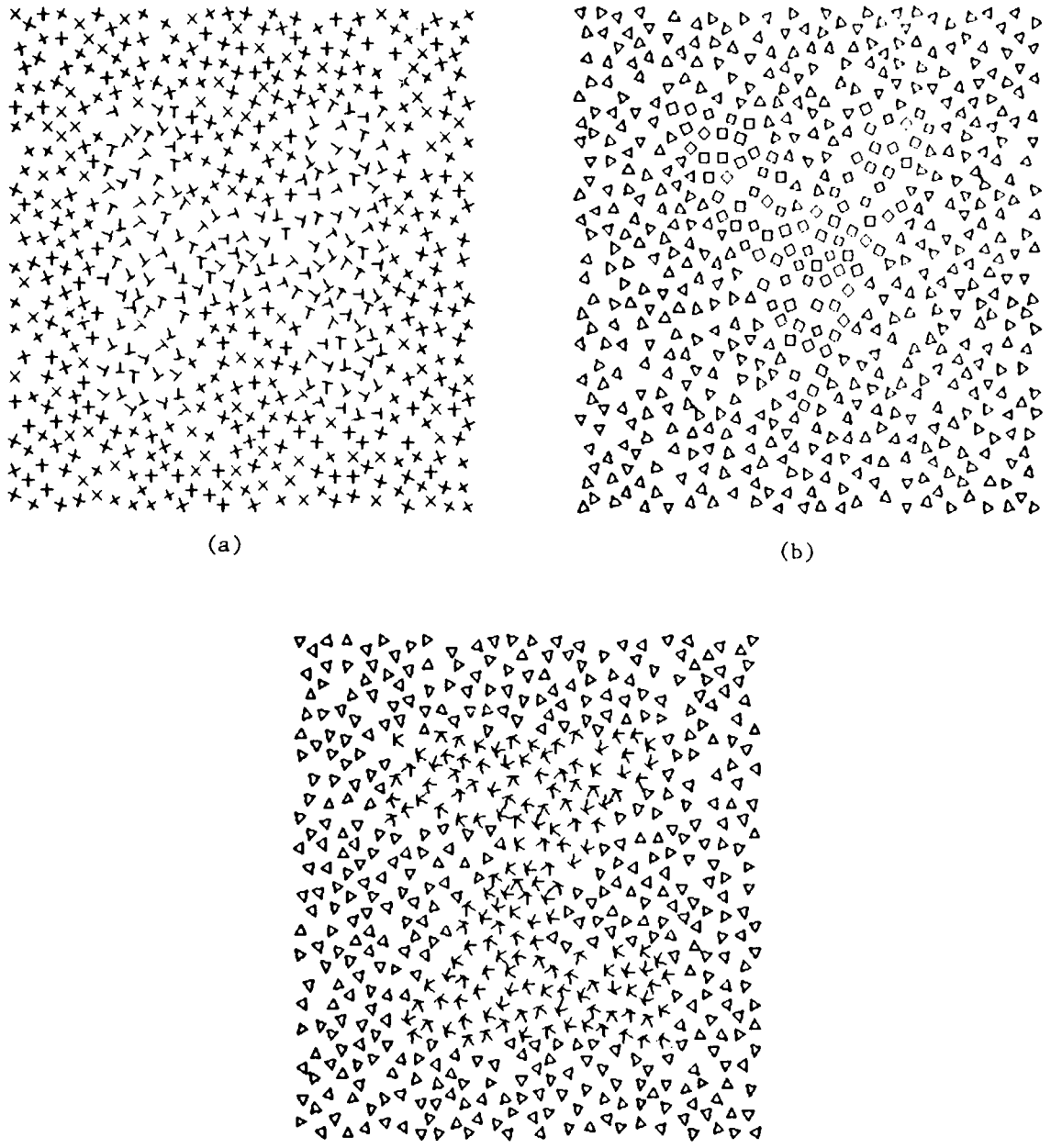


Figure 6.7 Several texture changes for which immediate shape recognition is difficult. Close examination of each pattern reveals that the embedded shape is (a) the letter H, (b) the letter V, and (c) the letter Z. Note that the texture change in (c) is the same as in the "discriminable" Figure 6.1.

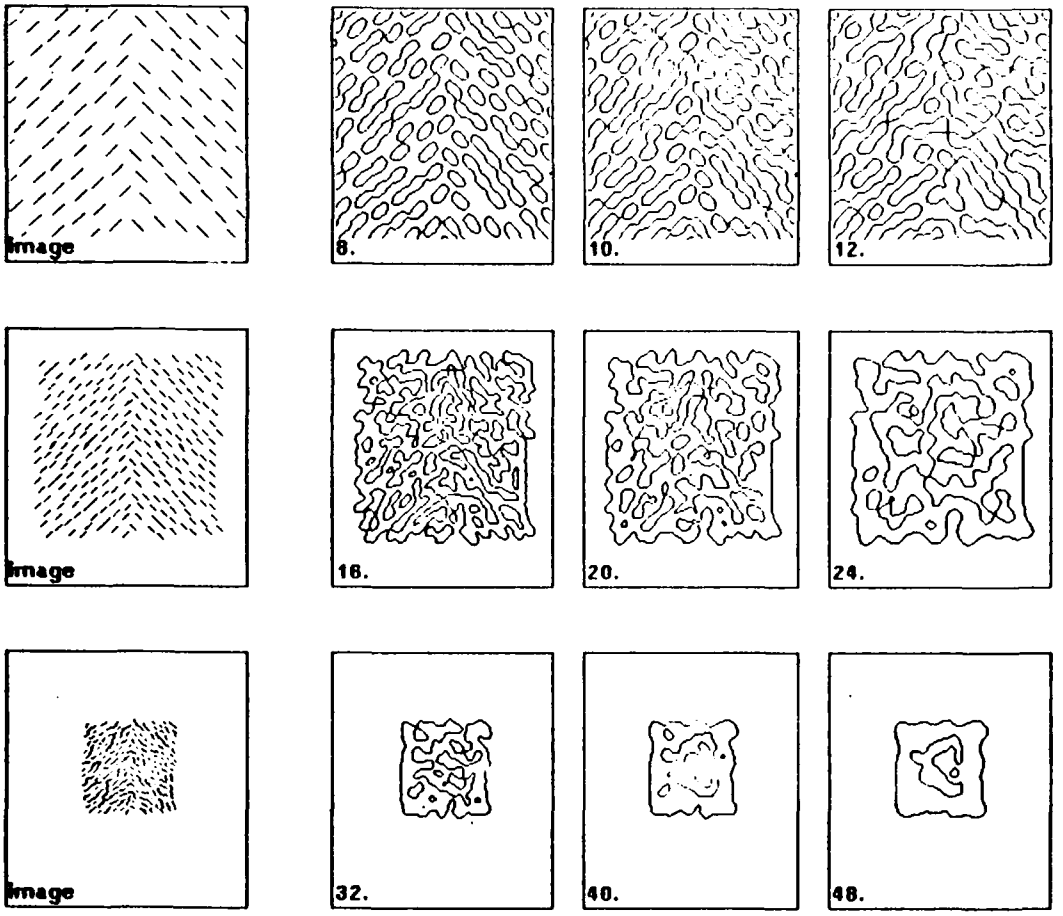


Figure 6.8 The zero-crossings for the texture change in Figure 6.5 using $\nabla^2 G$ operators of various sizes. The leftmost figure of each row depicts the image used to produce the zero-crossings in that row. The number adjacent to each figure gives the diameter of the excitatory region of the $\nabla^2 G$ operator used to produce the zero-crossings in that figure, where each line segment is 9 units long. At no scale is the boundary between the two regions of different line segment orientation explicitly demarked by a zero-crossing contour.

In Figure 6.6, there is a large scale intensity change that could be used to identify the embedded region's boundary (this easily is seen to be true by viewing the figure from far enough away that the individual dots are not resolvable -- the embedded shape can still be perceived due to the large scale intensity change). The fact that a texture boundary that is due to changing texture element density, length or width is often accompanied by a large scale intensity that coincides with the texture boundary makes it difficult to access experimentally if these texture changes result in perceived texture edges in the absence of these large scale intensity changes; further work is needed in this area. Orientation changes have been emphasized in this paper, since they are free of this complication.

Image range of the texture edge primitive

Motion correspondence and shape recognition can be achieved with these figures as large as 30-40° in visual angle; at this size, local scrutiny could reveal only a small portion of the boundary at a given time. But the motion correspondence is immediate, and shape recognition can still occur when a figure is briefly flashed (300 msec). This supports the hypothesis that many texture edges are being simultaneously found over a large portion of the image.

Characterizing those texture changes that produce perceived texture edges

A complete characterization of those texture changes that produce perceived texture edges and those that do not (as evidenced by the above apparent motion and shape recognition paradigms) has yet to emerge. A complete phenomenological characterization is difficult to obtain because there may be many attributes (e.g. contrast, color, orientation, density, length) that the visual system can use to detect texture edges, and new attributes can always be proposed that have yet to be tested psychophysically. Further, it is difficult to separate some attributes experimentally, such as texture

element density from average local intensity, as discussed above. Nevertheless, two rules seem to characterize many of those texture changes that can and cannot produce perceived texture edges.

The first rule is that *significant, abrupt changes in attributes that vary with changing surface geometry produce perceived texture edges.* This has already been shown to be the case above with the orientation of texture elements. Intensity, density, and size changes of texture elements can also produce perceived texture boundaries, but further work is needed to decouple the large scale intensity changes from the density and size changes to access each attribute's individual effect. Conversely, the textures in Figure 6.7 were generated by holding constant average local texture element density, orientation, length and width, but otherwise using different shaped texture elements across the texture boundary. Even though there are significant structural differences in the texture elements across the boundary, such as the number of terminations and corners, these changes alone do not produce perceived texture edges. In fact, texture element color and contrast are the only attributes that do not (usually) vary appreciably with changing surface geometry that have been found so far to produce perceived texture edges. This contrasts with Julesz's results for texture discrimination which indicate that changes in the number of terminations can apparently be used to discriminate textured regions [Julesz 1981]. As mentioned earlier, texture discriminability does not insure that a clear texture boundary will be perceived.

This first rule is not surprising in light of the discussion in Section 3 on the uses of texture edges. Reiterating what was said there, texture edges can identify discontinuities in surface geometry and surface structure. At a texture discontinuity where surface geometry changes but surface structure does not, it will be those image attributes that vary with surface geometry -- e.g. orientation, density, length, width -- that can be used to identify the discontinuity in the image. At a discontinuity where surface structure changes, everything is likely to change -- orientation, density

color, contrast, size. Further, the presence or absence of geometric invariants such as similarly oriented items at a given scale that remain oriented across a texture boundary can be used to distinguish between these two kinds of discontinuities. Thus, while structural attributes such as number of terminations and corners could help detect changes in surface structure when geometric attributes such as orientation, density, and size, all happen to be constant across a texture discontinuity, the visual system could consider such an occurrence too unlikely in natural images to justify its detection.

The second rule is that *the comparison of distributions of a given attribute of otherwise similar texture elements is kept simple.* This rule is detailed here only for the orientation attribute. Figure 6.9 shows that the oriented line segments at two fixed orientations (45° and -45°) found inside the embedded Z-shaped region are sufficient to match the randomly oriented line segments found outside the embedded region -- the embedded letter is difficult to recognize quickly. Likewise, the same texture change does not produce motion correspondence in the apparent motion paradigm. This suggests that the visual system may assume that the orientation distribution of items at a given scale either clusters around a single value or is, for all intents and purposes, random. A process that naturally produces, say, a distinct, two-peaked orientation distribution (as 45° and -45°) of otherwise identical items would be deemed too rare to be worth distinguishing from a random distribution. Incidentally, this contrasts with previous work by the author using texture discrimination instead, for which three orientations were found necessary to match random orientations [Riley 1977].

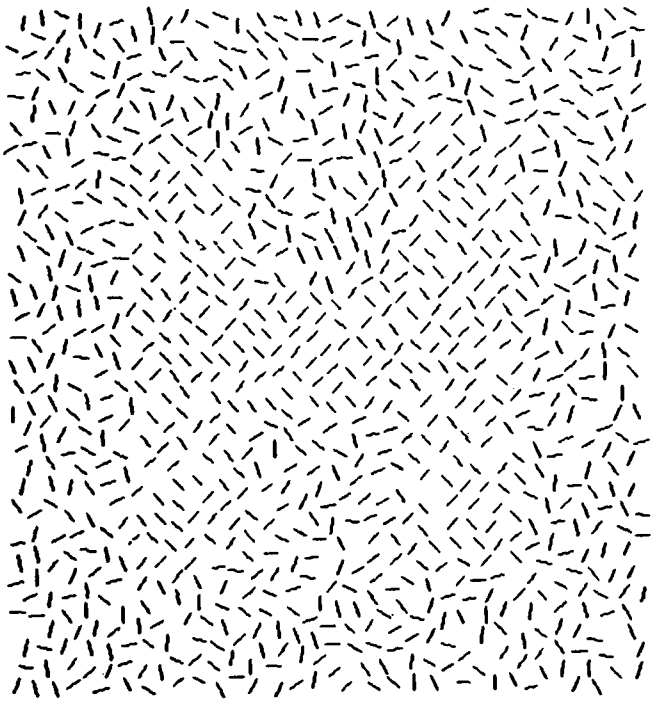


Figure 6.9 Two fixed orientations (45° and -45°) of the line segments inside the embedded region match the random orientations of the line segments outside the embedded region; the embedded shape is difficult to recognize initially as the letter **H**.

7. Texture Token Demonstrations

In this section, psychophysical demonstrations are presented that the elementary tokens that the human visual system uses to represent the local structure in image texture do not consist solely of the raw intensity changes at a variety of scales in an image. Specifically, demonstrations will be given that there are no significant changes in the orientation distribution of the $\nabla^2 G$ zero-crossings at any scale that can be used to detect some texture boundaries that humans can readily perceive. Two different approaches are taken to create these demonstrations.

The Checkerboard Paradigm

The first approach utilizes the checkerboard technique described in Section 4. The general idea is to use small black and white checkerboards as component items in larger scale groupings so that the larger scale groupings will not be explicit in the larger scale intensity changes due to the integrating effects of the $\nabla^2 G$ convolution operator. In particular, each dot in Figure 7.1 can be replaced by a small 2x2 black and white checkerboard and the entire figure given the matching grey background that is the psychophysical average of the black and white (see Figure 4.3). This match is achieved by viewing the checkerboards from sufficiently far away and adjusting the background grey until the checkerboards disappear. Under these conditions, the embedded letter, which can easily be perceived in the unmodified Figure 7.1, can still be immediately recognized in the so modified figure, provided the figure is viewed from sufficiently close in (otherwise, if the viewer moves back from the figure, the checkerboards eventually begin to disappear, with those toward the periphery being affected first).

The previous section argues that the above is evidence that a texture change consisting of a large change in the orientation of collinear triples of tiny 2x2 checkerboards can be identified by the

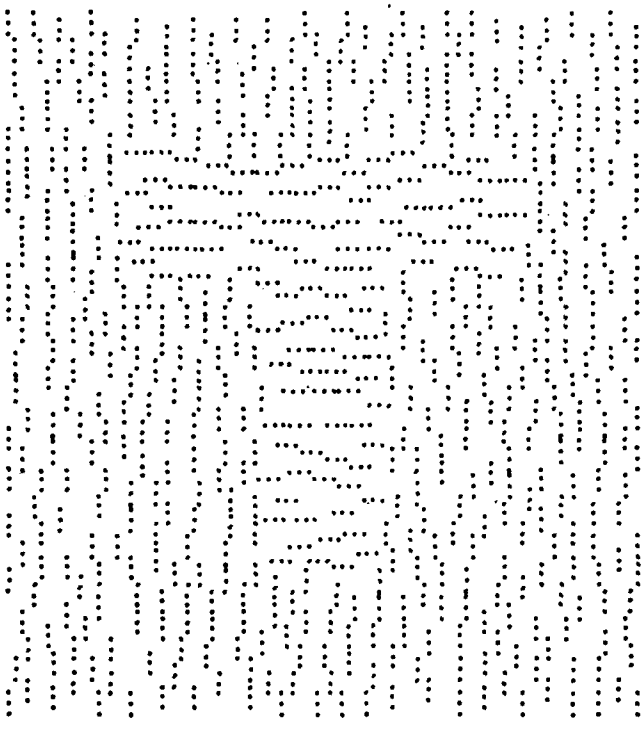
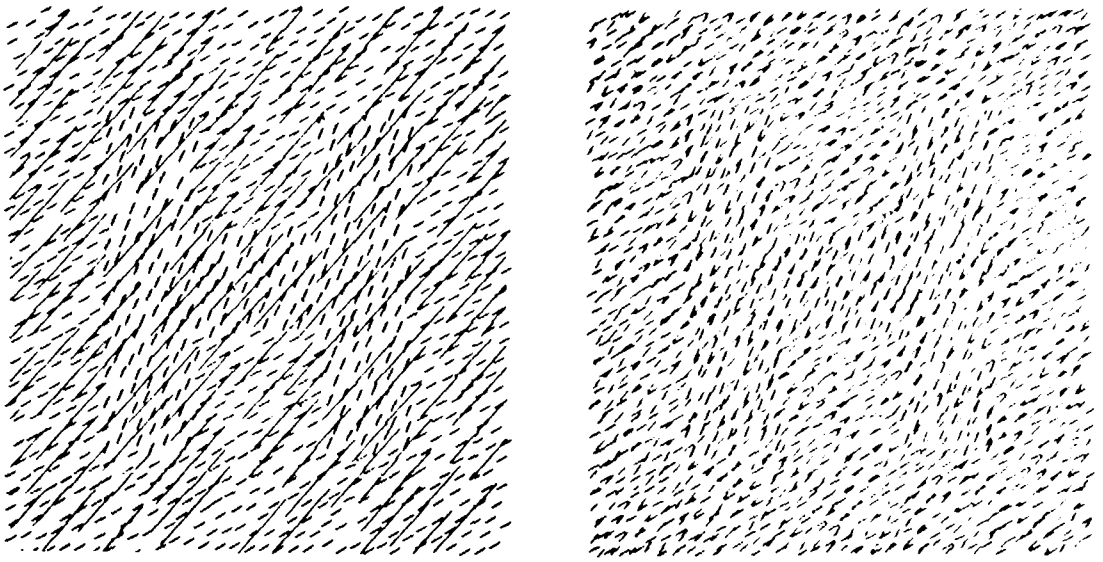


Figure 7.1 When the dots in this figure are replaced by small 2x2 black and white checkerboards and the entire figure is given the matching grey background that is the psychophysical average of the black and white, the embedded shape can still be recognized as the letter T. Figure 3.4 showed that at no scale is the boundary between the two regions of different checkerboard triple orientation explicitly demarked by a zero-crossing contour, nor is there a significant change in the local orientation distribution of the zero-crossings at the texture boundary.

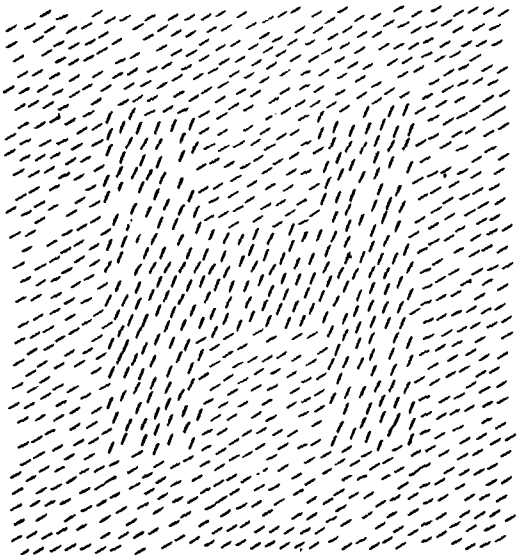
human visual system. Since any smooth spatial operator that encompasses several of these checkerboards will respond with the same output that is given to the grey background, there is no intensity change at any scale at the boundary between the two textured regions. Of crucial importance here is the fact that the orientation defined by checkerboard triples is not explicit in the intensity changes either. As shown in Section 4, the $\nabla^2 G$ zero-crossings at no scale make explicit the boundaries of individual triples while filtering out their internal structure, and thus the changing orientation of the triples at the texture boundary cannot be found by looking for a significant change there in the local orientation distributions of zero-crossings of $\nabla^2 G$ operators at some scale (see Figure 4.4).

Mixed lengths paradigm

The second approach taken to demonstrate that the raw intensity changes are not sufficient as the sole texture tokens utilizes texture elements of two different lengths. The general idea is that if one set of texture elements of a given length has, say, some oriented structure in a texture, then this oriented structure will be easier to detect in the presence of other texture elements of a very different length than in the presence of other texture elements of a similar length provided the texture elements are first separated on the basis of their length. Figure 7.2a shows a texture pattern composed of line segments of two different lengths. The shorter line segments are oriented at 65° inside the embedded H-shaped region and at 25° outside this region. The larger line segments are nine times as long as the shorter line segments and are oriented at 45° throughout the texture pattern. Figure 7.2c shows, for reference, just the shorter lines found in Figure 7.2a. Figure 7.2b contains an identical copy of the shorter line segments found in Figure 7.2a, but the larger line segments have been shrunk 1/9th in length (to the same length as the other line segments) with a corresponding nine fold increase in their density (i.e. number/area), thus keeping the total amount



(a) (b)



(c)

Figure 7.2 The creation of texture patterns (a) and (b) both begin with underlying pattern (c), which has line segments at 65° inside the embedded region and at 25° outside this region. Masking 45° line segments nine times as long as those in pattern (c) and with one ninth the density (number/area) are added to complete pattern (a). Masking 45° line segments of the same length and with the same density as pattern (c) are added to complete pattern (b). The embedded H in pattern (a) is easier to recognize than that in pattern (b); an effect that is accentuated at oblique or distant viewpoints. This result is difficult to explain if the raw intensity changes at various scales are the sole texture tokens.

of 45° contour over a given area constant. Thus, measuring the amount of contour at a given orientation per unit area taken from very local descriptions of the intensity changes found in an image of these figures would not show significant differences between Figure 7.2a and Figure 7.2b. Figure 7.3 and Figure 7.4 contain $\nabla^2 G$ zero-crossings at various scales near the embedded texture boundary of Figure 7.2a and of Figure 7.2b, respectively. They were generated to show that for no scale (operator size) is there a significant difference between the local orientation distributions of zero-crossings for Figure 7.3 and Figure 7.4 that would result in a noticeable difference between the detectability of the embedded region in Figure 7.2a and Figure 7.2b. At the smaller scales, the zero-crossings where the line segments of different orientations cross are very similar for Figure 7.2a and Figure 7.2b, and since the local amount of contour at each orientation is the same in both figures by design, the local zero-crossing distributions of the two figures at these smaller scales are quite similar. At the larger scales, the smaller line segments are not resolved; since the smaller line segments carry the orientation change that produces the texture boundary, differences in the local zero-crossing distributions of the two figures at larger scales are not relevant to the detectability of the texture boundary. Thus, if texture boundary detection were based on identifying significant changes in the distribution of zero-crossings at the boundary, the texture boundaries in Figure 7.2a and Figure 7.2b should have similar detectability. Note, however, that in Figure 7.2a, the embedded letter is easier to recognize than in Figure 7.2b, an effect that is accentuated at distant or oblique viewpoints. This suggests that the line segments are somehow first separated on the basis of their length.

This result may seem at odds with those due to Treisman [1977,1980]. She found, using a variety of techniques, that human observers were very poor at the pre-attentive selection of items having the conjunction of two or more attribute values (e.g. shape:H and color:red) in a field of distractors.

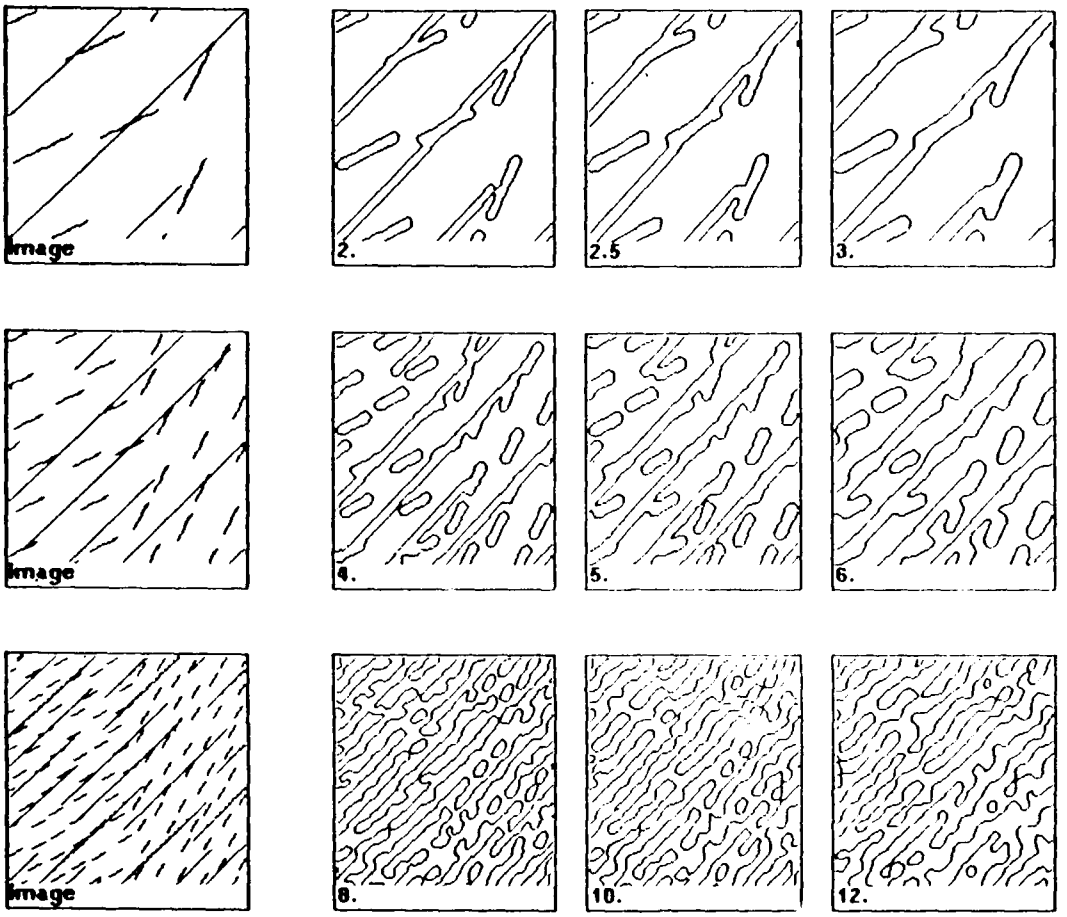


Figure 7.3 Zero-crossings for the texture change in Figure 7.2a using $\nabla^2 G$ operators of various sizes. Again, the leftmost figure of each row depicts the image used to produce the zero-crossings in that row, and the number adjacent to each figure gives the diameter of the excitatory region of the $\nabla^2 G$ operator used to produce the zero-crossings in that figure, where the shorter line segments are 9 units long. Comparison with Figure 7.4 reveals that at the smaller scales, there is no significant difference in the local orientation distribution of the zero-crossings between the two figures, while at the larger scales the smaller line segments, which contain the boundary-forming orientation change, are not resolved. Thus, the results in Figure 7.2 cannot be explained if the texture boundary is detected solely on the basis of significant changes in the local zero-crossing distribution across the boundary.

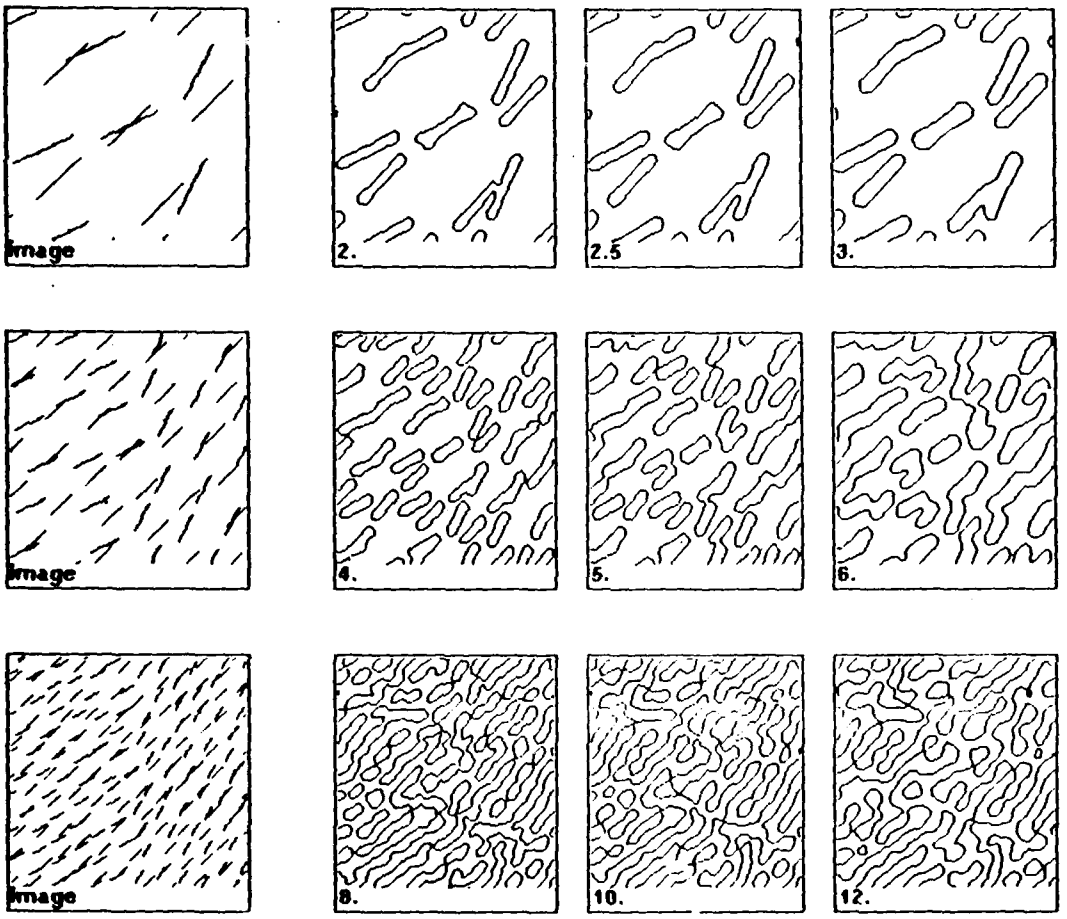


Figure 7.4 Zero-crossings for the texture change in Figure 7.2b using $\nabla^2 G$ operators of various sizes, with the same format as Figure 7.3.

In Figure 7.2, the selected attributes are orientation and scale (length of line segment). A possible explanation is that scale is indeed special as suggested earlier -- large differences in size may not be treated like other variations in attribute values, since they strongly suggest that different processes are responsible for the respective items.

8. Summary of Demonstrations

(1) Two different experimental paradigms -- one based on static shape recognition of a textured region embedded in a textured surround and one based on motion correspondence of texture boundaries -- support the hypothesis that some kinds of texture boundaries are detected by the visual system and are made explicit in a representation that covers a large range in an image.

(2) $\nabla^2 G$ zero-crossing results indicate that there are no significant intensity changes at any scale coincident with the texture boundaries in the above figures and thus the detection of these boundaries must be based on more abstract texture measures.

(3) Two rules characterize many of the texture changes that can and cannot produce perceived texture edges as evidenced by the experimental paradigms in (1):

(a) Significant, abrupt changes in texture element attributes that vary with changing surface geometry -- orientation, length, density, width -- produce perceived texture edges.

(b) The comparison of distributions of a given attribute of otherwise similar texture elements is kept simple -- e.g. two fixed orientations are sufficient to match random orientations in the texture boundary paradigms.

(4) Two different experimental paradigms -- one using oriented groupings of 2x2 checkerboards and one using line segments of two different lengths combined with $\nabla^2 G$ zero-crossing results cast doubt that the raw intensity changes at various scales would suffice as the sole texture tokens; there are no significant changes in the distribution of the $\nabla^2 G$ zero-crossings at any scale at the texture boundaries found in these demonstrations.

References

Beck, J. 1966. Effect of orientation and of shape similarity on perceptual grouping. *Perception and Psychophysics*, **1**, 300-302.

Horn, B. 1977. Understanding image intensities. *Artificial Intelligence*, **8**, 201-231.

Julesz, B. 1973. Inability of humans to discriminate between visual textures that agree in second-order statistics -- revisited. *Perception*, **2**, 391-405.

Julesz, B. 1981. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biol. Cybern.*, **41**, 131-138.

Kidd, A., Frisby, J., & Mayhew, J. 1979. Texture contours can facilitate stereopsis by initiating appropriate vergence eye movements. *Nature*, **280**, 829-832.

Marr, D. 1976. Early processing of visual information. *Phil. Trans. Roy. Soc. B.*, **275**, 483-524 .

Marr, D. 1977. Representing visual information. *AAAS 143rd Annual Meeting, Symposium on Some Mathematical Questions in Biology*, February. Also available as M.I.T. A.I. Lab. Memo 415.

Marr, D. 1981. *Vision: a computational investigation into the human representation and processing of visual information.* San Francisco: Freeman (in press).

Marr, D. & Hildreth, E. 1980. Theory of edge detection. *Proc. Roy. Soc. Lond.* B, **207**, 187-217.

Marr, D. & Nishihara, K. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Phil Trans. Roy. Soc. B.*, **200**, 269-294.

Marr, D. & Poggio, T. 1978. A theory of human stereo vision. *Proc. Roy. Soc. Lond. B.*, **204**, 301-328. Also available as M.I.T. A.I. Lab. Memo 451.

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

Ramachandran, V., Madhusudhan, V., & Vidyasagar, T. 1973. Apparent movement with subjective contours. *Vision Research*, **13**, 1399-1401.

Riley, M. 1977. Discrimination of bar textures with differing orientation and length distributions. *M.I.T. B.S. Thesis, Dept. of Elect. Eng. and Comp. Sci.*, June.

Stevens, K. 1981a. Information content of texture gradients. *Biol. Cybern.*, (in press). Also available as M.I.T. A.I. Lab. Technical Report 512.

Stevens, K. 1981b. Personal communication.

Treisman, A. 1977. Focused attention in the perception and retrieval of multidimensional stimuli. *Perception and Psychophysics*, **22**, 1-11.

Treisman, A. & Gelade, G. 1980. A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Ullman, S. 1976. On the visual detection of light sources. *Biol. Cybern.*, **21**, 205-212.

Ullman, S. 1979. *The interpretation of visual motion.* Cambridge, Ma.: M.I.T. Press.


DATE
ILME