Dan Brotsky

AI-TR-579

IMPLEMENTATION OF A THEORY OF EDGE DETECTION

ELLEN C. HILDRETH

April 1980

ARTIFICIAL INTELLIGENCE LABORATORY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

This blank page was inserted to preserve pagination.

IMPLEMENTATION OF A THEORY OF EDGE DETECTION

by

Ellen C. Hildreth

Massachusetts Institute of Technology

April, 1980

Revised version of a dissertation submitted to the Department of Electrical Engineering and Computer Science on January 3, 1980 in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

This report describes the implementation of a theory of edge detection, proposed by Marr and Hildreth (1979). According to this theory, the image is first processed independently through a set of different size filters, whose shape is the Laplacian of a Gaussian, $\nabla^2 G(x,y)$. Zero-crossings in the output of these filters mark the positions of intensity changes at different resolutions. Information about these zero-crossings is then used for deriving a full symbolic description of changes in intensity in the image, called the raw primal sketch. The theory is closely tied with early processing in the human visual system.

In this report, we first examine the critical properties of the initial filters used in the edge detection process, both from a theoretical and practical standpoint. The implementation is then used as a test bed for exploring aspects of the human visual system; in particular, acuity and hyperacuity. Finally, we present some preliminary results concerning the relationship between zero-crossings detected at different resolutions, and some observations relevant to the process by which the human visual system integrates descriptions of intensity changes obtained at different resolutions.

Thesis Supervisor: David C. Marr

Title: Professor of Psychology

ACKNOWLEDGEMENTS

I would like to give special thanks to my supervisor, David Marr, whose constant and generous support and guidance have been invaluable.

Eric Grimson, Shimon Ullman, and Chuck Rich have always provided help and encouragment when things became difficult.

I thank Whitman Richards, Berthold Horn, Keith Nishihara, Tomaso Poggio, Brian Schunck, and Matt Mason for their interest and many useful comments.

TABLE OF CONTENTS

Introduction	5
I. The Raw Primal Sketch	9
II. Computing the Single Channel Description	28
II.1 The Initial Filters	29
II.2 The Initial Assumptions	53
II.3 The Final Channel Description	56
III. Performance of the First Stage by the Human Visual System	62
IV. Combining the Channels	75
IV.1 Relationships Between the Channel Outputs	75
IV.2 Some Perceptual Experiments	83
IV.3 From the Raw Primal Sketch	89
V. Other Approaches to Edge Detection	103
VI. Summary	112

Introduction

Our study of early vision is a study of the first computations which are performed by a vision system in the analysis of an image. Ultimately, we are interested in understanding the way in which the human system begins to process visual information, but whether we are analyzing a biological vision system, or a machine vision system, we can view the system as an information processor, performing computations on an incoming array of intensity values, and there is a common set of problems each system is trying to solve (Marr 1976a, 1978). Our objective in studying early vision is to uncover those computations which must be common to any general vision processor, and consider possible algorithms for describing how these computations might take place. We can then turn to the particular question of how the human system performs these computations.

This report describes the implementation of a particular theory of early visual processing (Marr & Hildreth 1979), which grew from a set of ideas proposed by Marr (1976b). The theory argues that the first goal in analyzing an image is to describe, locally, the significant intensity changes. A change in intensity will, in general, be the result of a physical change in some property of a surface, such as reflectance, a change in illumination, or a discontinuity in the depth or orientation of a surface, such as that which occurs along the boundary between two objects separated in depth. In later visual processing, it will be necessary to make the nature of these physical changes explicit, but we begin by describing properties of the intensity changes to which they give rise. The properties which can be computed for a given intensity change are (1) two-dimensional orientation in the image, (2) contrast, the amount by which intensity has changed, (3) width, the distance across which the intensity is changing, and (3) length, defined locally to be the distance, along the orientation of the intensity change, over which other

properties remain roughly uniform. This first description of the image is called the raw primal sketch (Marr 1976b).

Since the primal sketch was first proposed, there has been extensive progress toward understanding the computations involved in this and other aspects of early vision, such as stereopsis and the analysis of motion (Marr & Poggio 1979, Grimson & Marr 1979, Ullman 1979, Marr & Ullman 1979). Stereopsis refers to the computation of depth information by the comparison of the relative positions of elements in two images of a scene, taken from slightly different viewing angles. The first step in this process is to compute a correspondence between elements in the two images. Marr and Ullman (1979) divide the tasks involved in the early analysis of motion into two classes: separation tasks, which require the instantaneous measurement of the position and velocity of elements in the image, such as the detection of sudden motion; and integration tasks, which require the integration of this information over time, such as in the recovery of structure and three-dimensional motion from an orthographic projection (Ullman 1979). A range of possible image representations could potentially form the input stage to the above tasks: from the initial intensity values, to a description of changes in intensity, to a high-level description of objects in the scene. Julesz' random dot stereograms (Julesz 1971) and Ullman's many motion demonstrations, such as the broken wheel (Ullman 1979, p. 22), illustrate that these tasks need not be high-level operations. An additional demonstration by Ullman (1979, p. 17), together with the requirement of Marr and Poggio (1979) that the basic elements for stereo matching represent unique physical locations, and Marr and Ullman's (1979) suggestion that the early detection of motion be combined with the analysis of contours, argue against the use of grey levels in these computations. Thus a primitive description of the location of changes in intensity remains the likely candidate for the input to these

secondary processes. The stereo matching computation, and detection of direction of motion are low-level operations, requiring only a description of the position and sign of intensity changes, at a set of different image resolutions. The recovery of structure from motion begins also with the computation of a correspondence between elements, now between images changing in time, but its input differs from the input to stereopsis and the detection of motion in at least two ways: the basic elements may also be primitive groupings of similar intensity changes, and the explicit properties of contrast, orientation, and size of elements are used in computing the correspondence. Understanding these requirements of stereo and motion analysis has helped to formalize what information should be made explicit in the raw primal sketch, and how the information should be represented. They now offer strong support of its basic goal of describing changes in intensity.

Precise quantitative studies of the operators used in early human vision (Wilson & Giese 1977, Wilson & Bergen 1979, Schiller et al. 1976a, 1976b) suggested a number of critical issues for the design of the operators used in the computation of the primal sketch. In developing this theory of early processing, we sought not only to model the visual processing of the human system, but to understand why these computations evolved as they did.

As the theory developed, its implementation played a key role. The implementation could provide strong evidence in support of some aspects of the theory, while at the same time, could uncover other areas where the theory was incomplete, imprecise, or unsatisfactory. The theory also makes basic assumptions about the nature of intensity functions arising from the natural world; in an implementation, we can test the validity of these assumptions. Finally, once the skeleton of the theory is implemented, it can serve as an experimental medium for testing the performance of the theory against the human system which we ultimately seek to understand. The medium for the

implementation is a computer system; the initial input is a two-dimensional array of intensity values, obtained by digitizing the photograph of a physical scene. The computations which the theory proposes for detecting and describing intensity changes will then be performed on this array of intensities. The focus of this report will be this role which the implementation has played in the development of the primal sketch theory.

An overview of the theory is presented in greater detail in Section I. The theory divides naturally into two parts; the implementation of the first stage is discussed in Section II. As suggested earlier in the introduction, this first stage can be closely tied with low-level operations performed by the human visual system. Section III reviews this relationship, with particular emphasis on quantitative aspects of the human operators. The main emphasis of this work will be on this first stage of the theory. Section IV discusses some aspects of the implementation of the second stage of the theory. In this area, I feel there still remain many open questions. In Section IV.1, I present some preliminary results; Section IV.2 discusses some perceptual experiments relevant to the way in which the human system performs this stage; and finally, in Section IV.3, I stress a necessary tool for addressing some of the open questions: examination of the subsequent processes, such as the grouping of primitive edge elements, lightness related computations, and motion correspondence, which will utilize the primal sketch at some stage in its development. Similar approaches to the early analysis of images, which have influenced the development of the primal sketch, or drawn from these ideas, will be discussed in Section V. Throughout this report, I will touch on the relationship between the present theory, and studies of the human visual system, but I would like to stress that the theory can be supported by computational arguments, which apply to any general vision processor, of which the human system is a particular example.

I. The Raw Primal Sketch

This section provides an overview of the theory proposed in Marr & Hildreth (1979). The theory is presented as a theory of *edge detection*, but in the introduction, our goal for early processing was described as the construction of a description of *intensity changes*. We make a distinction between these two terms; a change in intensity is the phenomenon that we will detect and describe in an image; edges are the physical changes that give rise to these intensity changes. We will see in this section that we can detect intensity changes at different image resolutions, and that in general, there may not exist a one-to-one correspondence between intensity changes detected at a particular scale, and edges in the physical scene. However, there is an assumption we can make about physical changes which will allow us to determine when these intensity changes do, in fact, reflect edges in the physical world. For this reason, we present the primal sketch theory as a theory of edge detection.

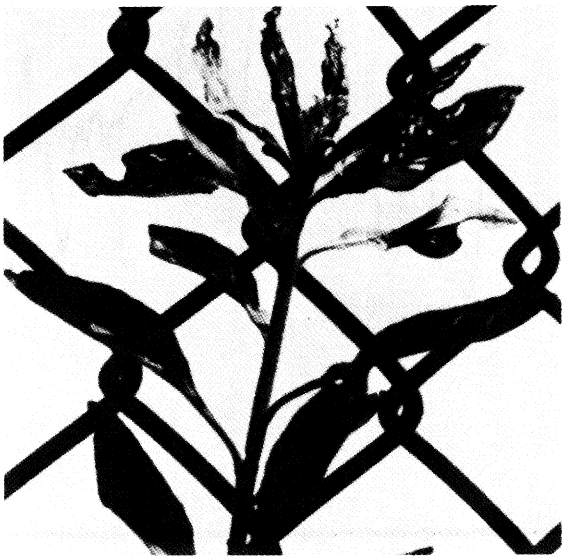
Given that our goal is to detect intensity changes in the image, we seek an operation to apply to the image which will allow us to extract these changes in a simple way. A number of considerations contributed to the design of this operator. First, changes in intensity will occur in the image at a range of different scales. If we look at individual picture elements (pixels), we find intensity changing from pixel to pixel. Often, there will be uniform changes over some distance. Most edges in the real world are sharp edges; the intensity function will be composed of a few steep changes over a small number of pixels. Other edges, such as shading edges, are very fuzzy; their corresponding intensity function will increase slowly in smaller steps over a large number of pixels. These different types of intensity change are not distinct in the image; it is common, for example, to find a slow intensity change, due to a shading effect, superimposed on a sharp, high

contrast change due to an abrupt change in the reflectance of the surface. We would like to separate changes taking place at different scales, so that we make explicit local changes, taking place over short distances, as well as gross changes in the image. That is, we first look at the intensity function at a resolution near the initial image resolution, and detect intensity changes occuring at this scale. We then smooth the intensity function over different size areas, detecting intensity changes that occur at larger and larger scales. This scheme has the implication that a formal definition of an intensity change incorporates the scale at which the change occurs. It now poses the question of what function is used in the smoothing process.

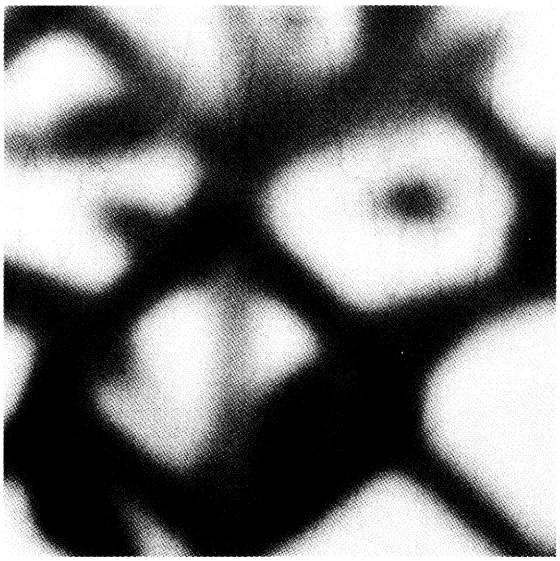
Two considerations come into play in the design of an appropriate smoothing filter. First, changes in the physical world are generally localized in space, so we desire that our initial operators also be spatially localized. Second, one of our goals for this operator is to restrict the scale at which intensity changes take place in the output of the operator; for example, to detect gross changes in intensity, we would like the frequency spectrum of the smoothed output to be localized about the low frequencies. This requirement of localization in frequency will, in general, conflict with the need for localization in space, but the two requirements can be optimally satisfied by the Gaussian distribution (Leipnik 1960); that is, the Gaussian minimizes the product of bandwidth in space and frequency. We have therefor chosen the Gaussian as our initial smoothing filter, thus beginning our detection stage by smoothing the image with a set of different size Gaussian functions. In Figure 1a, we have the image of a plant against a chain-link fence, viewed at two different resolutions. A single intensity profile appears in Figure 1b, with its counterparts from the smoothed images below it. At a small scale, we view many sharp changes, whereas at the lower resolution, only gross changes remain. We are now left with the problem of detecting

A.

1.

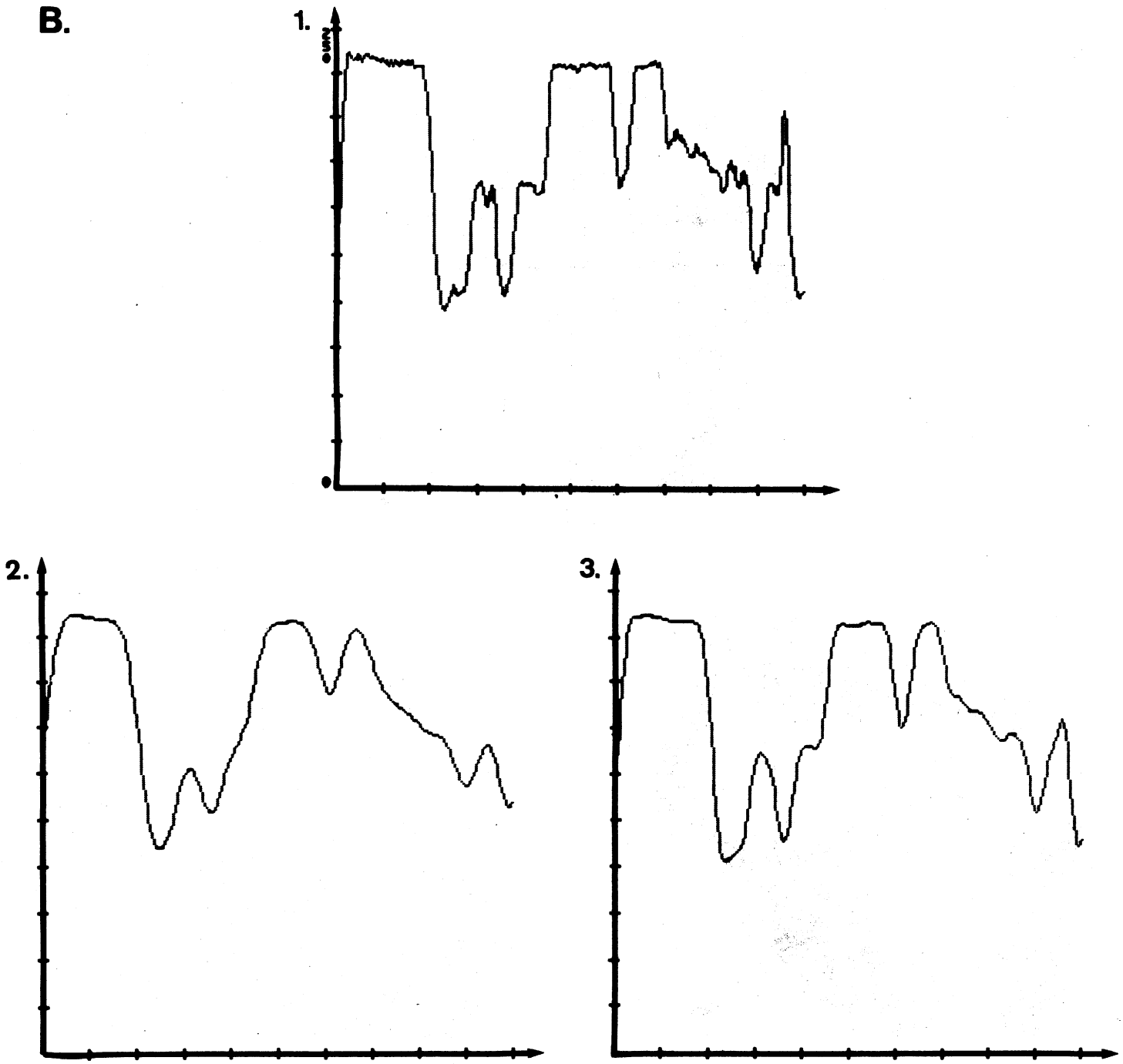


2.



3.

1a. Image smoothed with a Gaussian filter. The first stage in the primal sketch computation can be thought of as decomposing the original image into a set of images, each smoothed with a different size Gaussian filter, and detecting the intensity changes separately in each. The original image appears in (1); (2) shows the image filtered with a Gaussian having σ = 8 picture elements, and in (3), σ = 4. The image is 320 by 320 picture elements.



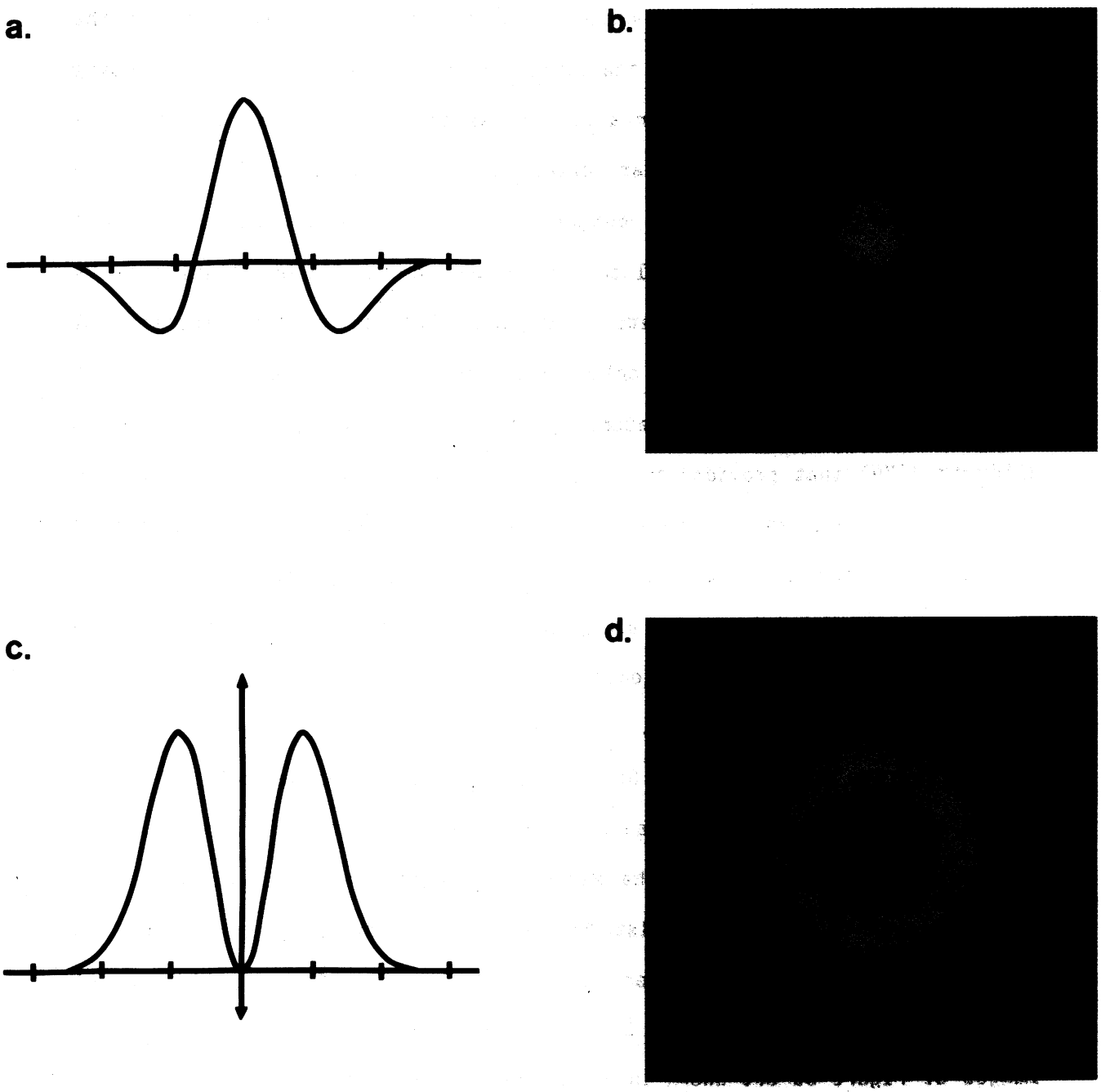
1b. Intensity profiles extracted from the original image, and filtered images of Figure 1a. (1) illustrates the image profile; (2) and (3) are extracted from the smoothed images with $\sigma = 8$ and $\sigma = 4$, respectively.

these changes.

If an intensity change occurs along a particular orientation in the image, there will be a peak in the first directional derivative of intensity measured across the change, or a zero-crossing in the second directional derivative. Thus at a particular scale, intensity changes can be located by finding zero-crossings in the output of a second directional derivative operator. A number of practical considerations, which will be illuminated in the discussion of the implementation, suggested that the initial operators not be directional operators. The only non-directional linear second derivative operator is the Laplacian operator. It was then shown (Appendix A, Marr & Hildreth 1979) that provided two simple conditions on the intensity function in the neighborhood of an edge are satisfied, the zero-crossings of the second directional derivative taken perpendicular to an edge will coincide with the zero-crossings of the Laplacian along that edge. Therefore, theoretically, we could detect intensity changes occuring at all orientations using the single non-oriented Laplacian operator.

The two operations, the Gaussian and Laplacian, can be combined into a single operator, so that one can now detect intensity changes occuring at a particular scale by locating the zero-crossings in the output of $\nabla^2 G(x,y)$, the Laplacian of a Gaussian distribution. The operator, together with its Fourier transform, is illustrated in Figure 2. Examples of the application of this operator appear in Figure 3. The convolution outputs for the original images of Figure 3a are shown in Figure 3b. Zero is represented by a medium grey, so that very positive values in the convolution are white, and negative values black. The zero-crossing contours appear in Figure 3d. The pictures are 320 x 320 picture elements. The size of the operator is defined by w, the diameter of its positive central region, which in this case is 9 pixels.

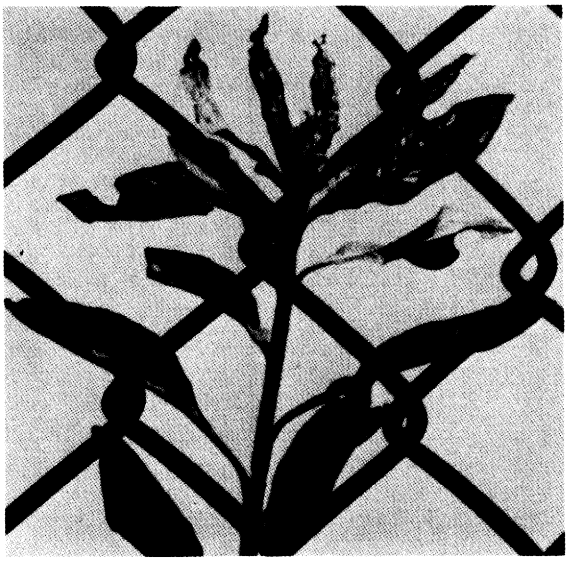
There are two additional properties which can be computed for the



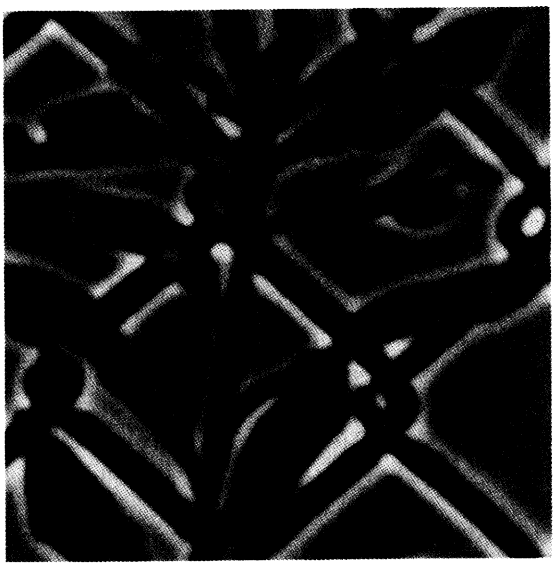
2. The initial filters. (a) illustrates D^2G, the one-dimensional second derivative of a Gaussian, and (b) shows its two-dimensional counterpart, ∇^2G. (c) and (d) illustrate the one and two-dimensional Fourier transforms.

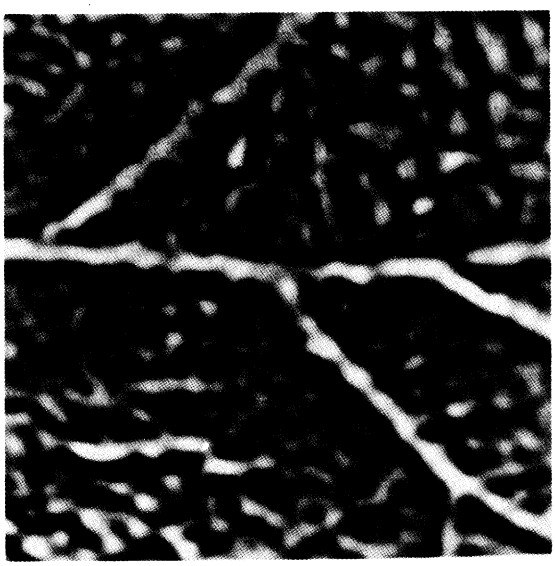
3. Examples of zero-crossing detection using $\nabla^2 G$. Column (a) shows three images, and column (b) shows the convolutions with the $\nabla^2 G$ operator of Figure 2 ($w = 8$), with zero being represented by a medium grey. In column (c), positive values of the convolution are shown in white, and negative values, black. In column (d), only the zero-crossing contours appear.

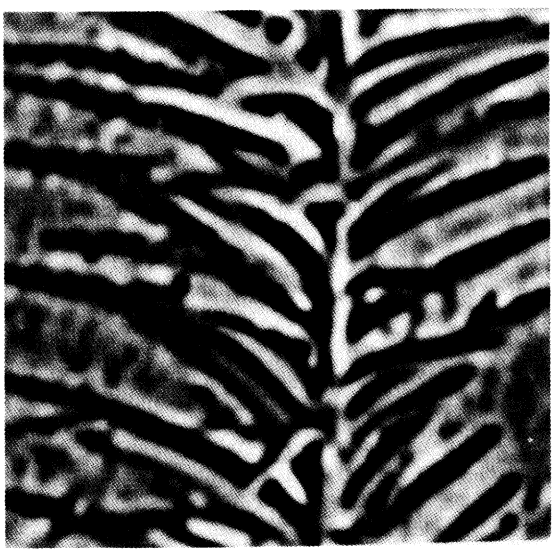
A.



B.

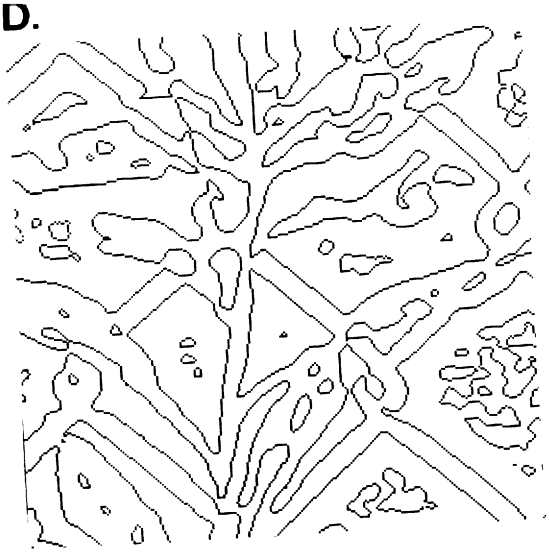


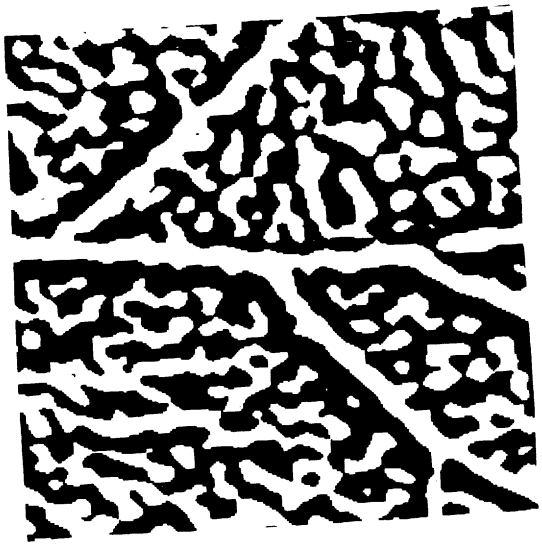


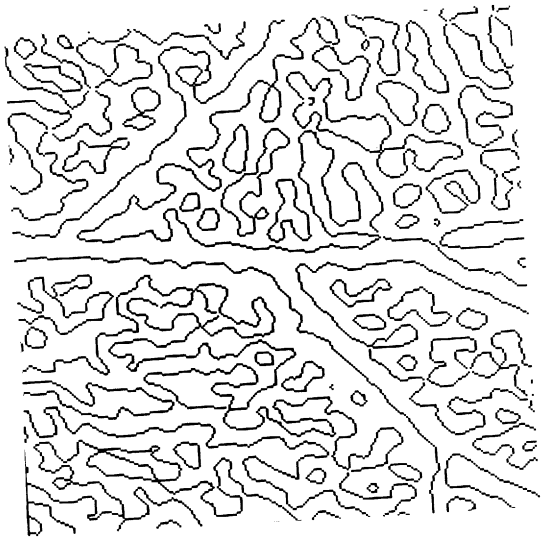


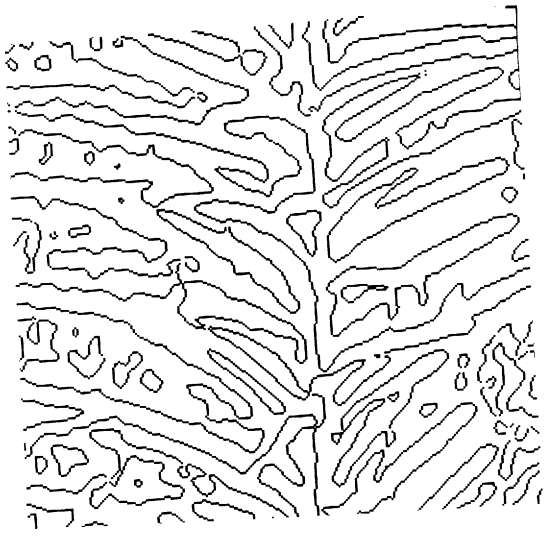
C.

D.









zero-crossings at a particular scale: slope and local orientation. The slope of a zero-crossing is the rate at which the convolution output changes as it crosses zero. Slope is related to the contrast and width of the intensity change, so will be necessary for later computation of these properties. In Figure 4, the magnitude of the slope of the zero-crossings is displayed as intensity; sharp, high contrast edges yield darker zero-crossing contours. Orientation is a two-dimensional property which can only be defined for a line of zero-crossings. Thus a *zero-crossing segment* was formally defined to be a linear, connected line of zero-crossings whose orientation is roughly uniform. The theory proposes that one mechanism for representing local orientation is to divide the zero-crossing contours into a set of short zero-crossing segments, each with an associated orientation, average slope, and length. Some zero-crossing contours will be small, closed contours which can not effectively be described by a set of zero-crossing segments. These contours are described as blobs, and a position, orientation, size, and average slope is computed for the entire structure. It is also proposed that we make explicit closely spaced, parallel zero-crossing segments (bars), and places where the contours reflect the termination of an edge or line. These additional elements will be discussed further in Sections II and III. The final description we obtain from a single channel is a set of zero-crossing contours, with symbolic descriptions of orientation, slope, and size attached to segments of the contours, or in the case of blobs, to the entire contour. An example of this first stage appears in Figure 5.

Figure 6 illustrates the zero-crossing contours from the output of several size operators applied to an image. We are now faced with the problem of integrating these descriptions which we obtained independently from the different size operators. The stereo matching computation, and the detection of motion take place early in visual processing, operating directly from the

4. Zero-crossings in the output of the $\nabla^2 G$ operator applied to an image, with the slope of the zero-crossing displayed as intensity; that is, a zero-crossing across which the convolution output changes more sharply will be displayed with a darker line. Slope provides a rough measure of contrast, so that one expects very high contrast edges, such as those running along the border between the buildings and the sky, to yield darker zero-crossing contours.

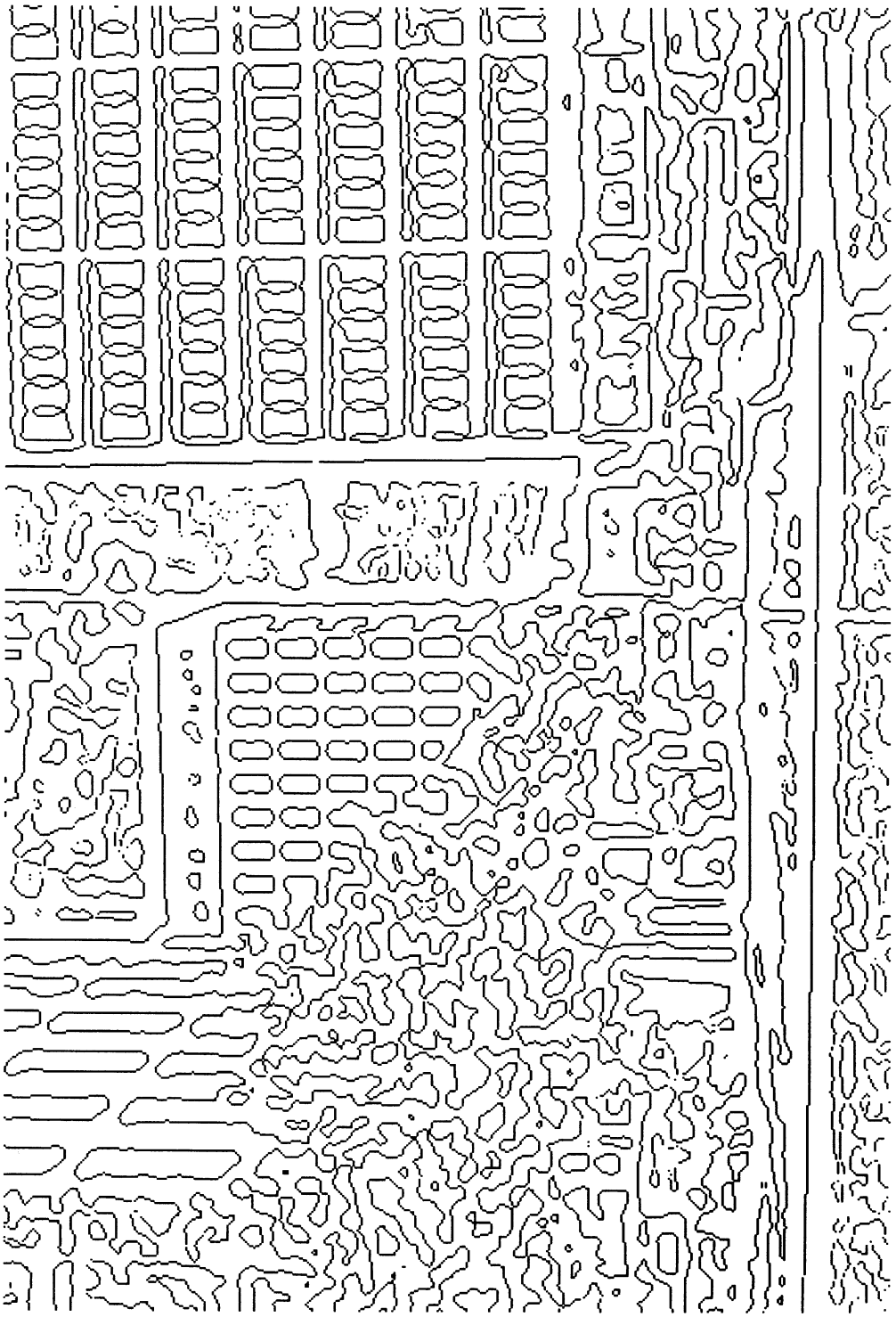
(4)

5. Symbolic illustration of the single channel description. (a) shows all of the zero-crossings from the convolution of the image in Figure 4 with the $\nabla^2 G$ operator, with w = 9 picture elements. (b) illustrates symbolically, the orientation attached to segments of the zero-crossing contours shown in (a). This diagram illustrates only the spatial information associated with these descriptors; below is an example of a full description of a segment of the contour:

```
(segment (position 104 23)
         (orientation 65)
         (slope -34)
         (length 18))
```

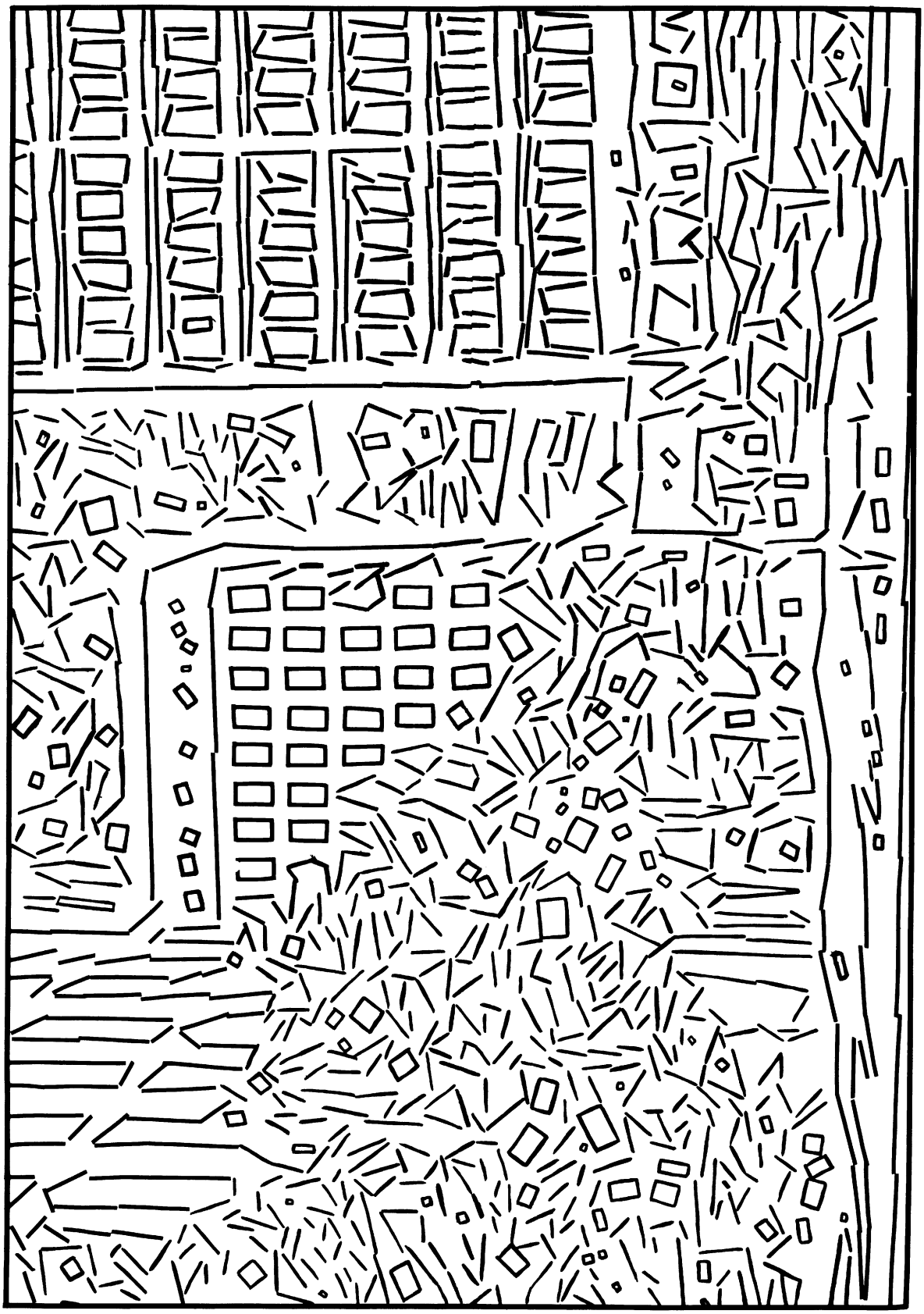
This additional information which we make explicit along the zero-crossing contours will be used in the later processing of the channel descriptions.

A.



(5)

B.

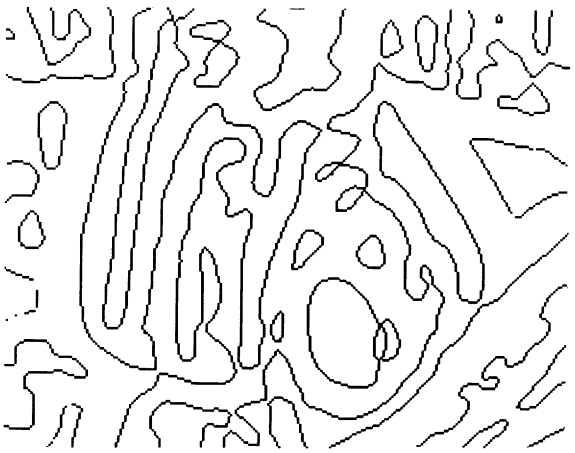


(5)

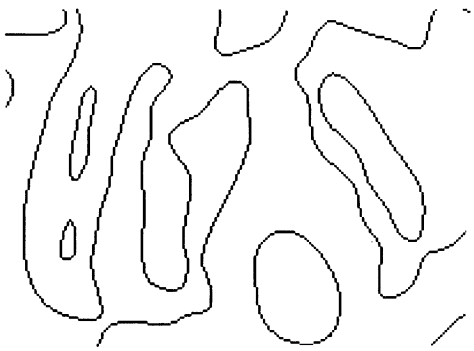
a.

b.

c.



d.



6. Zero-crossings from different size operators. The image in (a) has been convolved with $\nabla^2 G$ having w = 6, 12, and 24 pixels. These filters span approximately the range of filters that operate in the human fovea. (b), (c), and (d) show the zero-crossings obtained from these convolutions. In the full channel descriptions, also associated with these zero-crossings will be their slope and local orientation.

outputs of the single channels (Marr & Poggio 1979, Marr & Ullman 1979). That is, these processes combine channel outputs in a particular way to compute disparity information along the zero-crossing contours, or the direction of motion of these contours. The primal sketch theory proposes that the individual channel descriptions are also combined into a single description of intensity changes, in which the important properties of contrast and width are made explicit. From this single description, grouping and early texture analysis takes place, prior to other processes, such as motion correspondence. Unfortunately, the integration of these descriptions into a single raw primal sketch is not a simple matter. Some intensity changes will be detected over a range of adjacent scales, while others may be detected only at a single scale. A key observation aided in simplifying this problem; because changes in the physical world are localized, the intensity changes to which they give rise are also localized, so that if two adjacent channels detect a particular intensity change, indicated by the presence of zero-crossings, then the zero-crossings from the two channels will tend to be spatially localized. This critical observation formed the basis of the *spatial coincidence assumption*:

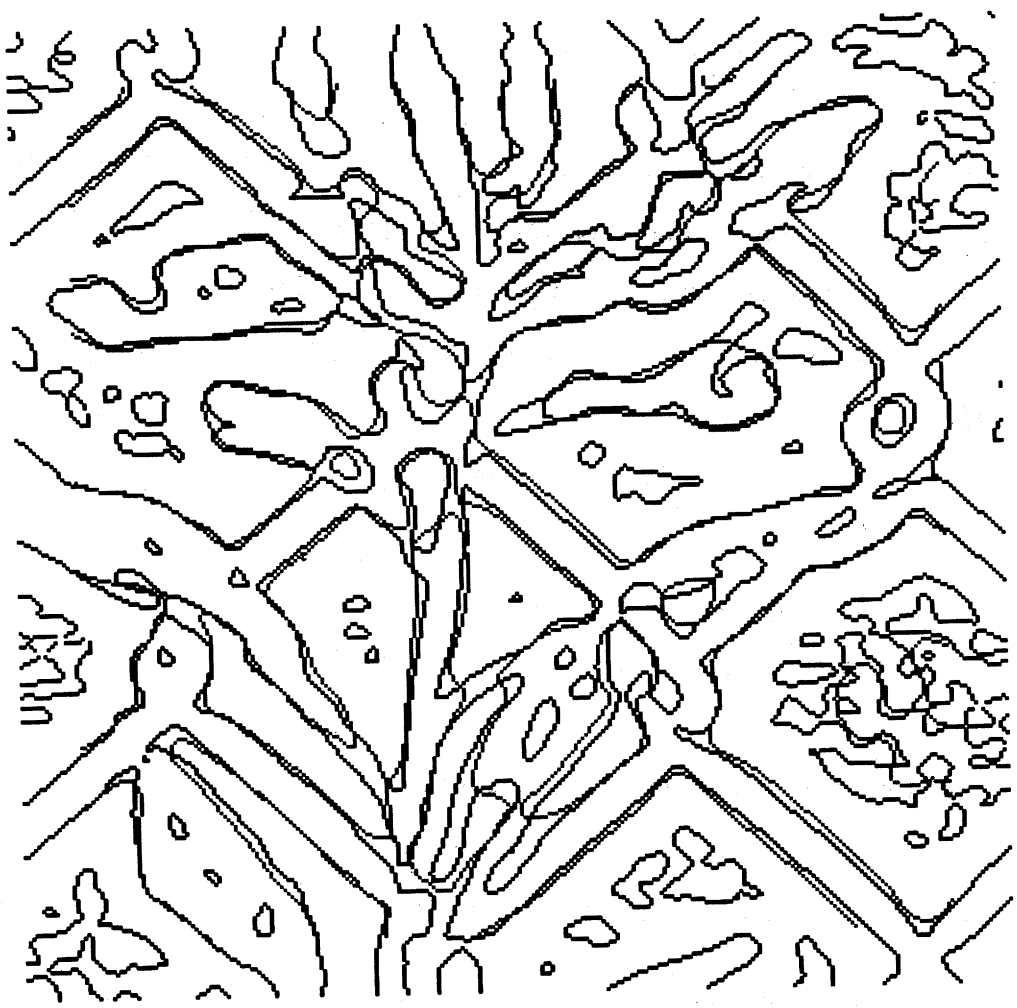
> "If a zero-crossing segment is present in a set of independent $\nabla^2 G(x,y)$ channels over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments may be taken to indicate the presence of an intensity change in the image due to a single physical phenomenon (a change in reflectance, illumination, depth or surface orientation)" (Marr & Hildreth 1979, p. 30)

Information in one channel which does not coincide with that from adjacent channels is assumed to arise from a physical phenomenon which can only be measured at that one scale, so it gives rise to an independent descriptive

element. Figure 7 illustrates this coincidence of the zero-crossing contours; zero-crossings from the smaller channels appear black, those from the larger channel are light grey, and those which exactly coincide are medium grey.

This problem of integrating the channel descriptions is a difficult one, which I will take a slightly more conservative view toward in this work. What is essential is that we work toward an understanding of the relation between the zero-crossings from the separate channels, their correspondence with the physical changes that give rise to them, and what information we could extract by examining the individual channels. Spatial coincidence may certainly play a key role, in deciding the presence of a physical edge, and computing the necessary edge properties. I would only like to suggest that it is premature to ask whether the information concerning intensity changes exists explicitly in a single description, and what the particular form of the representation might be, before understanding how much information we could possibly extract from the integration of the channels, and the necessary constraints on the representation imposed by grouping, texture, motion correspondence, or lightness.

To summarize this section, a theory of edge detection has been described, which proposes that intensity changes first be detected at a range of different scales, by localizing the position, slope, and orientation of zero-crossings in the outputs of a set of $\nabla^2 G(x,y)$ operators. These independent descriptions are then integrated in some way, in order to make additional information explicit for subsequent processing.



7. The spatial coincidence of zero-crossings from different size channels. The zero-crossings obtained from the convolution of the image in Figure 1a with $\nabla^2 G$ operators of size w = 9 and 18 pixels are shown superimposed, with the smaller channel zero-crossings displayed in black, and larger channel zero-crossings in light grey. Points at which the zero-crossings exactly coincide are shown in medium grey. The smaller operator detects edges with much finer detail, but for strong physical edges, both operators detect the edge, with zero-crossing contours which roughly coincide.

II. Computing the Single Channel Description

We begin this stage with an image, which is a two-dimensional array of intensity values. As it is implemented serially, the basic algorithm for computing a single channel description is as follows:

1. Convolve the image with a set of two-dimensional $\nabla^2 G$ functions, which vary in size. The size of a $\nabla^2 G$ operator is defined by the diameter of its central, positive region, w. The operator can be constructed by setting the space constant for the Gaussian, $\sigma = w/2\sqrt{2}$:

$$\nabla^2 G = [2 - r^2/\sigma^2]e^{-r^2/2\sigma^2}$$

r is the distance from the center of the mask.[1]

2. Locate the position and local slope of zero-crossings in each of the convolution outputs, by scanning the output in the horizontal and vertical directions.

3. Follow the contours formed locally by the zero-crossings, constructing a description of the orientation, average slope, and size of zero-crossing segments, and blobs.

There are four issues concerning this early stage which I would like to discuss in more detail. The first two are the shape and size of the initial operators. We will see that for the reliable detection of intensity changes, the shape of the operators is highly constrained, whereas the size is less constrained. For the development of a machine vision system, the range of operator sizes may vary with the range and resolution of edge information required by the particular application. Interestingly, if we do a comparitive study of animals with developed visual systems, we see that the basic shape of the operators in the initial processing in the retina is similar (see, for

example, mudpuppy: Dowling & Werblin 1969, Werblin & Dowling 1969; cat: Rodieck & Stone 1965, Cleland, Dubin & Levick 1971a; monkey: Hubel & Wiesel 1968, deMonasterio & Gouras 1975, deMonasterio 1978a), but the range of sizes, and geometry of the eye varies to suit the survival needs of the particular animal.

The third topic of discussion in this section is the assumptions made in the theoretical analysis of the detection of intensity changes. The assumptions and their consequence, together with a test for their validity, will be examined.

Finally, I would like to describe, in more detail, the construction of the symbolic description of the zero-crossing contours as a set of segments, blobs, and bars, with properties of orientation, slope, and size.

II.1 The Initial Filters

Shape of the Filters

As the theory of the primal sketch developed, there was much experimentation in the implementation with the shape of the initial filters; a set of properties slowly emerged which appeared critical for these operators. In Section 1, we have seen the theoretical development of the $\nabla^2 G$ operator. However, many of the theoretical advancements were motivated by the results of this experimentation with different types of operators. Theory and practise finally converged on the $\nabla^2 G$ operator. As a result of our experimentation, we can now state three critical properties for the shape of this operator which can be strongly supported, both in theory and practise:

1. localization in space
2. localization in frequency
3. no orientation dependence

To appreciate the importance of each of these properties in practise, we can use the general framework of locating zero-crossings in the output of a second derivative operator, and examine the performance of this algorithm in locating

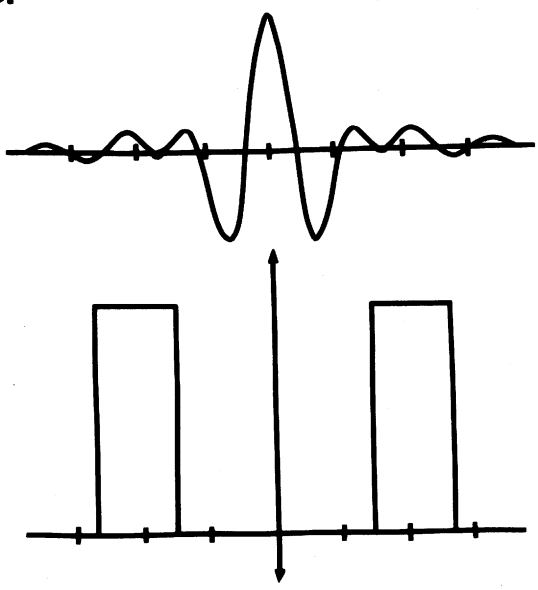
the position of intensity changes as each of the critical properties of the operator is violated. The first demonstration appears in Figure 8, reproduced from Marr & Hildreth (1979). In the first filter, we have violated the property of spatial localization, with an extended sinc function, whose frequency spectrum approaches an ideal bandpass filter. Side lobes in the spatial filter give rise to echoing zero-crossing contours; in Figure 8b, these multiple contours can be seen running parallel to the contours outlining the bars of the chain-link fence. This operator is unreliable as an edge detector, as it yields zero-crossings in regions where there are no changes in intensity in the original image.

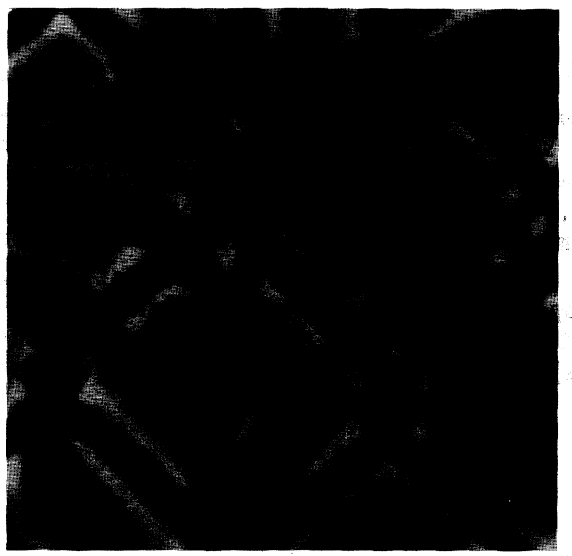
Frequency localization was violated with an operator which is the square-wave approximation to the second derivative. Its affect on the zero-crossing contours, illustrated in Figure 8d, is a smoothing. For a particular size, the square-wave operator does not detect intensity changes with the resolution of the $\nabla^2 G$ operator. For a large class of one-dimensional signals, an increase in the bandwidth of the signal is coupled with a decrease in the number of zero-crossings, provided the center frequency is held constant (Logan 1977). If this result extends to the case of two-dimensional signals, it may offer a theoretical explanation for the decreased detail in the output of this operator. Practically, one property of edges which, with contrast, must be decoupled from the measurement of slope across an intensity change, is edge width. A rough analysis of the frequency spectrum of an edge, which can be obtained by measuring the strength of an edge through filters tuned to different spatial frequencies, constrains the computation of edge width. Using the $\nabla^2 G$ operator, contrast and width can be computed directly for an isolated edge (see Section IV). The reliability of possible schemes for computing these properties is a factor in the evaluation of masks with different frequency characteristics. For some applications of edge detection, the constraint of frequency localization may not be as critical as spatial localization; although the Gaussian optimally

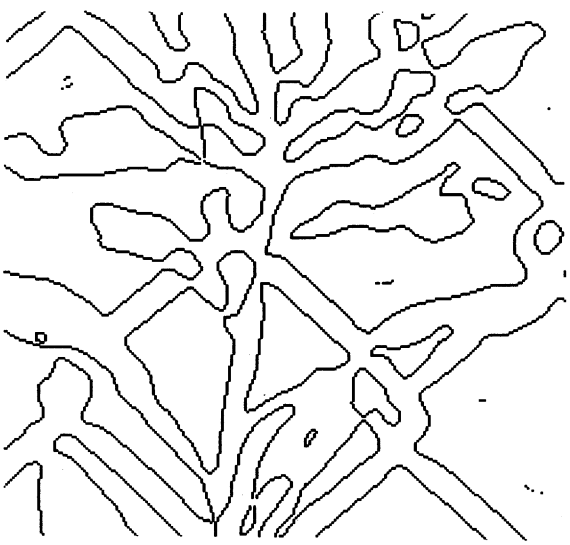
8. Comparison of the performance of the $\nabla^2 G$ filter with that of similar filters. Column (a) shows an image, its convolution with $\nabla^2 G$, and the resulting zero-crossings. Column (b) contains the same sequence, but for an extended sinc function, whose frequency spectrum approaches the ideal bandpass filter, shown in the top picture. In the zero-crossings, we can see echoing zero-crossing contours along the strong edges. Columns (c) and (d) exhibit a similar comparison, now between the $\nabla^2 G$ operator, and the square-wave approximation to the second derivative. The square-wave operator sees relatively fewer zero-crossings. The widths of the central excitatory regions of the filters are the same for each pair, being 12 pixels for (a) and (b), and 18 pixels for (c) and (d).

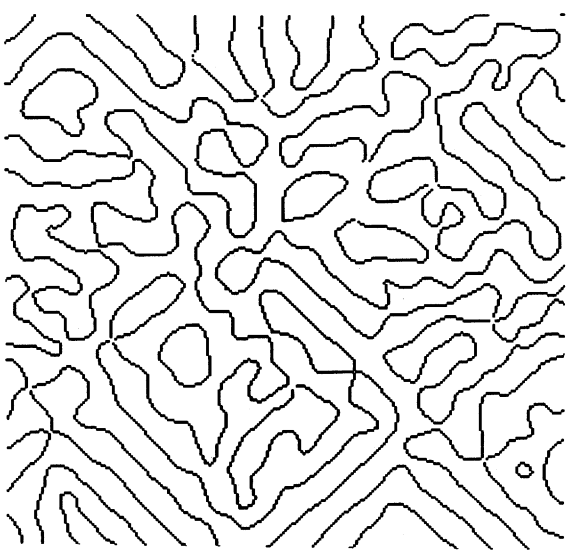
A.

B.

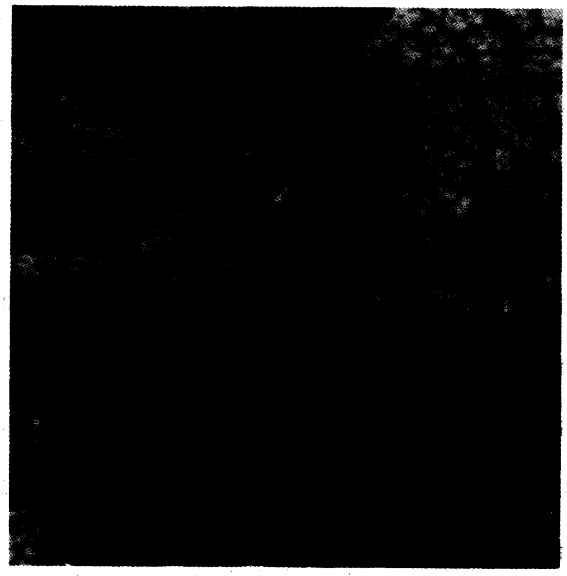




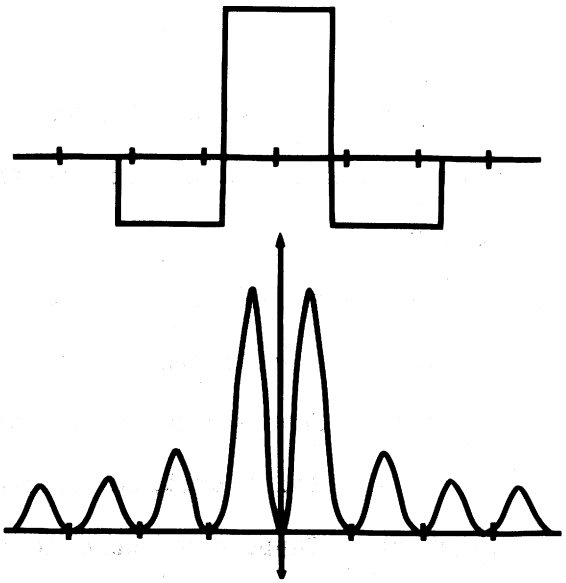


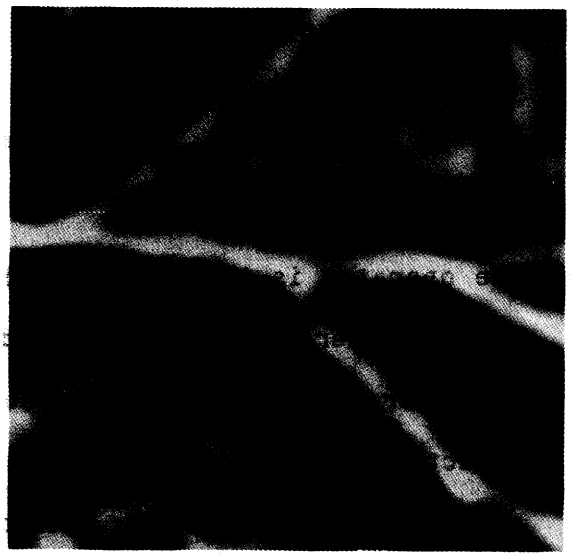


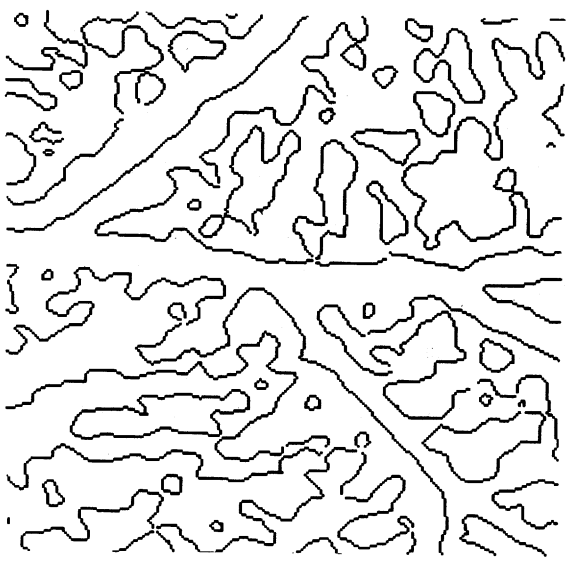
C.

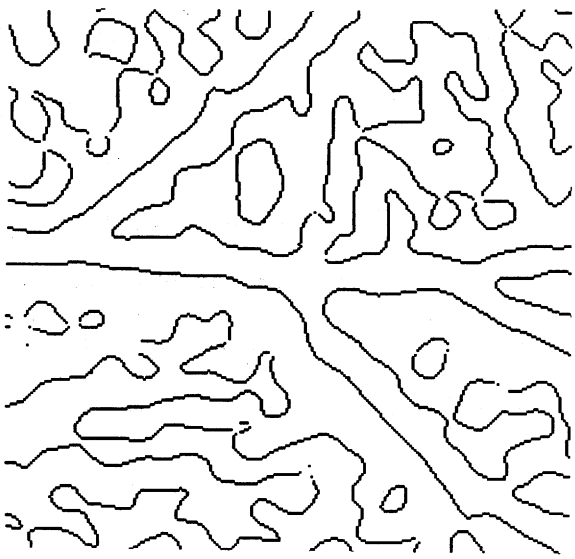


D.









satisfies these two constraints, the breakdown in behavior of the zero-crossings is not as significant in practice, as in the case of violation of spatial localization.

The above two examples admittedly represent extreme violation of our localization requirements, but clearly convey the need for examining the shape of the filters from the two perspectives of space and frequency.

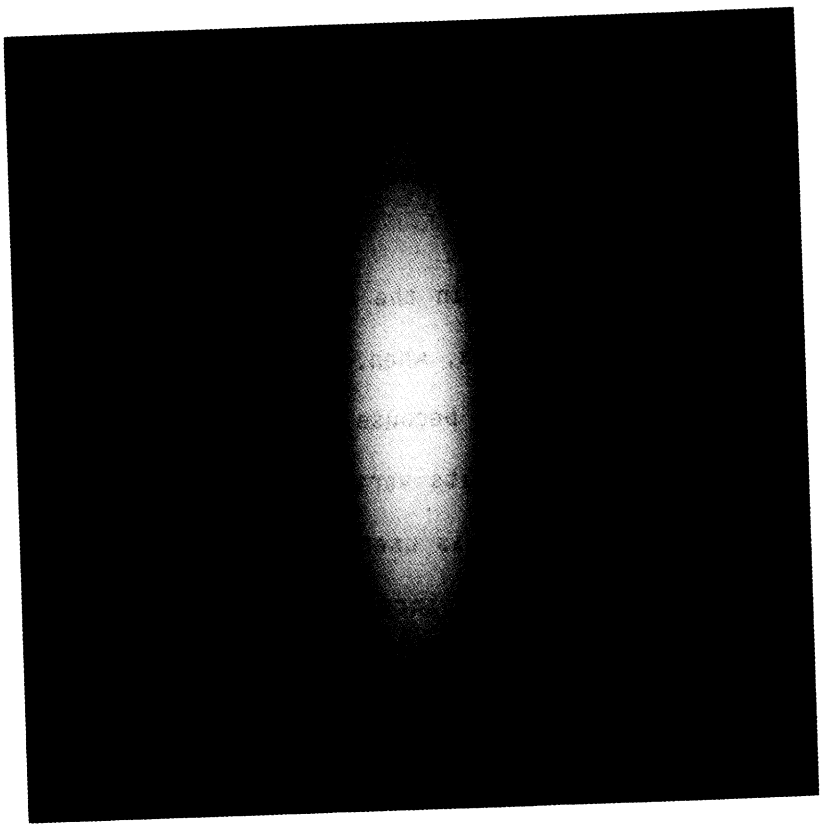
The third critical property of the mask is that it not be orientation dependent. Suppose we were to use oriented operators. Without violating our first two criteria, we can examine the behavior of a vertically oriented mask whose horizontal cross-section is the second derivative of a Gaussian, and whose vertical cross-section is a Gaussian:

$$M(x,y) = e^{-y^2/2\sigma_v^2}[e^{-x^2/2\sigma^2} - (x^2/\sigma^2)e^{-x^2/2\sigma^2}]$$

The operator is illustrated in Figure 9. (A similar operator, with an 'edge' shaped rather than 'bar' shaped field, has been used by Macleod (1972).) Such an operator will compute the second derivative of intensity in a particular direction. If an edge appears in the image at an orientation θ, the orientation of the derivative operator used to measure this edge should also be θ. Provided that the *Condition of Linear Variation* (discussed in Section II.2) is satisfied, the direction of this derivative will be the direction in which the zero-crossings have maximum slope. This suggests the following algorithm for detecting intensity changes at all orientations, using directional derivatives:

1. Convolve the image with oriented operators at all orientations.
2. For each output, extract those intensity changes whose orientation aligns with the orientation of the operator.

We will find that in practise, such a simple scheme is inadequate. Let w_x, the central panel width for the horizontal cross-section, be 9 pixels, so that we can

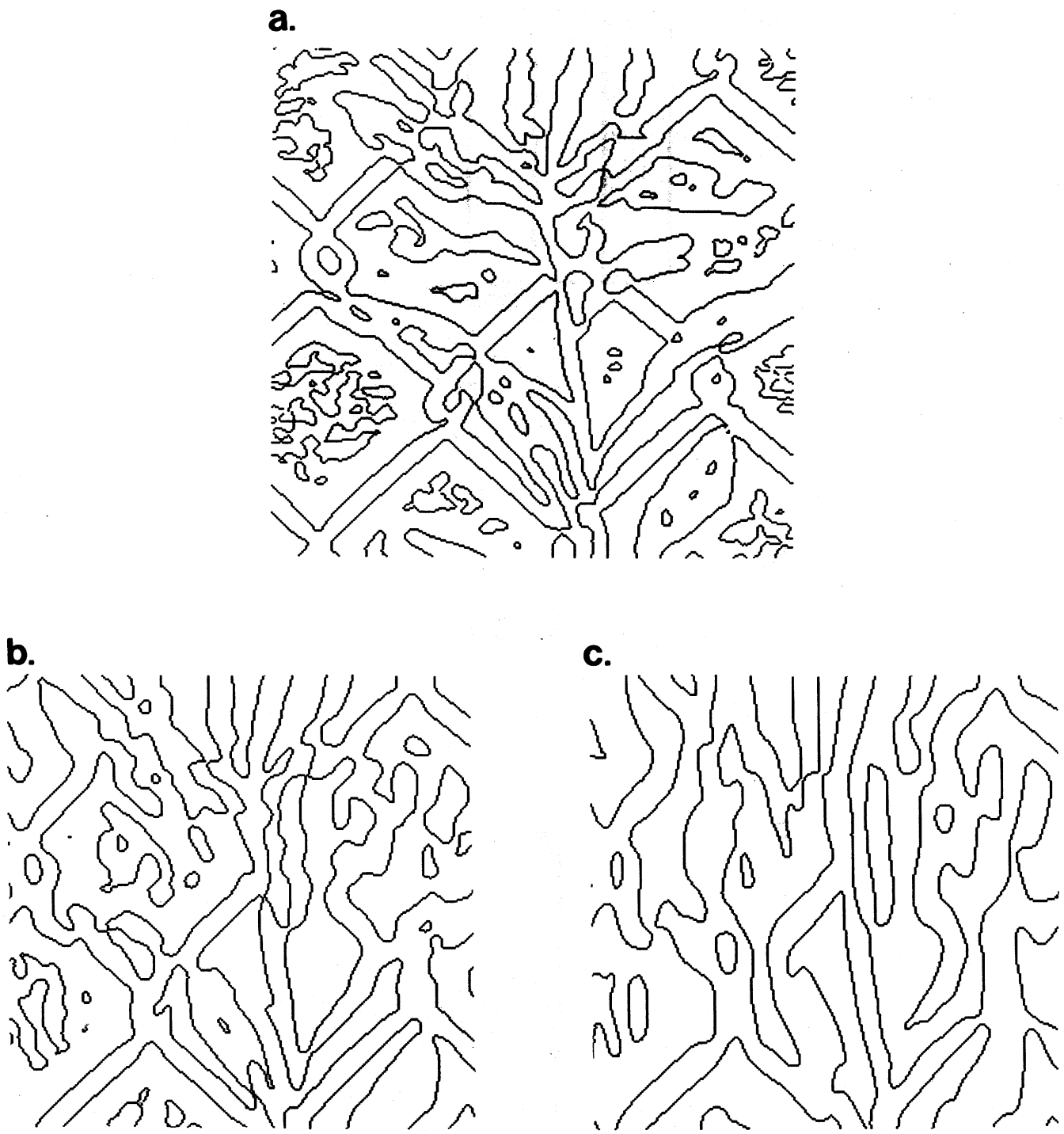


9. Vertical operator. The horizontal cross-section of this operator is the second derivative of a Gaussian, illustrated in Figure 2a; its vertical cross-section is a Gaussian.

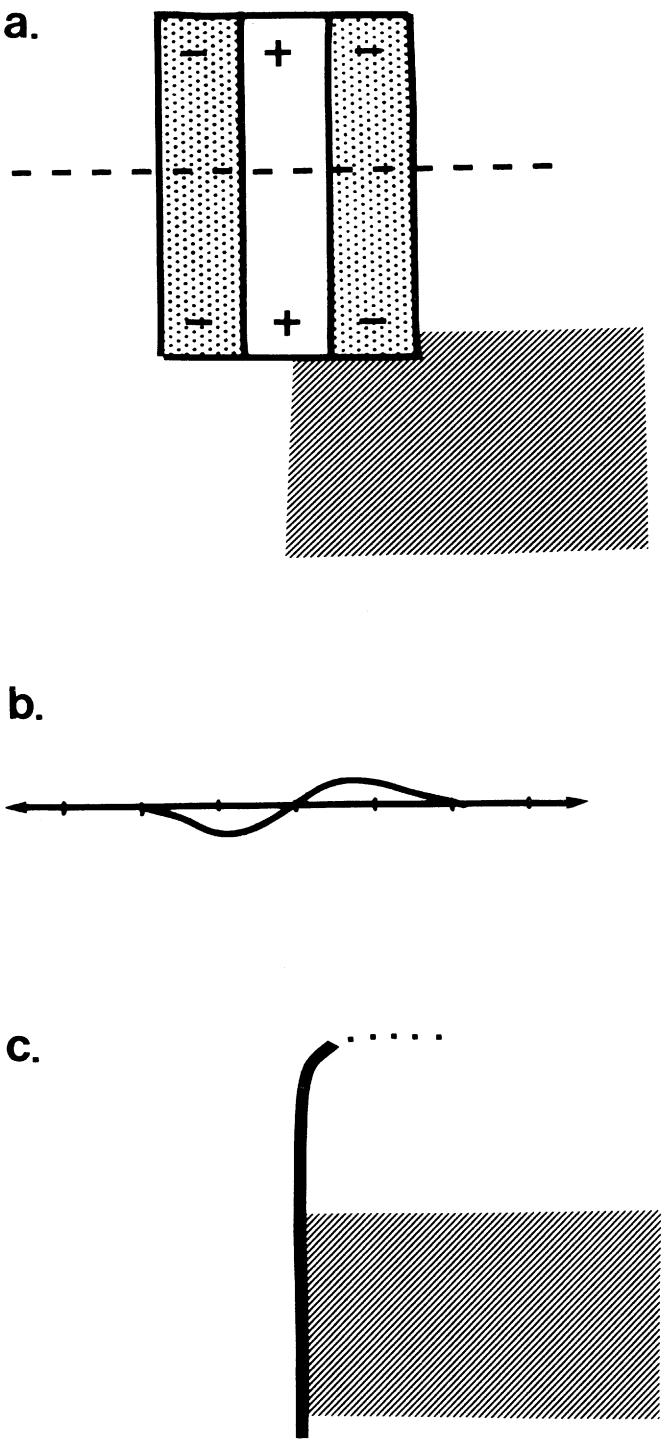
draw a direct comparison between the output of the vertical mask and the non-oriented Laplacian we have used previously. The remaining parameter is σ_v, the space constant for the vertical Gaussian. Define an aspect ratio for the mask as $4\sigma_v/w_x$. A Gaussian function will cover about 90% of its area over a distance of roughly 4σ, so this measure provides an approximate length-to-width ratio for the central positive region of the mask. Figure 10 shows a comparison between the zero-crossings in the output of the two-dimensional non-oriented operator, with w = 9, and two vertically-oriented masks with w_x = 9 and aspect ratios 4:1 and 2:1.

The most evident problem in the output of these filters is the vertical smearing of zero-crossing contours, when, for example, an edge terminates in the vertical direction. This occurs because the mask will begin to respond to an edge as soon as the edge enters its vertical field, as shown in Figure 11. If the output of the vertical mask is used for making assertions about vertical zero-crossing segments, it will incorrectly assert the presence of these elongated contours associated with terminations.

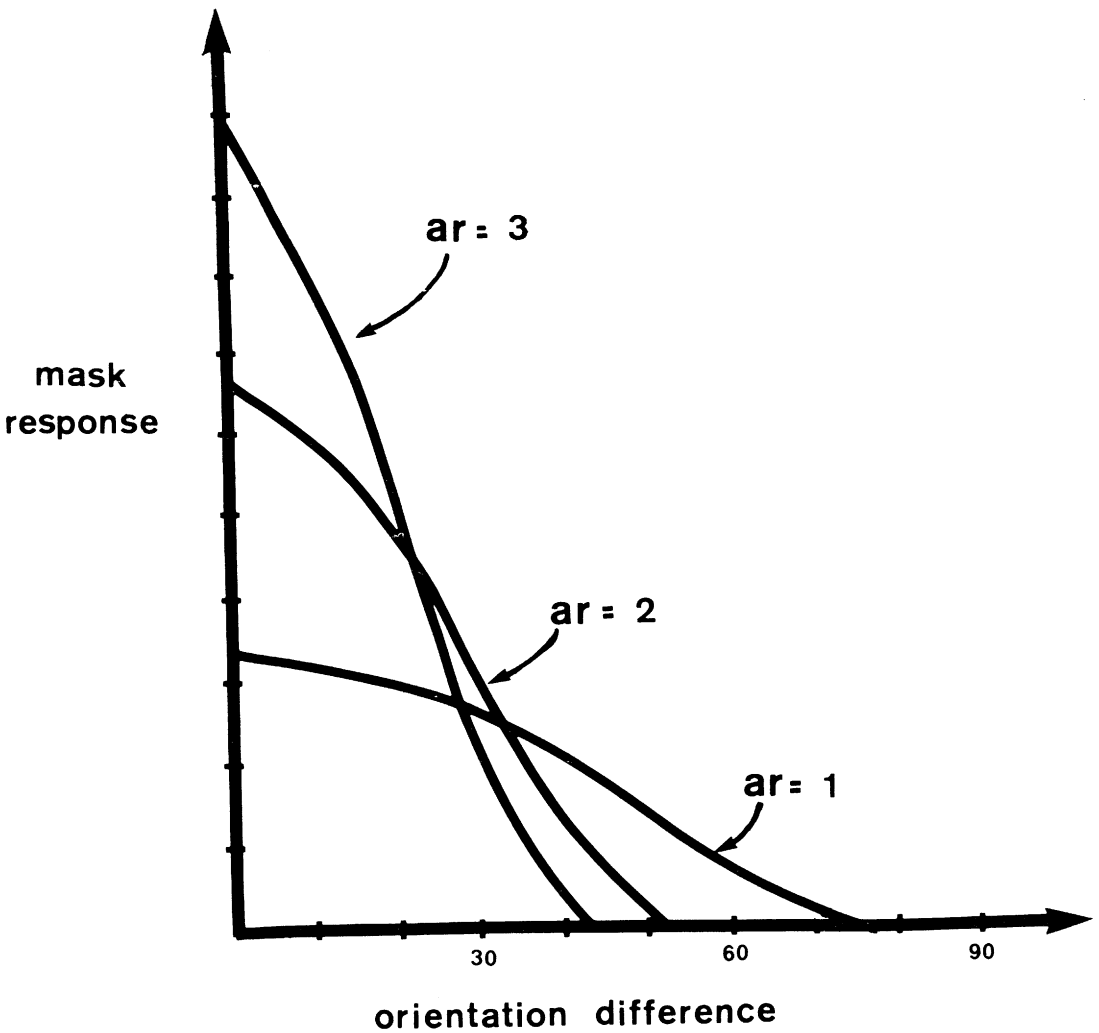
Incorrect zero-crossing segments, whose orientation again coincides with the orientation of the mask, may also appear when the difference between the orientations of the mask and edge approaches 90°. Although a particular derivative operator will optimally detect an edge whose orientation coincides with its own, it will detect the position of an edge for a range of orientations. In the case of a continuous ideal step edge, the oriented mask should yield a smooth zero-crossing contour along the edge, unless the edge is exactly perpendicular to the mask. As the orientation difference between mask and edge increases, the response of the mask decreases, as in Figure 12 (for a range of aspect ratios). In practice, when the response of the mask is near zero, it becomes more sensitive to minor changes in intensity due to noise, discretization, and quantization, which combine to cause the output to fluctuate



10. Comparison between the zero-crossings in the output of the $\nabla^2 G$ operator, and the vertical operator illustrated in Figure 9. In each case, the width of the central positive regions are the same ($w = w_x = 9$ picture elements). The aspect ratios for the vertical masks are 2:1 and 4:1 for (b) and (c), respectively. The extension of zero-crossing contours along the vertical direction is especially apparent in this example.



11. An oriented operator will begin to respond to an edge as soon as the edge enters its vertical field. (a) indicates that we wish to compute convolution values for the slice of the image along the dotted line. In (b), we have the convolution output along this line plotted in one dimension. The final zero-crossing contour is partially displayed in (c). As we can see, the contour extends beyond the termination of the edge.



12. The magnitude of the zero-crossing slope, plotted against the difference in orientation between a vertical mask, and an oriented edge in two dimensions. As we increase the aspect ratio of the mask, its response becomes more narrowly tuned.

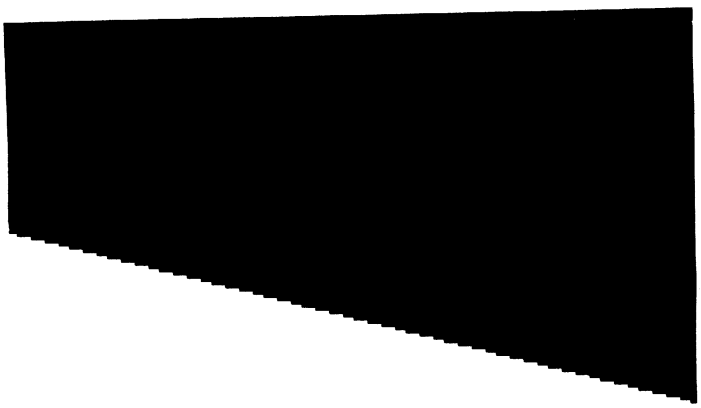
around zero, producing a number of very shallow zero-crossings. The vertical extent of the mask then causes the effect of these fluctuations to be extended along the vertical direction, yielding a zero-crossing contour containing components along the orientation of the mask rather than the orientation of the edge. This effect is illustrated in Figure 13. For a mask with aspect ratio 2:1, and maximum contrast edge, this breakdown in the behavior of the zero-crossing contours becomes evident at an orientation difference of about 50°; for an aspect ratio of 4:1, this difference is about 35°.

Although these effects of oriented masks are reduced as we decrease the aspect ratio, Figure 14 indicates that even for an aspect ratio as small as 1:3, they are still evident. In Figure 18, the zero-crossings of this operator can be seen superimposed on the zero-crossings of the $\nabla^2 G$ operator.

The use of oriented masks greatly complicates the edge detection process. First, it is not sufficient to simply extract from their output, all zero-crossing segments whose local orientation aligns with the operator orientation. One must be able to recognize the above two situations, carefully designing a set of criteria for determining the reliability of a particular zero-crossing contour. Orientation is a property of an edge which is important to make explicit for later processing, but we are suggesting that orientation is computed after the convolution of an image with the non-oriented Laplacian operator. The implementation of a more knowledgeable scheme for combining outputs from several orientations (Marr 1976b), which attempted to deal with the above problems directly, proved to be extremely difficult. Finally, convolutions are very costly, so it is certainly more efficient to use the non-oriented operator, and obtain a map of intensity changes at all orientations in a single step.

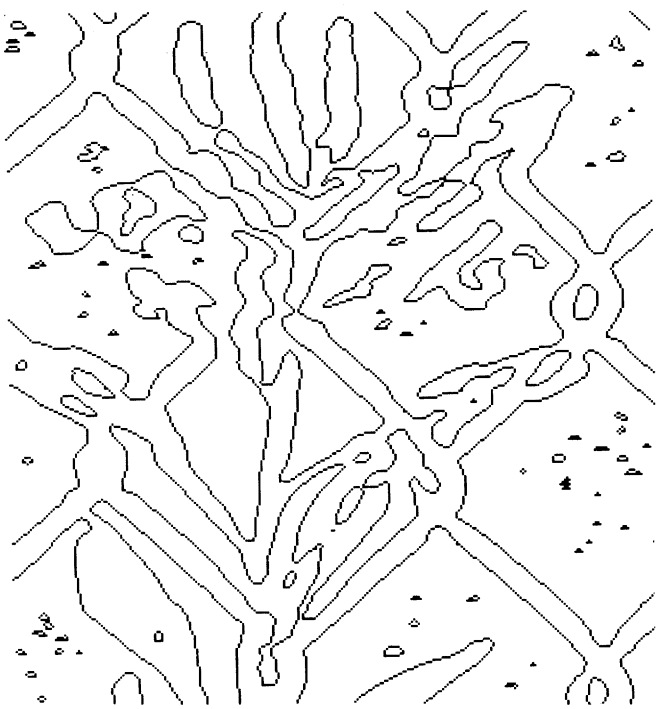
The above experiments, together with the theoretical arguments posed in Section 1, provide strong evidence in support of the initial $\nabla^2 G$ operators proposed in this theory. Of the three properties shown to be critical for these

a.



b.

13. In (a), we have a simple step edge, oriented 75 degrees away from the vertical. The zero-crossings from the output of this edge convolved with a vertical mask, with aspect ratio 4:1 are illustrated in (b). The contour of zero-crossings has a general orientation along the edge, however has a significant component along the vertical direction. This effect is in general due to the discretization of an edge, and quantization of the intensities.



14. Zero-crossings in the output of a vertical mask whose aspect ratio is 1:3. Although the problems resulting from the smoothing of intensity changes along the vertical direction are significantly reduced, they are still apparent here.

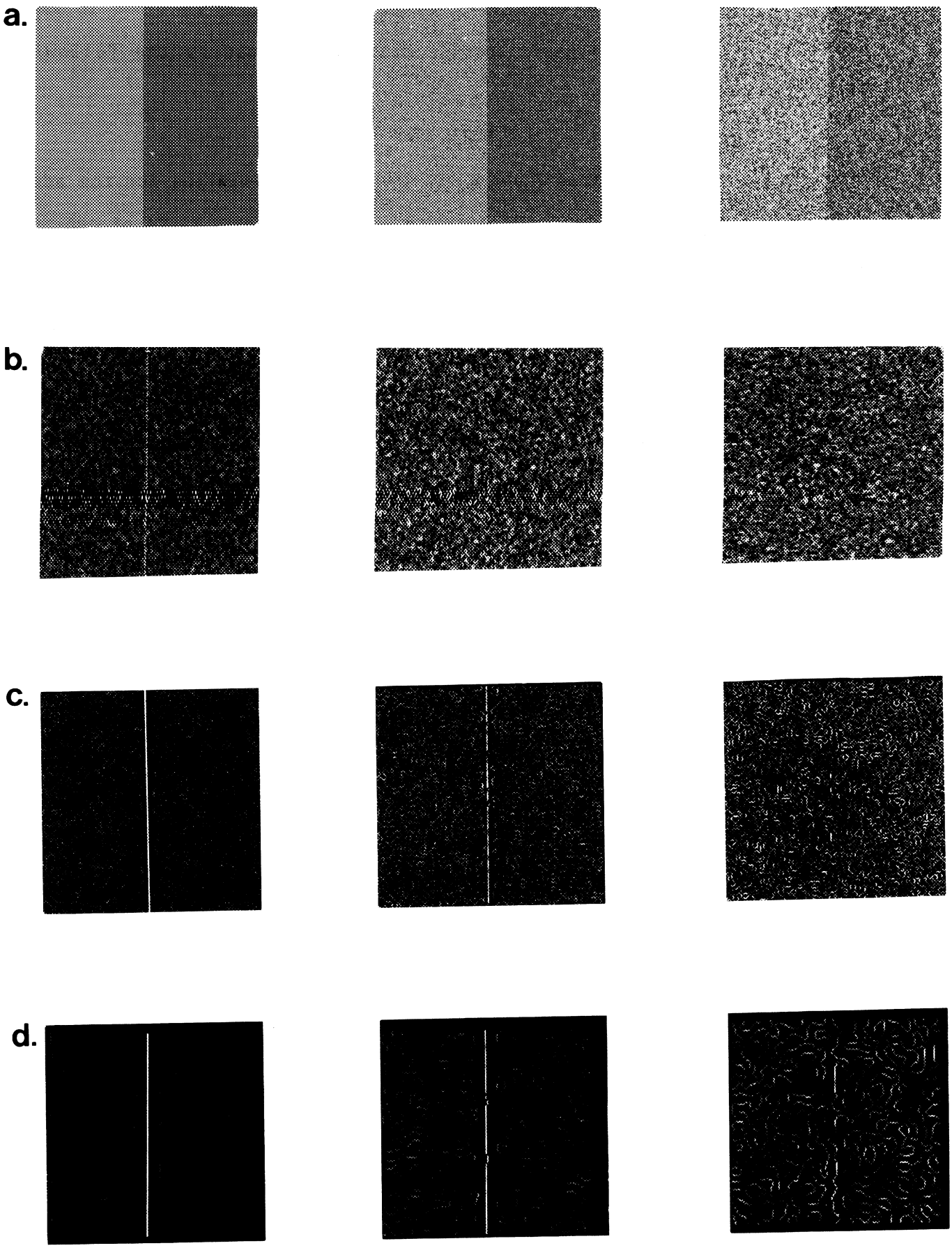
operators, the first two properties of localization in space and frequency, are best satisfied by the Gaussian distribution (Leipnik 1960); and use of the Laplacian satisfies the property of having no orientation dependence.

Size of the operators

Concerning the sizes of these operators, there are two parameters: the overall range of sizes, and the separation between sizes. Here, there is a lower bound on the smallest size operator one should use for the reliable detection of edges, and a loose constraint on the maximum separation between sizes. Other constraints will arise from the range and resolution of information required by the particular application of the edge detection process. For example, for the stereo matching process, the size of the largest mask constrains the largest disparity, at a particular location, which can be fused for a given eye position, while the size of the smallest operator determines the resolution to which disparity information is computed (Grimson 1980). The human visual system analyses its input at an extremely fine resolution, both in its original sampling of visual information (roughly one receptor per 20" of visual arc for the central fovea (Polyak 1941)), and in its smallest operator size (1'30", proposed in (Marr, Poggio, & Hildreth 1979)). For many applications in machine vision, such as the detection of parts on a moving conveyor, or counting cells in a cell culture, such fine resolution is not essential.

A primary factor governing the smallest size operator used in the edge detection process is the noise level of the imaging system. Noise will generally be restricted to very high spatial frequencies, so by restricting the smallest size operator, we are placing a limit on the highest spatial frequencies allowed to pass through the channels. Figure 15 demonstrates the need for this lower limit; Figure 15a shows a simple two-dimensional step change in intensity, to which three levels of Gaussian noise have been added. A similar demonstration,

15. The relationship between mask size, and its sensitivity to noise. In row (a) are three step changes in intensity, to which different levels of Gaussian noise have been added. The images were then convolved with $\nabla^2 G$ operators with w = 1, 4, and 8 pixels. The zero-crossings of these outputs are displayed in rows (b), (c), and (d), respectively. The smallest operator is very sensitive to even small amounts of noise, whereas the larger operator is fairly robust, in that it detects the underlying step change in the presence of large amounts of noise.


a.
b.
c.
d.

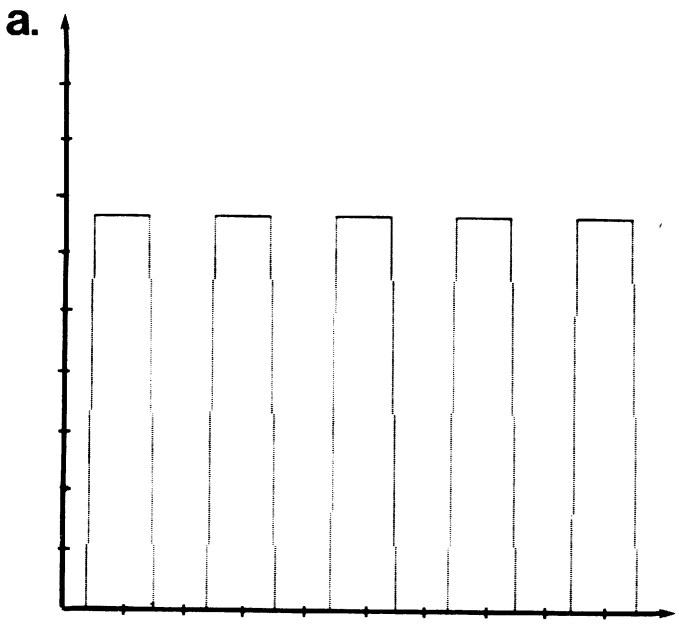
used in the evaluation of other edge detection schemes, can be found in Pratt (1978, p. 498). The signal-to-noise ratio is defined by the following expression:

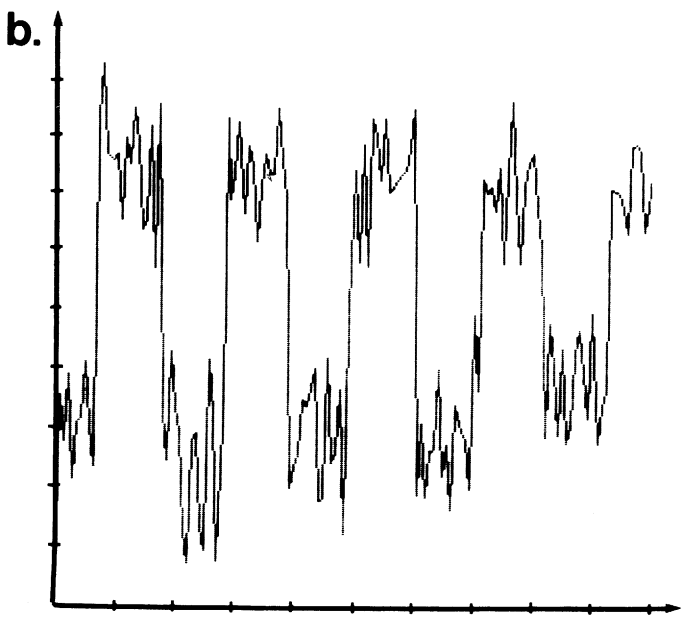
$$SNR = c^2/\sigma_n^2$$

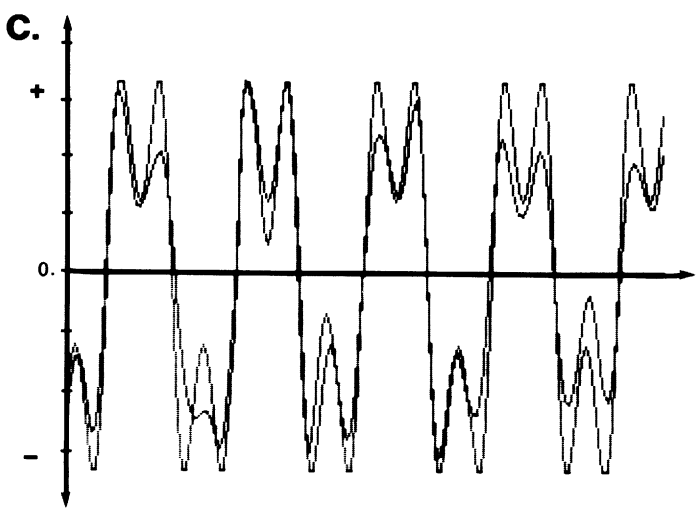
where c is the edge contrast, and σ_n is the standard deviation for the added noise. The three edges in Figure 15a have a signal-to-noise ratio of 100, 10, and 1, respectively. In Figures 15b, c, and d are the zero-crossings in the output of the three edges convolved with ∇^2G operators with w = 1, 4, and 8 pixels. The zero-crossings with maximum slope in each case are displayed at maximum intensity; the slope values are not correlated between outputs. For a signal-to-noise ratio of 100, all sizes are capable of detecting the primary edge, and respond more strongly to this edge than those due to the added noise. However, the smallest mask breaks down very fast, and for a signal-to-noise ratio of 1, it is unable to detect the primary edge behind the noise. The largest mask here corresponds to the smallest size presently used in our implementation; in this example, it is capable of reponding more strongly to the underlying edge, even in the presence of large amounts of noise.

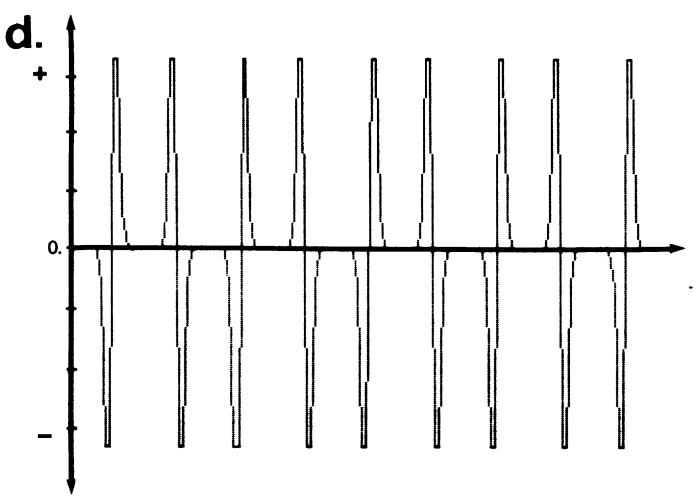
Figure 16 provides a second demonstration, now in one dimension. In Figure 16a is a series of bars, to which a high level of Gaussian noise has been added. This example is intended to emphasize that in evaluating the output of various size channels, we are interested in the position and slope of the zero-crossings, and how well they reflect the significant changes in our input profile. In Figure 16b, the outputs of the one-dimensional D^2G operator, with w_x = 6 pixels, convolved with the ideal bar profile and noisy profile are superimposed. Although there is deviation in the overall noise output, the position and slope of the zero-crossings are well preserved.[2] As we decrease the mask size, the error in localizing the edges of the bars does not change significantly, but the error in measuring slope increases, and additional zero-

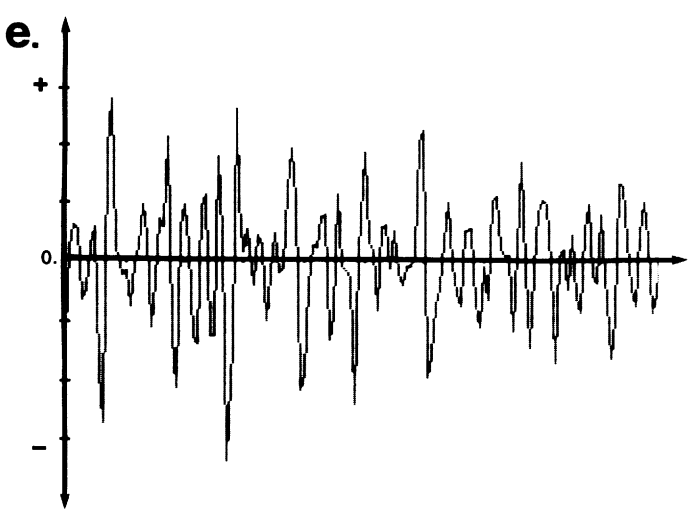
16. The relationship between mask size, and its sensitivity to noise. (a) and (b) illustrate the ideal profile for a set of bars, and the bar profile with a large amount of Gaussian noise added. As (c) illustrates, a large mask (in this case, w = 6 pixels) will still detect the edges of the bars with zero-crossings, although the non-zero portion of the convolution output may contain significant error. The two outputs are shown here superimposed. For a smaller mask, this is not the case. In (d) and (e), we have the response of a mask, with w = 2, to the ideal and noisy profiles, respectively. Many more zero-crossings are seen, reflecting the mask's sensitivity to edges created by noise.


a.


b.


c.
+
0.
-


d.
+
0.
-

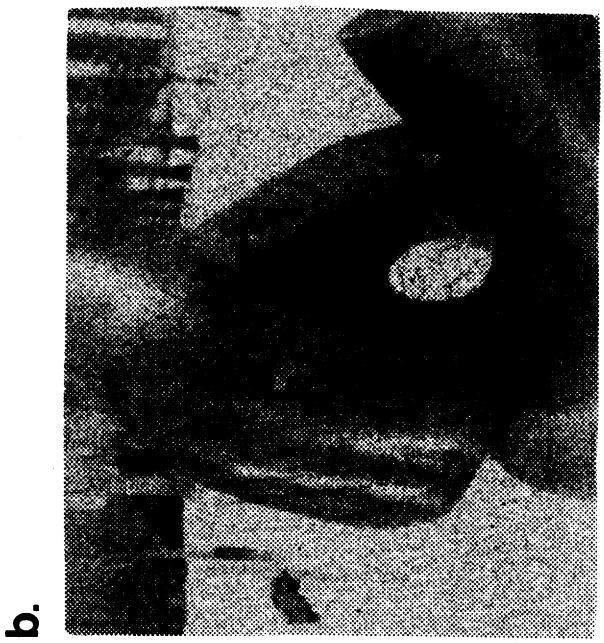

e.
+
0.
-

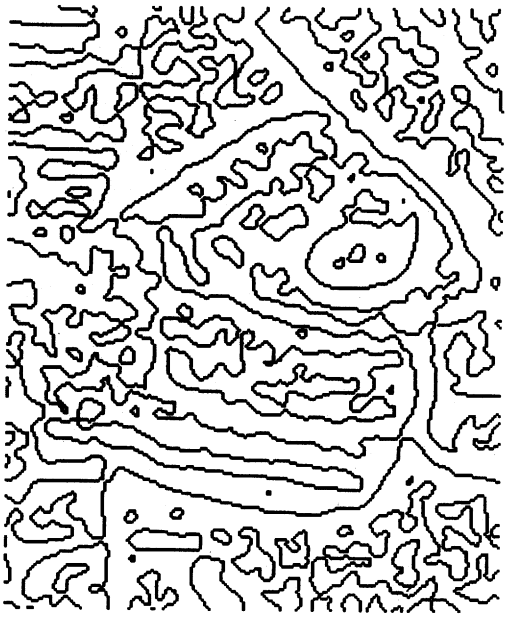
crossings, reflecting the noisy edges, are introduced.[3] The slopes associated with these noisy zero-crossings are comparable to those associated with significant edges, so that a simple thresholding technique can not distinguish between the two.

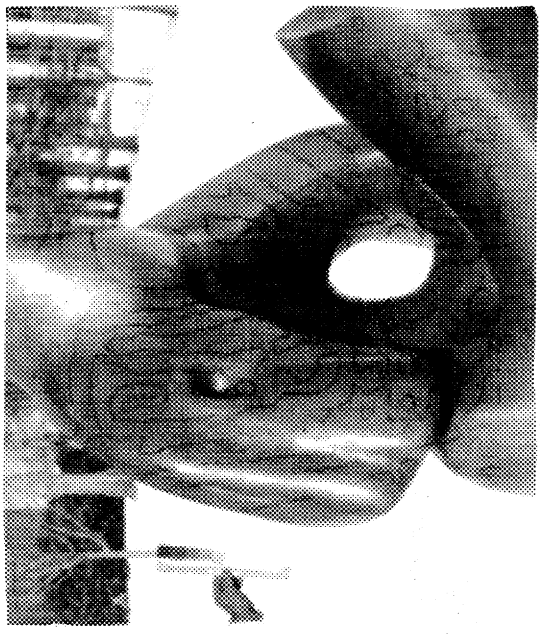
Secondary factors in limiting the size of the smallest operator are the combined effects of diffraction, discretization of the signal, and quantization of the intensities. For a particular combination of line spread, sampling interval, and quantization levels, there will be some scale at which two adjacent, discrete changes of the same contrast sign, are most likely to reflect the discretization of a single physical intensity change, or the spatial extent of a single physical mark on a surface. In both cases, the two changes together are likely to function as a unit under motion, or lie on the same depth plane for stereopsis, so computationally, it would be more efficient to describe them as a single primitive change. The finest resolution at which discrete changes are made explicit will be controlled by the size of the smallest mask.

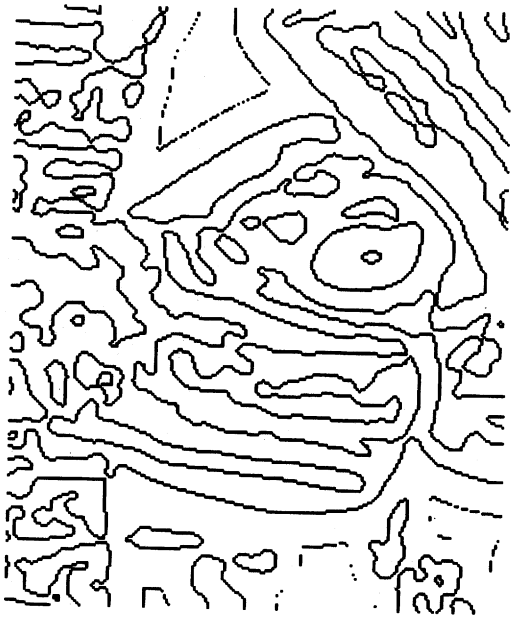
I have not defined a precise quantitative lower bound on size, because it varies with many factors of the imaging system. However, in many applications of edge detection, which I will discuss further in Section 6, particularly in the uses of the Laplacian (see, for example, Rosenfeld & Kak 1976, Pratt 1978), the operators have been extremely small, with a central diameter of one or two pixels. These tiny operators suffer from too much sensitivity to noise and quantization, and have been generally unsatisfactory for the reliable detection of edges. For our own imaging system, we have found that a ∇^2G operator of size $w = 6$ pixels (with a total support of roughly 900 pixels) adequately suppresses these noise factors. A final demonstration of the robustness of the larger ∇^2G operators is given in Figure 17. The original image, with its zero-crossings from the output of an operator with $w = 8$ pixels, is shown in Figure 17a. In Figure 17b, a large amount of Gaussian noise has been added to the image, and the

17. Demonstration of the robustness of a large $\nabla^2 G$ operator. We have the original image in (a), with the zero-crossings from the output of a $\nabla^2 G$ operator with w = 8 pixels. A large amount of Gaussian noise is added to the image in (b), and the image is quantized to 16 and 8 grey levels in (c) and (d). In each case, the zero-crossings from the same operator are displayed. Just as our own perception of the sculpture is robust, through these transformations, so are the resulting zero-crossing contours.



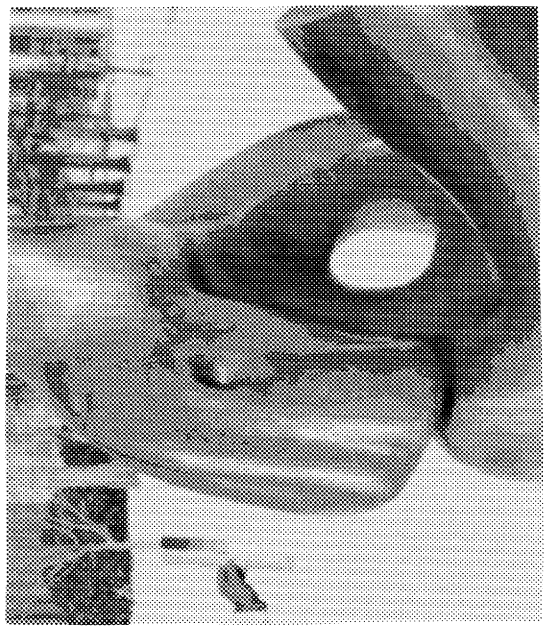


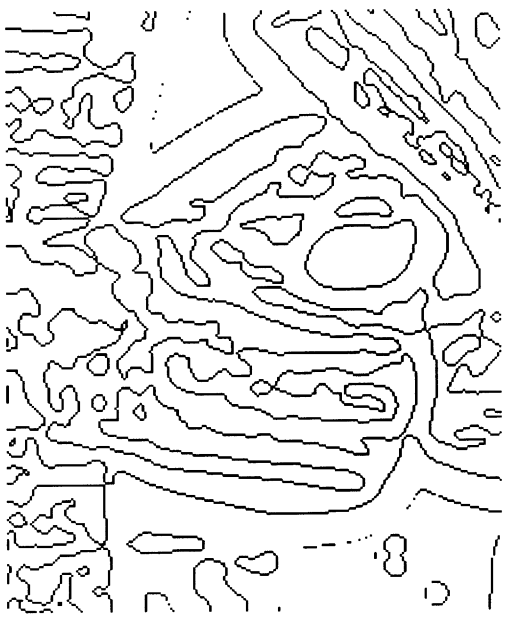




(17)

d.





c.

(17)

resulting zero-crossings displayed. Finally, in Figures 17c and d, the image is quantized to 16 and 8 intensity levels (the image was originally 256 intensity levels). The zero-crossings are fairly robust; their behavior does not break down with these changes in the image, just as our own perception of the images does not break down.

Our only constraint on the separation of sizes is that the sum of the Fourier spectra of the set of channels together be relatively flat, falling off only at the high and low frequency ends. The separation between sizes used here is one octave; chosen to provide sufficient sensitivity to all frequencies within the spectrum spanned by the set of operators, while minimizing the overlap of responses from adjacent channels. It is based on an analysis of peak frequency response and bandwidth of the difference of two Gaussian functions, which can be used to approximate $\nabla^2 G$ (Appendix B, Marr & Hildreth 1979).

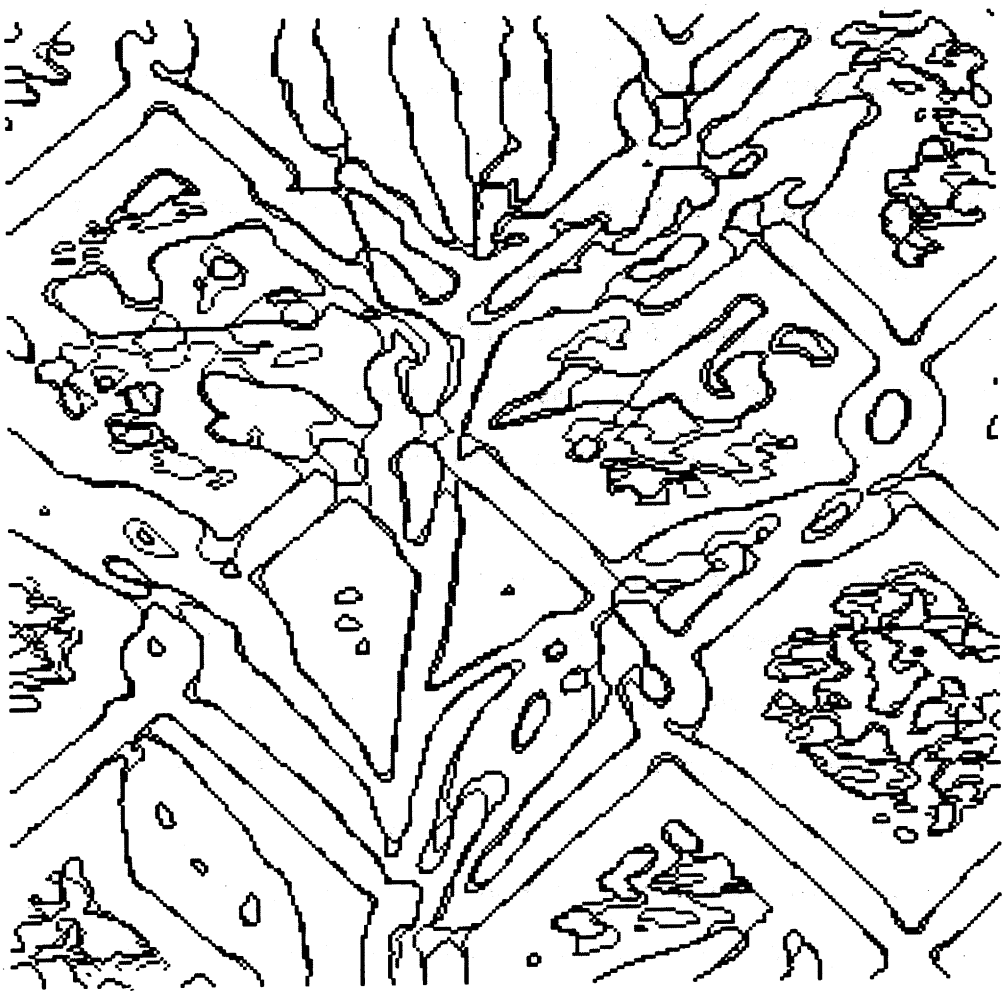
II.2 The Initial Assumptions

There are two conditions on the intensity function in the neighborhood of an intensity change which must be satisfied if we are to detect intensity changes by locating the zero-crossings in the output of a Laplacian operator. The first is the *Condition of Linear Variation* which states that the intensity function near and parallel to the line of zero-crossings should locally be linear. If this condition is satisfied, then for an edge whose orientation is θ, the directional derivative which yields a zero-crossing with maximum slope will also have orientation θ. The second condition requires that the intensity function be linear along, but not necessarily near, the line of zero-crossings. If this condition is satisfied, then the zero-crossings of the directional derivative measured perpendicular to the intensity change will coincide with the zero-crossings of the Laplacian operator. If the two conditions are satisfied, then an intensity change at any orientation in the image will give rise to a line of

zero-crossings in the Laplacian output, along the orientation of the change. In the development of the theory, it has been assumed that these conditions are generally satisfied in natural images.

One method for checking the validity of these assumptions is a direct statistical test of the conditions for a large number of intensity samples. Suppose we have an edge in a natural image, whose orientation is θ. To show that the *Condition of Linear Variation* holds, it would be necessary to show that within a narrow neighborhood of the zero-crossing segment in $\nabla^2 G$ to which it gives rise, $\partial(G*I)/\partial\theta$ is roughly constant everywhere. For the second condition to be true, this derivative must be constant only along the line corresponding to the position of the zero-crossing segment. A difficulty we have with such a statistical test is that at the resolution we are sampling these images, visual features are closely packed; in the smoothing process, interaction between nearby edges can yield strong variation in the convolution output between zero-crossings, which will negatively influence these statistics.

A second approach to testing these assumptions might be to test their consequence (Ullman, personal communication); that is, if an intensity change whose orientation is θ, satisfies the two conditions, then the zero-crossings of the directional derivative along the change will exactly coincide with the zero-crossings of the Laplacian. A possible test of the assumptions might then be to compute both a directional derivative and the Laplacian for an image, and compare the positions of the zero-crossings from the two outputs to see how closely they coincide for intensity changes whose orientation aligns with that of the operator orientation. Figure 18 illustrates this test. Superimposed are the zero-crossings in the outputs of the vertical operator described in Section II.1 (with aspect ratio 1:3), displayed in light grey, and the zero-crossings in the output of the $\nabla^2 G$ operator, displayed in medium grey. Zero-crossings which exactly coincide are displayed in black. Qualitatively, intensity changes whose



18. Comparison of zero-crossings in the output of the Laplacian operator with those of a directional derivative. Zero-crossings resulting from convolution with a vertical operator with aspect ratio 1:3 are displayed in light grey; those of the Laplacian in medium grey, and zero-crossings from the two outputs which exactly coincide are shown in black. Pieces of the zero-crossing contours in the Laplacian output whose orientation is within about 45° of the vertical correspond well with those of the directional derivative.

orientation is within about 45° of the vertical give rise to zero-crossings in the output of the two operators which correspond well.

II.3 The Final Channel Description

The remaining analysis in a single channel marks the beginning of the computation of a symbolic description of the low-level features in the image. We describe blobs formed by small contours; extended contours by a set of zero-crossing segments, each with an associated orientation, length, and average slope; and combine closely-spaced, parallel zero-crossing segments into descriptions of bars. Two motivations for making explicit elements such as blobs and bars is that first, their size is such that the elements most likely arise from a surface mark, so for the purposes of later processing, it can be treated as a primitive unit, which will be seen roughly the same by both eyes, and move as a unit, as the object moves. Second, due to the proximity of the intensity changes which give rise to bars and blobs at one scale, the zero-crossings from a larger channel are likely not to coincide with the zero-crossings of a smaller channel. Making explicit this proximity of intensity changes in one channel is useful for determining the reliability of zero-crossings from a larger channel.

To maintain the continuity and accuracy of position information provided by the initial zero-crossing contours, each channel description contains a binary map indicating the positions of zero-crossings. Attached to this map are the symbolic descriptors computed for pieces of the contour.

II.3.1 Zero-Crossing Segments

A zero-crossing segment is a line of zero-crossings whose local orientation is roughly uniform. Presently, we require that the length of these segments be at least w, so that we are measuring orientation over a significant distance. The maximum length has been chosen to be roughly 3w. This value is

motivated by studies of the human system (Nakayama & Roberts 1972), and is not a critical factor. In the implementation, the segment description is constructed as follows. The beginning of a zero-crossing contour which has not yet been described is located. Following the contour, a line of approximately w points is formed. The fit of the points to a line through their center of gravity is tested; if the fit is below a particular threshold, we continue to extend the line of points as long as its local orientation remains roughly uniform, up to a length 3w. If the fit is above this threshold, the initial segment is shifted around the contour, and the process continued. When a segment has been completed, a descriptor is placed at the middle of the set of points, with the following attributes:

1. type of descriptor: EDGE
2. average SLOPE of points along the contour.
3. LENGTH of the segment.
4. ORIENTATION, defined, from the horizontal, to be the orientation of the line which is a least squares fit to the set of points.

Figure 19b illustrates symbolically with oriented lines, the position and orientation of segments along the zero-crossing contours.

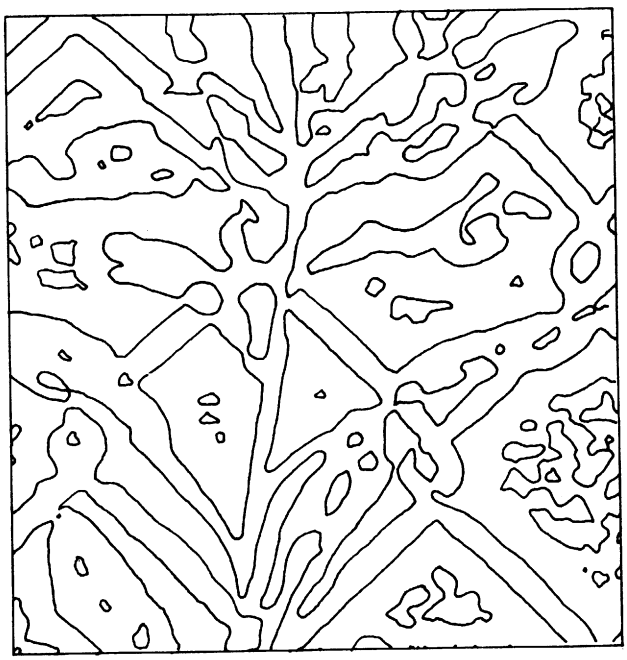
II.3.2 The Blob Description

To simplify the process of locating blob-like structures, any small closed contour whose total spatial extent fits within a square of size 3w x 3w is considered to be a blob. This particular size is again motivated by studies of the human system, and does not appear to be critical. A descriptor, with the following attributes, is placed at the point defining the center of gravity for points on the contour:

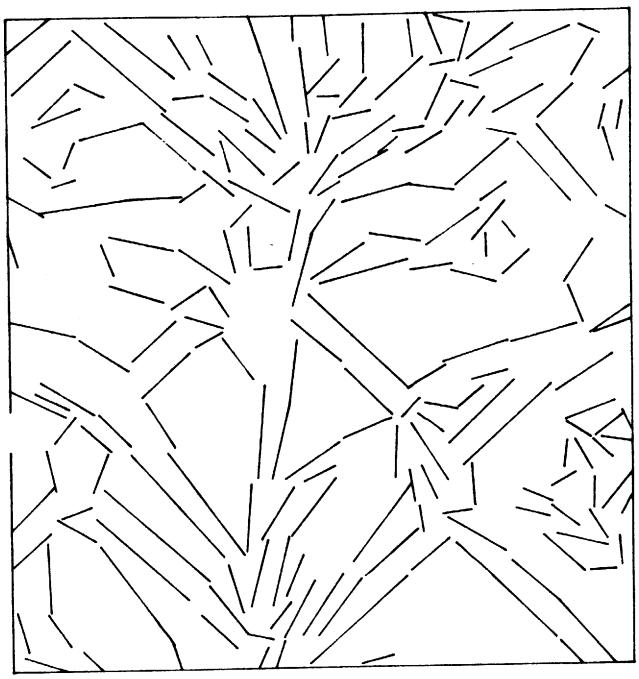
1. type of descriptor: BLOB

19. Single channel description. In (a) are the original zero-crossings for the plant image of Figure 1a. (b), (c), and (d) illustrate, symbolically, the descriptors attached to parts of the zero-crossing contours of (a).(b) illustrates the zero-crossing segments, in a manner similar to Figure 5. (c) shows the blobs; their rough spatial dimensions and orientation are represented in the dimensions and orientation of the rectangles. Finally, two oriented segments, which are roughly parallel form a bar. The bars in this case are shown symbolically in (d). Again, we are only conveying the spatial properties of the primitives here; also attached to the primitives would be the average slope of the convolution output across the primitive.

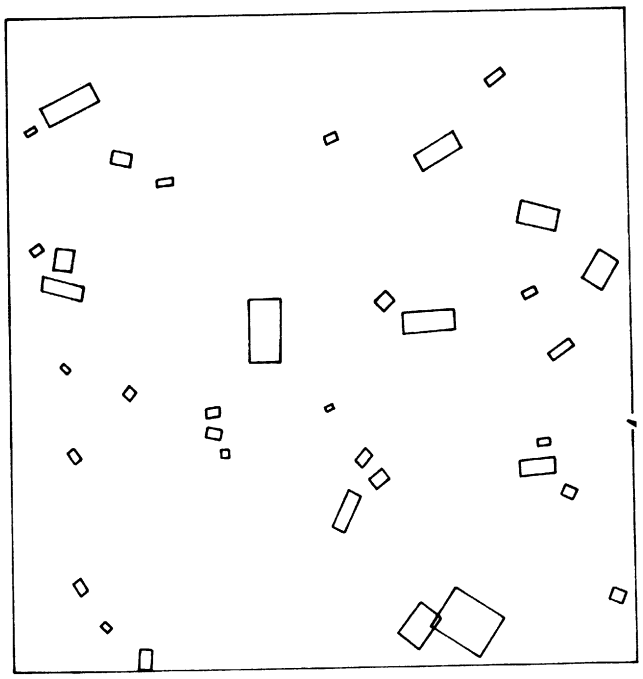
a.



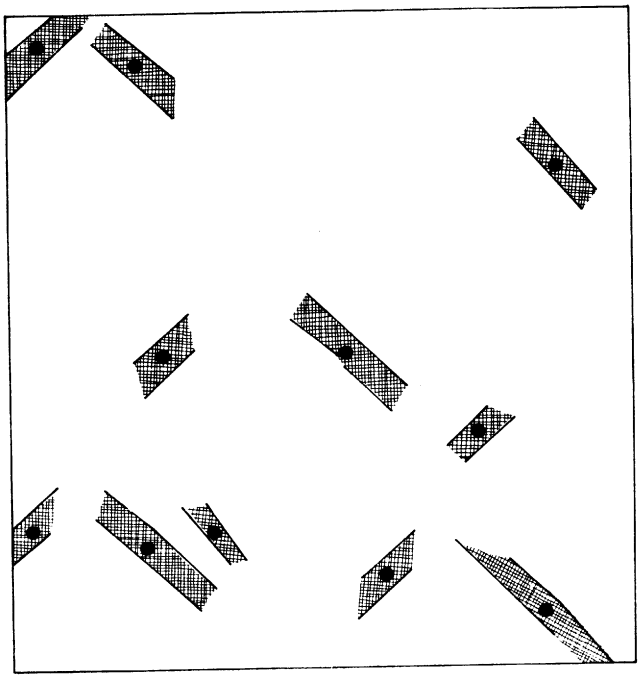
b.



c.



d.



(19)

2. average SLOPE of points along the contour, with its sign specified by the sign of the convolution output inside the closed contour.

3. ORIENTATION, which is always specified from the horizontal axis, and defined to be the orientation of the line through the center of gravity of the contour points which best fits these points (using a least squares fit).

4. LENGTH of the blob; a rough measure of the extent of the blob along its major axis.

5. WIDTH; a rough measure of the extent of the blob along its minor axis.

In Figure 19c, the blobs detected for the plant image of Figure 3a are represented symbolically with oriented rectangles.

II.3.3 Bars

A bar consists of two roughly parallel zero-crossing segments, separated by a distance about w, which extend over similar lengths. Bars are easily detected in a serial scheme by searching in a direction perpendicular to a particular segment for a second segment with similar orientation (presently, the difference in orientation is allowed to be as much as 5°). The associated descriptor, placed at the center of the two segments, maintains the following attributes:

1. type of descriptor: BAR

2. SLOPE of both segments, with the sign of the bar defined by the sign of the convolution output between the two segments.

3. average ORIENTATION of the two segments, again measured from the horizontal.

4. LENGTH of the bar.

5. WIDTH of the bar (separation between the two segments).

Examples of bars in the plant image appear in Figure 19d.

This computer implementation of the description of zero-crossing segments, blobs, and bars uses a serial, contour following scheme. A local, parallel scheme for detecting and describing these elements (Marr & Hildreth 1979), more attractive from a biological standpoint, is discussed further in the next section. The implementation of a parallel scheme is likely to raise issues of representation, and communication between local operators, which do not arise in this serial scheme.

There is also evidence that early in this detection stage, the human visual system makes explicit the termination of edges (see Section III), but the computational definition of terminations is still an open problem, which I will not be dealing with here.

III. Performance of the First Stage by the Human Visual System

Studies in neurophysiology and psychophysics offer much support for this first stage of early processing. Cell recordings in the retina of cat and monkey (Kuffler 1953, Rodieck & Stone 1965, Enroth-Cugell & Robson 1966, Cleland, Dubin & Levick 1971a, deMonasterio 1978a, 1978b) have uncovered two classes of retinal ganglion cells, whose axons form the optic nerve fibers, along which visual information is carried to the lateral geniculate nucleus, before reaching visual cortex. These two cell types, termed X and Y cells, both have receptive fields with an antagonistic center-surround organization, whose shape is a difference of two Gaussian functions (DOG).[4] The DOG is an approximation to the $\nabla^2 G$ function proposed here (see Appendix B, Marr & Hildreth 1979). X cells are distinguished by their smaller size, linearity, sustained response to changes in their visual input, and selectivity for color. In contrast, Y cells are larger, nonlinear, have a transient response pattern, and no wavelength specificity. These studies have also indicated that at each location on the retina, there is only one size for each cell class, although the size increases with eccentricity. Marr and Ullman (1979) have proposed that the primary function of the transient cells is to compute the time derivative of their input: $d/dt(\nabla^2 G * I)$. They provide an excellent demonstration of the close correspondence between the measured output of the retinal ganglion X and Y cells to the predicted responses of the proposed operators (Marr & Ullman 1979, p. 32, 37).

At layer 4c of visual cortex, which receives input fibers from the LGN, there is a greater scatter of receptive field sizes. The scatter seems to reflect a range of sizes of about 4:1 (Hubel & Wiesel 1974b, Figure 4). This scatter is believed to arise at the LGN (Cleland, Dubin, & Levick 1971b), where other properties are preserved. Computationally, it requires the output of only a few DOGS to yield a DOG which is twice the size (Marr & Ullman 1979). Simultaneous recordings from the retina and LGN (Cleland, Dubin, & Levick 1971b)

indicate that only a few ganglion cell inputs are necessary to account for the output of most LGN cells studied.

At the cortex, there is again a distinction of cell types, originally classified by Hubel & Wiesel (1962) as simple, complex, and hypercomplex. The class of primary interest here are the simple cells, which were described as linear, with 'bar' or 'edge' shaped receptive fields, clear orientation specificity,[5] smaller in size, and with distinct subregions of their receptive fields responding to light increment and light decrement. The size of their receptive fields increases with eccentricity. Subsequent studies (Schiller et al. 1976a) have further divided simple cells according to their selectivity for the direction of motion. They were also shown not to be strictly linear devices, as their orientation tuning did not change significantly with an increase in the strength of flanking subfields (Schiller et al. 1976b).

Marr & Hildreth (1979) first proposed that the function of simple cells is to detect segments of zero-crossings in the $\nabla^2 G$ output which forms the input to these cells. The model was elaborated by Marr and Ullman (1979) to include a mechanism for directional selectivity. It was also proposed that the function of 'bar' shaped simple cells is the detection of closely spaced zero-crossing segments. It is interesting that these 'bar' and 'edge' shaped cells first found by Hubel and Wiesel offered strong motivation for the first primitive operators proposed by Marr in the early primal sketch theory (Marr 1976b). Computational experiments later showed that if we apply these operators directly to the image, attempting to model the dimensions of the simple cells (the aspect ratio is roughly 4:1), the description we obtain for the changes in intensity are very unreliable. So it was found, computationally, to be necessary to separate the two tasks of initial filtering with a non-oriented operator, and detection of oriented intensity changes by an oriented operator, as the human system divides these two tasks. In the serial implementation described in the previous section,

the local detection of zero-crossings, and assignment of orientation to segments of the zero-crossing contour were separated. The simple cell operator combines these two tasks, and can be applied in parallel over the $\nabla^2 G$ output.

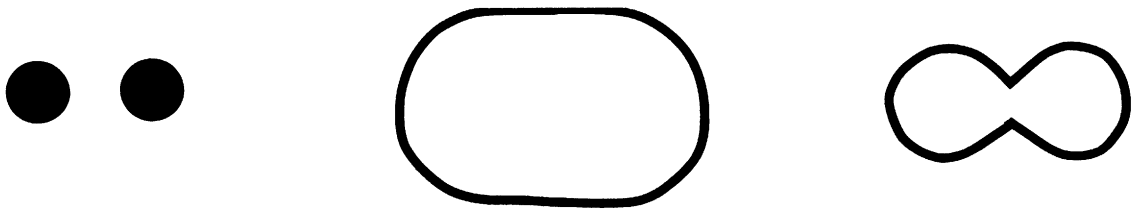
It has long been known, through psychophysical experimentation, that there exist separate, orientation sensitive channels through which visual information is processed, each tuned to different ranges of spatial frequency (for example, Campbell & Robson 1968, Blakemore & Campbell 1969, Sachs, Nachmias, & Robson 1971, Tolhurst 1973, Cowan 1977). There still remains considerable controversy concerning the number and frequency bandwidth of these channels. Wilson and Giese (1977) and Wilson and Bergen (1979) have recently completed quantitative studies which suggest that there exist four, fairly broadly tuned channels. In their studies they tested the contrast sensitivity of their subjects to particular vertical patterns, modulated with time functions intended to selectively stimulate the sustained and transient mechanisms. Two channels, one sustained and one transient, were measured directly by these experiments, while the other two were implied by general data concerning the overall modulation transfer function of the human system. The shape of the response of these channels can be realized by oriented operators whose horizontal cross-section is the difference of two Gaussian functions; for the smaller, sustained channels, the ratio of excitatory to inhibitory space constants is 1:1.75, and half-power bandwidth is roughly 1-3 octaves; for the larger, transient channels, this ratio is 1:3.0, and the bandwidth may be slightly larger. The four channel sizes are separated by about one octave, and increase linearly with eccentricity. The sizes at the fovea range from 3' of arc to 21' of arc. The above experiments used vertical patterns as stimuli; if the results represent processing through an initial circular center-surround operator before processing by oriented operators, the experiments must be interpreted as measuring the one-dimensional projection of these two-dimensional circular operators. The two-dimensional

counterpart of a D^2G operator with central width w_x is a ∇^2G operator with central diameter $w = \sqrt{2}w_x$. The measured sizes of the Wilson and Bergen channels then imply initial operators with diameters in the central fovea ranging from $3.1\sqrt{2} = 4.38'$ to $21\sqrt{2} = 29.69'$.

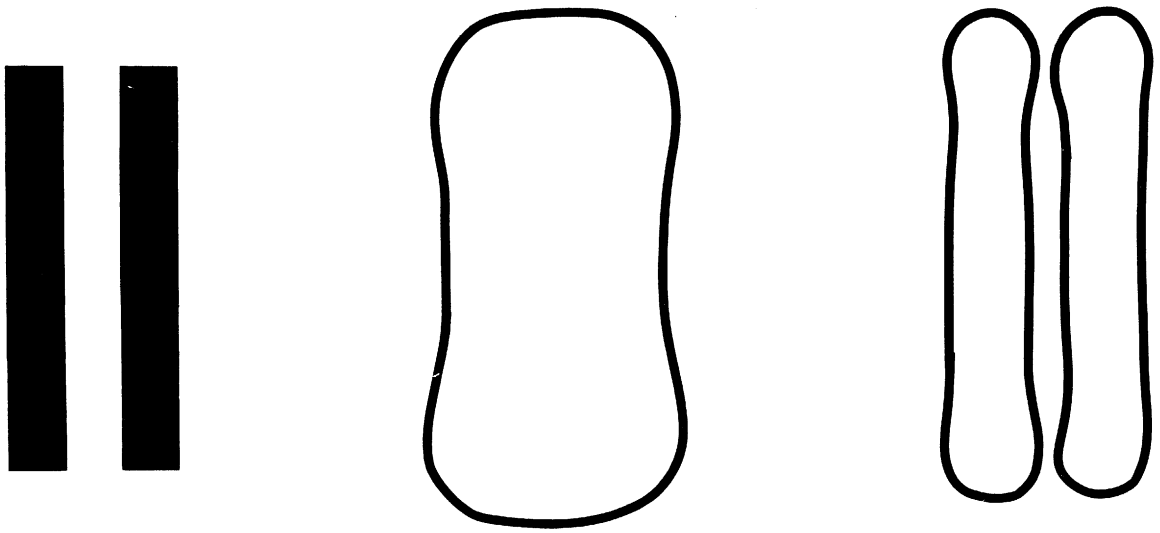
Psychophysical experimentation has also revealed a remarkable ability of the human visual system to resolve fine detail. The basic visual acuity experiments test our ability to resolve two fine dots or bars, with small separation between. Under the best viewing conditions, two dots, at maximum contrast and with diameter 1' of arc, can be resolved with 75% confidence at a separation of slightly more than a minute; two 1' wide bars can be resolved with 75% confidence at a separation of 1' (Westheimer 1977). It is interesting to ask whether such fine acuity can be explained with a smallest channel size of 3.1' proposed by Wilson & Bergen (1979). If our limit of resolution of the two dots or two bars were determined by the smallest separation which yielded distinct zero-crossings between the two elements, a smaller channel is required to account for the experimental threshold (Marr, Poggio, & Hildreth 1979). Figure 20 illustrates the zero-crossings from a 3.1' and 1.5' channel, given the two point and two bar configurations. Also illustrated are the intensity profiles for a cross-section of the bars example. Acuity experiments generally use a paradigm of forced choice, which means that the subject is presented with some set of configurations, one at a time, and is asked to respond in one of two ways, such as "one" or "two". One can argue that other available information, such as peaks in the convolution output, or a difference in the position of the outer zero-crossings, might be used in making these binary decisions. Other available information does not, however, appear to predict the experimental threshold as well as the use of distinct zero-crossings, and a smaller channel.[6] Because of the availability of other information, we can not consider these experiments as proof of the existence of a smaller channel.

20. Acuity experiments. In (a), we have represented two dots, at maximum contrast, with diameter 1' of arc, separated by 1' of arc. Similarly, in (b) are two 1' wide bars, separated by 1' of arc. In both cases, we have the zero-crossings from the output of these patterns convolved with $\nabla^2 G$ operators with w = 3.1' and 1.5'. (c) shows a cross-section of the intensity profile for the bars example. Our perception of these patterns corresponds well to what is seen by the smaller channel.

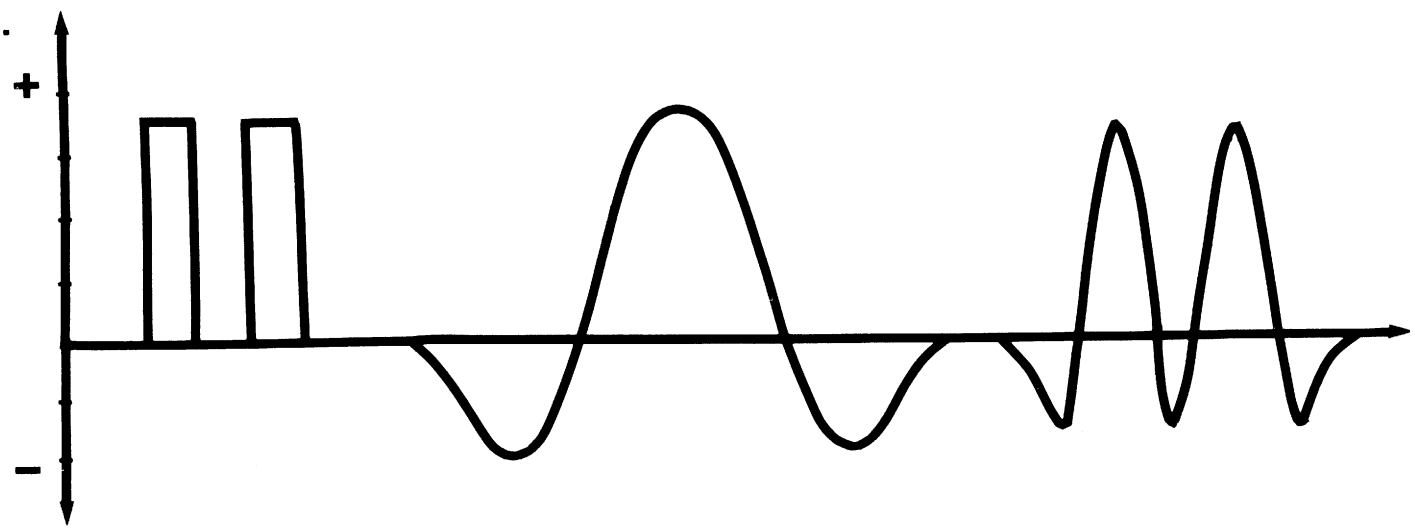
a.



b.



c.



(20)

It has been shown (Marr, Poggio, & Hildreth 1979) that a smallest channel size of 1.5' may also be reconciled with the basic optics of the eye, which places a theoretical limit on resolution, and known physiological data concerning midget ganglion cells in the retina, believed to be driven by a single cone.

There is a second type of visual acuity, labeled hyperacuity, which refers to an ability to make accurate judgements requiring the localization of some visual feature to a resolution of a few seconds of arc, roughly 1/5 the size of the finest foveal cones (Westheimer & McKee 1975, 1976, 1977, Westheimer 1976, 1977, Beck & Schwartz 1978, Burr 1979). The LGN input to visual cortex represents an even coarser sampling than the initial cones, providing samples roughly every 1' of arc in the central fovea (D. Marr, personal communication). Vernier acuity and stereo acuity fall into this hyperacuity range. Barlow (1979) suggested that a necessary component of this fine localization is the reconstruction of the smallest channel, at a higher resolution, after initial processing by the retina and LGN. The neurophysiology is particularly well-suited for this task; in layer 4c of visual cortex, where input fibers from the LGN are recieved, there is a myriad of granule cells which outnumber the input fibers 30-100 to one (see Barlow 1979). Barlow suggests that an interpolation takes place between the input samples, with the output of the granule cells representing the interpolated values. If this interpolation took place, the positions of the zero-crossings could be localized to a precision of a few seconds of arc (Crick, Marr & Poggio 1980, Marr, Poggio, & Hildreth 1979). This poses the question of how this interpolation might take place; that is, what interpolation function is best suited for computing the positions of the zero-crossings at this precision, with the least error. An important constraint on this interpolation scheme is that it be biologically feasible.

Statistical experiments, run on a wide range of intensity profiles, show that the interpolation can be performed simply and locally, with a range of

possible interpolation functions, yielding small error in positioning the zero-crossings. The precise experiments were as follows. Several one-dimensional intensity profiles were first convolved with a one-dimensional D^2G function with w_x ($w_x = 2\sigma$) corresponding to the 1.5' channel proposed. In two different experiments, the interval between samples of the intensity function (and D^2G output) roughly corresponded to 5" and 20", respectively (20" is the approximate spacing between foveal cones). The D^2G outputs were then sampled at intervals corresponding to 20", 40", and 60", and reconstructed at a 5" resolution. In the experiment where initial D^2G output samples were computed every 5", the reconstructed signals were compared against this original convolution output. In the second case (convolution samples originally every 20"), the reconstructions were compared against interpolation with an extended sinc function, whose frequency spectrum approaches the ideal lowpass filter required by the sampling theorem. Other interpolation functions were first, truncated sinc functions with n = 2, 4, and 6:

$$\begin{cases} (\sin(\pi x/int))/(\pi x/int) & |x| < 2n \\ 0 & \text{otherwise} \end{cases}$$

int is the spacing between input samples. The output was also convolved with a Gaussian function with space constant $\sigma = int/2$, and a triangular function of width $2*int$ (linear interpolation). As our main concern is the positioning of the zero-crossings, expected values for the deviation, in pixel units, of their position from the ideal positioning were computed; statistics for the second experiment appear in Table 1.[7] For the first experiment, the statistics were slightly better. Most striking about these figures is the fact that the deviations are quite small. Qualitatively, it was observed that for a relatively strong edge, positioning of the zero-crossings by all interpolation functions

Table I

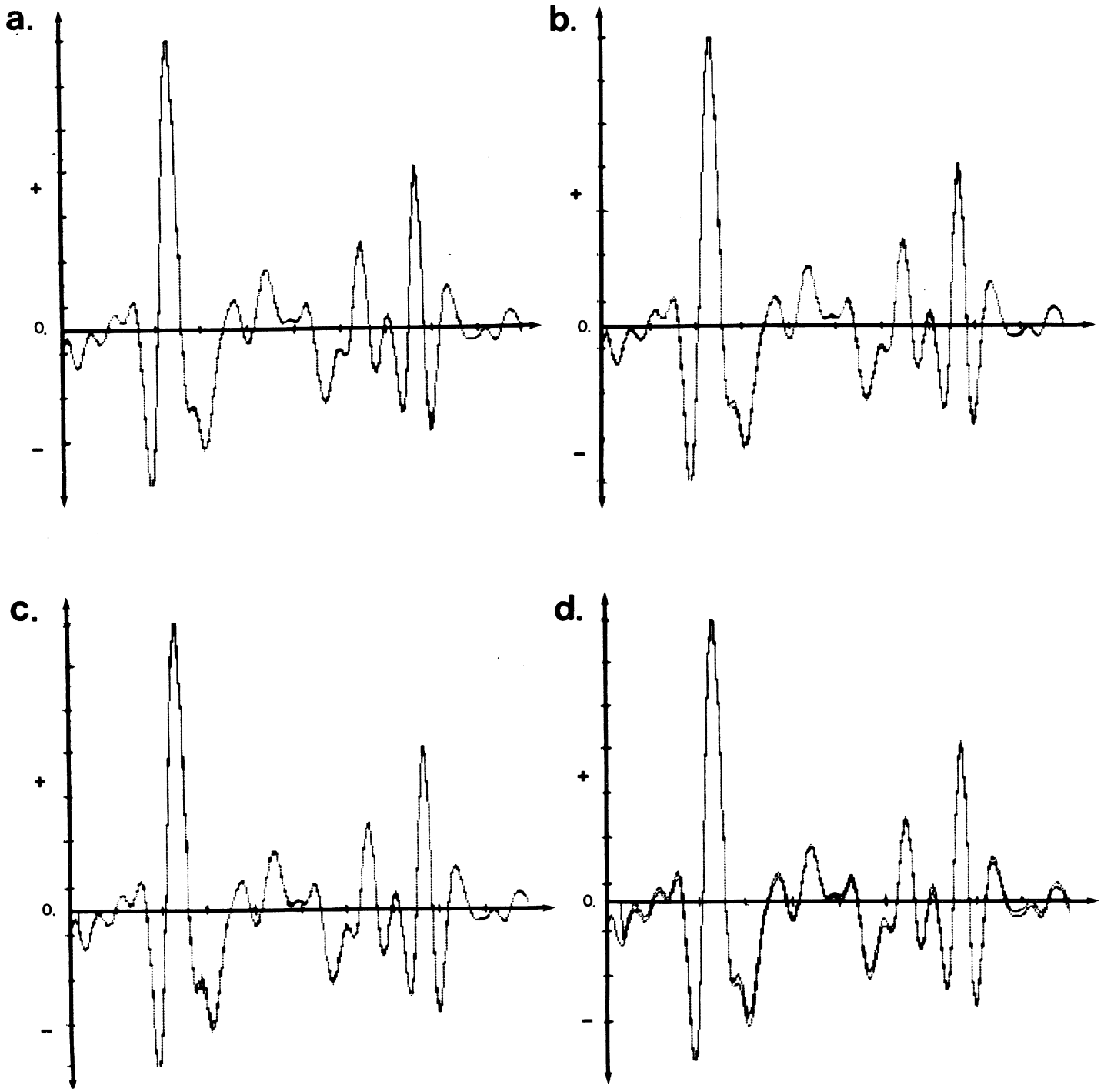
Case 1: no sampling of the convolution output			
truncated sinc, n = 2	0.037	5.45	-
linear	0.001	-	0.96
gaussian	0.005	-	1.92
Case 2: convolution sampled every 40 seconds			
truncated sinc, n = 2	0.523	6.19	1.03
truncated sinc, n = 4	0.410	3.09	1.03
truncated sinc, n = 6	0.479	3.09	1.03
linear	0.659	-	2.06
gaussian	0.889	-	5.15
Case 3: convolution sampled every 60 seconds			
truncated sinc, n = 2	0.855	2.50	8.75
truncated sinc, n = 4	0.798	6.25	3.75
truncated sinc, n = 6	0.729	6.25	3.75
linear	0.841	-	10.01
gaussian	0.961	1.25	11.25

Comparison of the reconstruction of D^2G outputs using different interpolation functions. The first column provides the mean displacement of the zero-crossings, in pixel units (which corresponded roughly to 5" spacing between samples), between the given interpolation function, and the extended sinc (used as the comparison interpolation). In the second and third columns are the percentage of additional zero-crossings in the interpolated output, and missing zero-crossings, respectively.

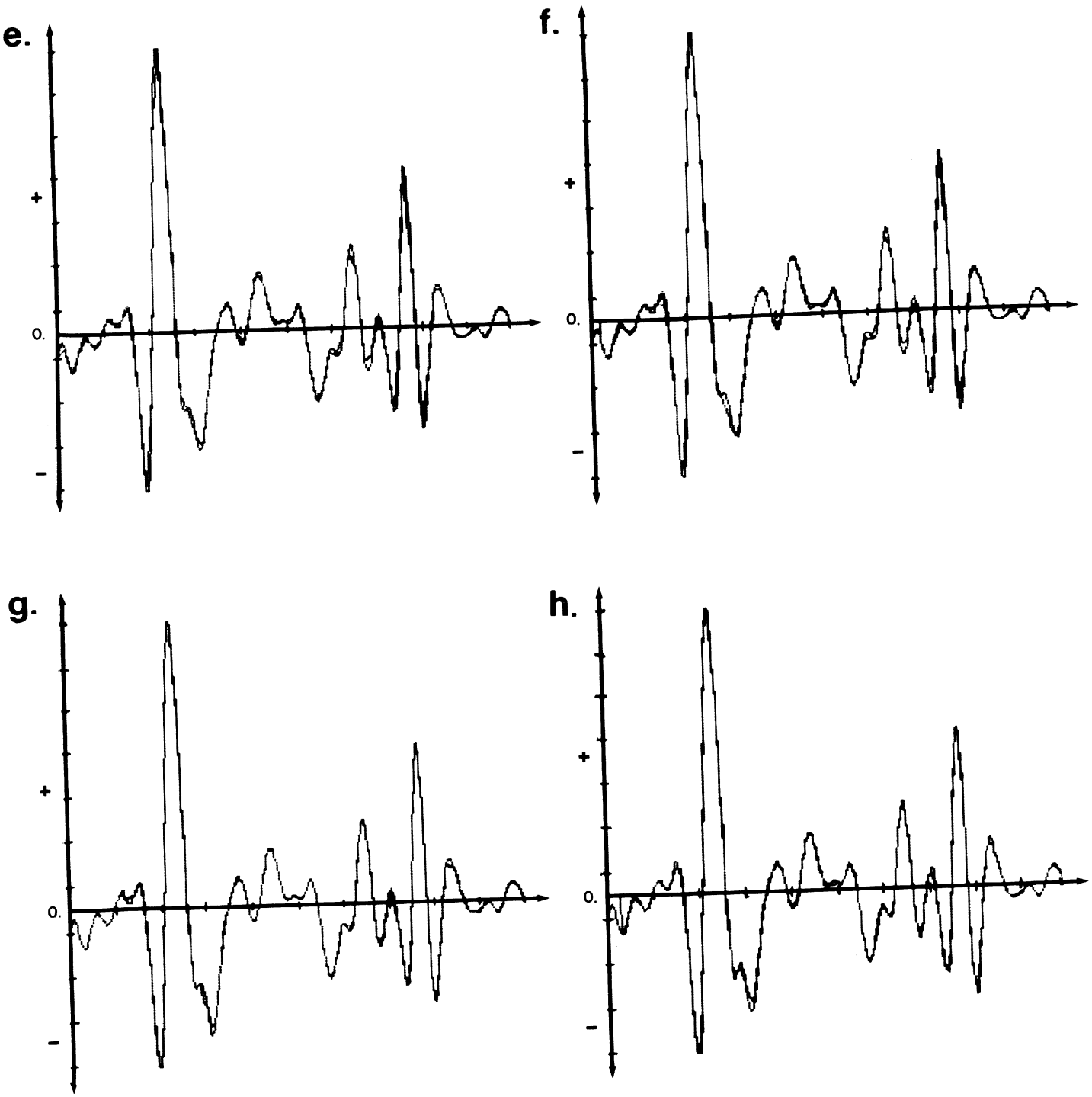
exactly coincided with the ideal position, at these precisions. Deviation from the ideal position generally occured in regions of high-frequency, low contrast change. The stimuli typically used in hyperacuity experiments would not distinguish between the possible interpolation functions. Some examples of the reconstructed signals appear in Figure 21.

For the more interesting case of reconstructing the outputs after sampling every 40" and 60" (closer to the sampling of LGN inputs in the human system), the truncated sincs provide the least error, with the error not decreasing significantly as we increase n. One factor in considering the biological feasibility of these interpolation functions is the size of the support required for their computation; linear and Gaussian interpolation are more local operations, requiring only two input samples in the one-dimensional case, while the truncated sinc (with n = 2), requires four. For simple intensity distributions, such as an ideal step, the sinc functions would yield shallow secondary zero-crossings to the side of the primary zero-crossing marking the position of the edge. In terms of a biological implementation, whether or not this poses a real problem would depend on whether interpolation takes place before or after (or simultaneously with) the detection of zero-crossings by the more coarsely spaced simple cells. On the other hand, with slightly larger error, one could use Gaussian or linear interpolation, which would not introduce these secondary zero-crossings.

To summarize this discussion, in order to account for the experimentally determined visual acuity and hyperacuity of the human visual system, the combination of a smaller, 1.5' channel operating in the central fovea, and interpolation of the $\nabla^2 G$ samples had been proposed. With the smaller channel, acuity is determined by the ability to discriminate visual input by the presence of distinct zero-crossings in the $\nabla^2 G$ output. Concerning interpolation, it has been shown that very simple, local functions are adequate to account for the



21. Reconstructed convolution outputs. An ideal reconstruction of a convolution output, at 5 times the original resolution of the output, is shown in (a). In (b), (c), and (d), the output was reconstructed (without sampling) by a truncated sinc with n = 2, linear interpolation, and a Gaussian, respectively. The results are shown superimposed on the ideal reconstruction. We can see that the error in positioning the zero-crossings is small. In (e), (f), (g), and (h), the output was sampled every other point, and then reconstructed using truncated sincs with n = 4 and 2, linear interpolation, and a Gaussian, respectively. Again, the outputs are shown superimposed on the ideal reconstruction of (a). The error in positioning the zero-crossings does not increase significantly.


e.
+
0.
-
f.
+
0.
-
g.
+
0.
-
h.
+
0.
-

hyperacuity of the human visual system.

There is also evidence from psychophysics that the human system makes explicit termination points, and that these primitives may at least be used in stereopsis (Frisby & Julesz 1976) and motion correspondence (Ullman 1979). Evidence so far only shows the endpoints of lines to be early primitives. For the special case of lines, even a simple, logical operator, designed in a similar manner as the simple cell operator, but whose optimal stimulus is a blob-like structure, would be able to detect their endpoints. However, further investigation is needed in this area, to determine whether terminations should be extended to a wider class of elements, and how they may be detected in the general case.

In conclusion, there are many parallels between this first stage of the primal sketch theory, and processing in the human system. Historically, the finding of cortical simple cells which behaved roughly as bar or edge-shaped differential operators provided motivation for the original primal sketch theory. Later, the distinction between the initial differential operation performed by the retina, and the role of simple cells, provided support for dividing these tasks in the computation. The known shape of the response of retinal ganglion cells, and the shape of the channels found by Wilson and Giese (1977), and Wilson and Bergen (1979) motivated a careful study of the necessary constraints on shape, leading to the requirements of localization in space and frequency. Given the general framework of the channels, the implementation then became a testing ground for phenomena such as acuity, leading to the suggestion of a smaller channel in the human visual system, and the proposal of simple interpolation schemes for the fine positioning of zero-crossing contours. It is this close interaction between the computational model, and the human processor which makes this area of research so exciting.

IV. Combining the Channels

In Section I, I described the second stage of the primal sketch computation: the integration of individual channel descriptions into a single raw primal sketch of the image. In Section IV.1, I will first discuss some of the known relationships that exist between zero-crossings in adjacent channels. The spatial coincidence assumption captures their spatial relationship; here, we also look at the slopes of coincident zero-crossings, and relate them to the properties of width and contrast of the original intensity change. These properties can be recovered easily for simple, relatively isolated intensity changes; this work is not able to offer a solution to the recovery of these properties in the general case of intensity changes for which there is complex spatial interaction between nearby intensity changes, or changes operating at different scales. In this case, I will only suggest additional sources of information which might contribute toward this recovery. The results of this first section are relevant to the integration of information between channels, regardless of whether we explicitly represent this information in a single description of intensity changes. In Section IV.2, I will examine some perceptual experiments relating to this question of combining the channels. Grouping, lightness computations, and motion correspondence will each impose requirements on the description of intensity changes which forms their input. In Section IV.3, it is suggested that understanding these requirements may be necessary for taking the work on channel integration further.

IV.1 Relationships Between the Channel Outputs

We can begin to study the relationship between zero-crossings from different channels by first examining the case of one-dimensional signals. Here, our input will be the zero-crossings and slopes from the output of $D^2G(x)*I(x)$, where D^2G is now the one-dimensional operator of Figure 2a. We can assume for

now that the finest resolution with which we can resolve individual changes in the intensity profile is the resolution provided by the smallest channel; we can not distinguish two nearby intensity changes which are not represented by distinct zero-crossings in the smallest channel. If we let the convolution output of the smallest channel be $O_1(x)$, then the output of the larger channels is just the result of Gaussian filtering $O_1(x)$ (by the derivative rule for convolutions). We are interested in understanding the inverse of this operation; that is, given local zero-crossing information from a set of channels, each of which is the result of additional smoothing of a single intensity profile, what can we say about the changes occuring in the original profile?

We can begin a systematic analysis of the zero-crossings with the case of a single, isolated edge in the image. Here the relationship between the zero-crossings in different channels is simple; there will exist a single zero-crossing at the location marking the middle of the edge in each channel output, whose slope is related to the contrast and width of the edge, and size of the operator, by the following equation:

$$s_i = ce^{-d^2/8\sigma_i^2}$$

c and d are the edge contrast and width, and σ_i is the space constant for the Gaussian used in forming the D^2G operator. Thus given zero-crossings from two channels, we have two equations which we can solve for c and d:

$$s_1 = ce^{-d^2/8\sigma_1^2}$$
$$s_2 = ce^{-d^2/8\sigma_2^2}$$

If we choose these channels to be separated in size by one octave, as in the case of adjacent channels used in the primal sketch computation, then $\sigma_2 = 2\sigma_1$. We are

dealing here with discrete, quantized signals, which pose an additional constraint on this solution. If d is very large in comparison with σ, there will be a neighborhood in which the convolution output will fluctuate around zero, yielding multiple zero-crossings, rather than a single zero-crossing at the location corresponding to the midpoint of the original edge. For this situation, illustrated in Figure 22, the above equation no longer holds. In order to solve for contrast and width using zero-crossings from adjacent channels, we must be careful to choose those channels for which w, the size of the filter, is close to d. For this we can use the *selection criterion* (Marr 1976b, Marr & Hildreth 1979). Using the above equation, we can relate the ratio of slopes of zero-crossings from adjacent channels to d:

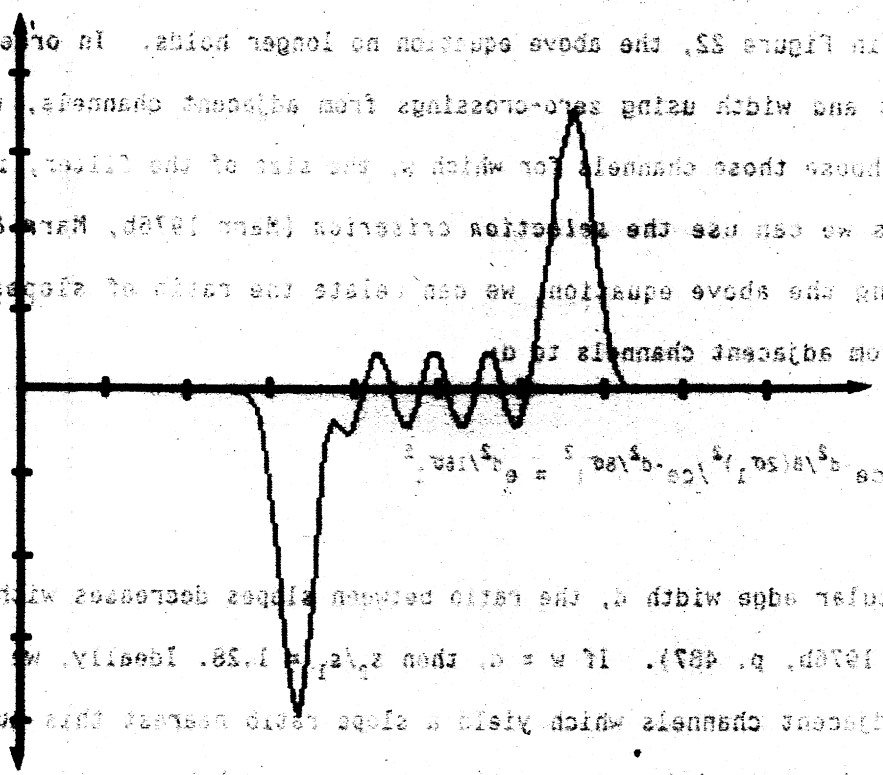
$$s_2/s_1 = ce^{-d^2/8(2\sigma_1)^2}/ce^{-d^2/8\sigma_1^2} = e^{d^2/16\sigma_1^2}$$

For a particular edge width d, the ratio between slopes decreases with mask size (as in Marr 1976b, p. 487). If w = d, then $s_2/s_1 = 1.28$. Ideally, we would like to choose adjacent channels which yield a slope ratio nearest this number. The requirement that w = d is conservative, however, and in practice we can use somewhat smaller channels for computing contrast and width.

The two-dimensional extension to the computation of contrast and width is straightforward; in this case, the slope will also be a function of the two-dimensional extent of the mask, here labeled x:[8]

$$s_i = cxe^{-d^2/4\sigma_i^2}$$

As before, we can then compute c and d explicitly from the measured slopes from two different size channels, as long as their size satisfies the selection criterion.

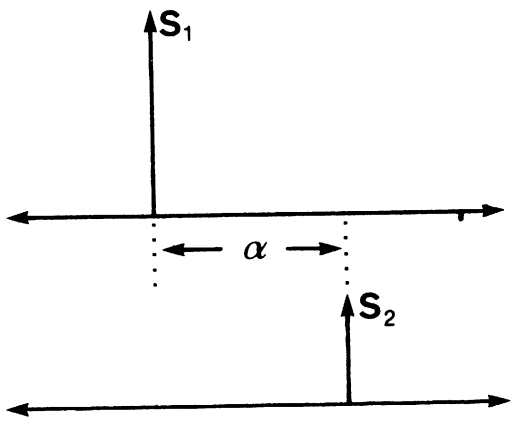


22. The breakdown of the zero-crossings when the mask is small in comparison to edge width. Due to quantization and discretization, a small mask will give a noisy response to even very wide edges, yielding multiple zero-crossings, as shown here.

Using the above scheme for computing contrast and width from the slopes of zero-crossings, we can propose a simple algorithm for combining two channel descriptions, which rests heavily on the spatial coincidence assumption described in Section I. The algorithm compares the position and orientation of the zero-crossing segments contained in corresponding neighborhoods of the two arrays representing two adjacent channels. If similar segments are found in the two neighborhoods, the descriptions are integrated into a single edge description, which now includes the calculated contrast and width, and together with the piece of zero-crossing contour described, this description is written into the array containing the final raw primal sketch. If a segment is found in only one channel, it may be written into the raw primal sketch independently.

The above algorithm is adequate for simple scenes consisting of relatively isolated,[9] simple intensity changes, but unfortunately, many edges in the real world will not be sufficiently isolated for this scheme to provide an accurate computation of the edge properties. As operator size increases, and the image is smoothed over larger areas, distinct intensity changes will be smoothed together, while smaller changes may disappear entirely. There is a corresponding change in the position and slope of the zero-crossings, and the above scheme for computing contrast and width is quite sensitive to these changes.[10] We are faced with a situation where the position and slope of local zero-crossings no longer have a simple dependence on the properties of a single, local edge, but also depend on the properties of nearby edges. We would like our computation of edge properties to remain a local operation. One remaining question is then whether we can compute, using local operations, a good approximation for the properties of a particular edge, in the case where the response of the D^2G operator to nearby edges interacts with the response to the local edge.

Suppose that we have zero-crossings in two adjacent channels, whose positions are separated by some small amount α:



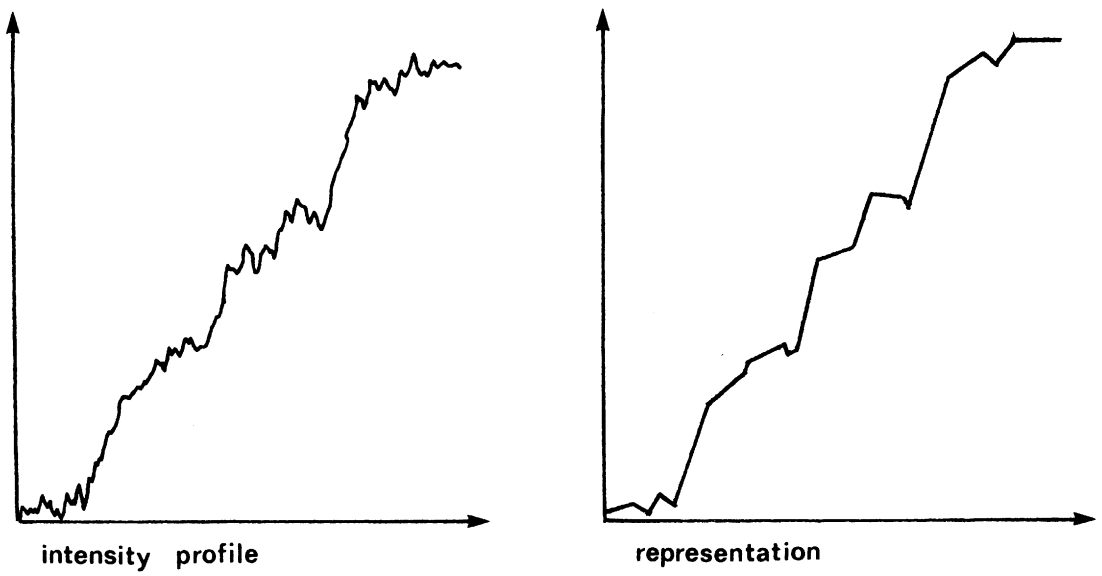
The spatial coincidence assumption allows us to conclude that there exists a physical edge somewhere in this vicinity; however, the three pieces of information given here (s_1, s_2, α) are not sufficient to unambiguously determine the contrast and width of this physical edge. There is a family of possible local edges (each with a different position, contrast, and width), which combined with nearby edges, could have given rise to this local information. To resolve the properties of the local edge, we need additional information. There are at least three potential sources of information which might help in the solution of this problem:

1. Properties of other zero-crossings within a neighborhood of these; the BAR operator described in Marr (1976b) and Marr & Hildreth (1979) is a possible mechanism for detecting closely spaced zero-crossings.
2. Information from other channels.
3. Assumptions about the local nature of intensity profiles, which are generally valid for natural scenes. An example of such an assumption might concern the rate at which intensity can change locally.

The extent to which each of these sources of information might be used, and how they are used remains an open question. A difficulty we have in advancing this work is an insufficient understanding of what is required by lightness-related computations. One might expect that for grouping, or motion correspondence, a measurement of contrast need not be very accurate. In the case of lightness

computations, however, are these measurements of contrast and width the appropriate quantities to make explicit at this stage, and how accurately must these properties be computed?

I would like to note that if we are to describe intensity changes occuring at a range of scales in a single representation, the design of this representation is now an issue. Marr and Nishihara (1977) presented a number of general criteria for judging a shape representation, of which three may be relevant here: (1) accessibility (computability) and (2) stability and sensitivity, and (3) organization of primitives within the representation. The meaning of accessibility in this case is clear. The second criterion, applied here, means that we would like the representation to be sensitive to small changes in intensity, but minor fluctuations in intensity, due to noise, or small illumination changes (such as would occur in time) should not significantly disrupt our description. For the organization of our primitives, we have at least two possible approaches; at one extreme is no organization; that is, all intensity changes in the description have the same status. The changes might be represented somewhat like this:



Each individual channel provides a description of intensity changes at a

particular scale, in a representation much like this; as a series of adjacent, non-overlapping changes. (For example, in Figure 1b, we showed the smoothed intensity profiles seen through different channels. There would occur a zero-crossing in the output of the second derivative of this profile at each of its inflection points. One could then think of the profile as a series of adjacent, non-overlapping intensity changes, whose location may be marked by the location of an inflection point in intensity.) This is also the simplest representation into which one could place the description obtained by integrating the channels using the local scheme discussed previously, which relies primarily on spatial coincidence. That is, our ultimate description most closely resembles the structure of the description obtained from the smallest channel, but with contrast and slope of the intensity changes made explicit, using the slopes of coincident zero-crossing contours from larger channels. It is yet unclear, using a representation such as this, how one would include the description of an intensity change which was only detected at one of the larger scales (such as an explicit shading edge). A second approach for the design of this intermediate representation is a hierarchical structure; at some level, we describe overall changes that take place; each overall change summarizes a group of smaller changes which take place at another scale. For the above intensity profile, the hierarchical organization would place a description of the overall rise in intensity at one level, and descriptions of the smaller changes superimposed on this change, at another level. The separate channels already provide one such intermediate hierarchical organization of intensity changes. In Section IV.3, there is an example of an image of a leaf, in which larger channel zero-crossing contours explicitly follow shading lines caused by the nature of the illumination. It is suggested there that for some processes, it might be more useful to maintain the distinction of channels.

To summarize, if we want to characterize an intensity change by the

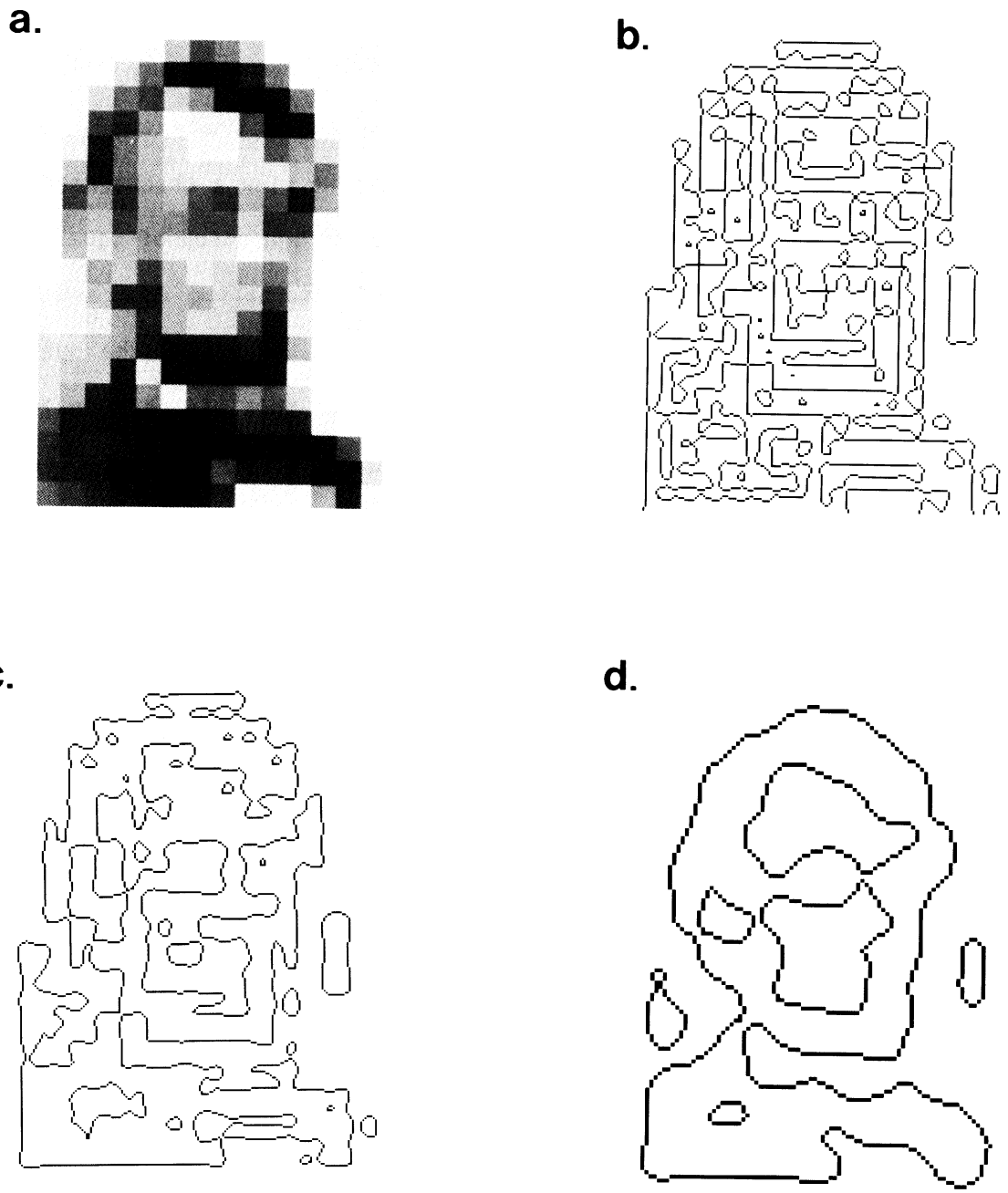
properties of contrast and width, we require zero-crossing information from more than one channel. Careful examination of the simpler case of one-dimensional signals contributed toward an understanding of the details involved in computing the needed edge properties. In the idealized situation of a single, isolated edge, this computation is simple, and we could develop an algorithm for combining channel descriptions, based on spatial coincidence, and our simple scheme for recovering contrast and width. Such an algorithm is adequate for simple images, and might be useful in certain visual processing tasks, but in natural images, local zero-crossings will generally not depend simply on the properties of a single edge, so other sources of information must be used to recover these properties. In the following sections, some perceptual experiments, and the role of subsequent processing in carrying this work further are discussed.

IV.2 Some Perceptual Experiments

We have a precise model of the shape and size of the individual channels, and propose that it is the zero-crossings in the output of these channels that carry essential information to subsequent processes. With this model, we can explore some human perceptual experiments by computing the descriptions we believe to be obtained by the individual channels, and observing the way in which they relate to our eventual perception. If this perception were the consequence of low-level interactions between channels, understanding the way in which they must interact in order to yield this perception might be useful. If we then have some understanding of how the channels interact in the human visual system, it might contribute toward an understanding of why they should interact in this way. Experimenting with this approach must, however, be taken with extreme care, as it will not necessarily be the low-level interactions between channels which account for this eventual perception. Two experiments which I explore here are Harmon and Julesz' (1973) quantized Lincoln figure, and our perception of a checkerboard

pattern.

In their 1973 paper, Harmon and Julesz present a quantized portrait of Abraham Lincoln, reproduced here in Figure 23a. The interesting property of this picture is that from the block portrait, we are unable to recognize Lincoln; however, if we squint our eyes, or stand a far distance from the image, we do recognize the face. Depicted in Figures 23b, c, and d are the zero-crossings in the output of three different size $\nabla^2 G$ channels. If we let the size of a single block be denoted by b, then the three mask sizes are b/2, b, and 2*b in Figures 23b, c, and d, respectively. They correspond roughly to what is seen by the human channels viewing the picture from a distance of 6 feet (in this case, the blocks would subtend about 6' of arc). It is clear from observing these outputs that the smaller channels provide information which emphasizes the quantization of the image, which closely resembles what we perceive at shorter viewing distances. However, for viewing distances greater than about 3 feet, larger channels provide zero-crossing information from which we could recognize Lincoln. The implication of this demonstration is that we have available to us, information from the larger channels which is not utilized in our final description of this image. We can not conclude where, or how, this apparent selection of the smaller channels takes place, because we are looking at the behavior of the entire visual system, in which a number of processes are acting together, either in sequence, or in parallel. (For example, it might be the case that in producing a description such as the raw primal sketch, which is concerned primarily with static, monocular analysis of changes in intensity, the smaller, sustained channels play a more significant role. The larger, transient channels may be more important for stereopsis and early motion analysis.) The system, as a whole, may be too complex to draw very detailed conclusions from this type of experiment, but I feel we can safely say that at some stage in processing the image, channel information is integrated, and once this stage is reached, we no



23. The quantized portrait of Abraham Lincoln, from Harmon & Julesz (1973). Block size in this experiment was 24 pixels. (b), (c), and (d) show the zero-crossings in the output of the block portrait convolved with $\nabla^2 G$ filters with w = 12, 24, and 48 pixels, respectively.

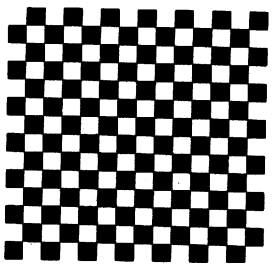
longer have access to the output of the individual channels. That is, the integration of the channels is an aspect of our nonattentive vision, and is most likely to preceed recognition.

The second experiment is simply a more objective example of the phenomenon illustrated by the Lincoln example. When we view a checkerboard from a close distance, such as in Figure 24a, we see the configuration of squares. As we increase our viewing distance, the diagonals of alternating contrast become more apparent. This can be seen in Figure 24b. Again, examining the output of the different size channels offers one explanation for this effect. In Figures 24c, d, and e, we have the output of $\nabla^2 G$ filters applied to a checkerboard pattern, with w = b/2, b, and 2*b (b is again the block size). The convolution outputs are shown in the first column; in the second column are the contours of zero-crossings, with the slope of the zero-crossings displayed as intensity. I chose here to display positive and negative contrast by light and dark intensity values. The particular convention for defining the direction of positive and negative contrast is arbitrary. Column 3 indicates all zero-crossing contours, displayed at constant intensity levels, and finally, the last column provides cross-sections of the convolution outputs across the zero-crossing contours. What we can observe here is that for all channel sizes, zero-crossings occur along the outline of the squares, but as we increase the channel size, that part of the zero-crossing contour with strongest slope becomes focused more narrowly around the midpoint of the edges dividing two squares. For very large channels, the slopes of the zero-crossing contours are extremely weak around the edge intersections.

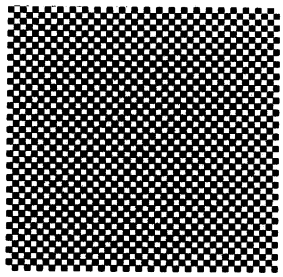
If we assume that oriented intensity changes are detected by simple cells operating at roughly the same scale as their $\nabla^2 G$ inputs, then for smaller channels, the horizontal and vertical zero-crossing detectors would respond very strongly, producing a description of the pattern of squares, illustrated

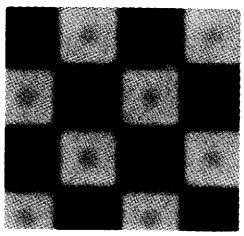
24. Analysis of the checkerboard pattern through different size operators. (a) and (b) show the checkerboard at two different resolutions. In (a) we clearly see the pattern of squares, whereas in (b), the diagonals dominate our perception. Block size for the experiment was 24 pixels. In (c), (d), and (e) we have the analysis of zero-crossings from filters of size w = 12, 24, and 48 pixels, respectively. In the first column are the convolution outputs. The second column shows the zero-crossings, with slope displayed as intensity (light and dark intensities represent positive and negative contrast; the particular convention used to define the direction of these contrasts is arbitrary). In the third column, we have all the zero-crossings displayed at uniform intensity, and finally, the fourth column provides a cross-section of the convolution output near the zero-crossing contours. (f) and (g) indicate symbolically, the description obtained by channels much larger, and much smaller, than the block size.

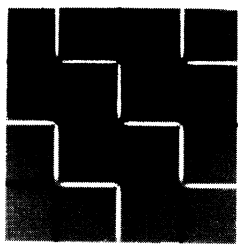
a.

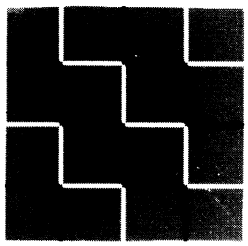


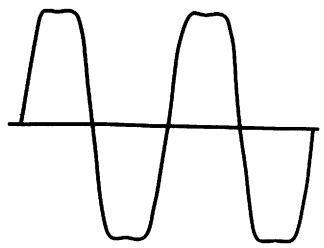
b.

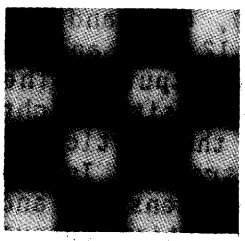


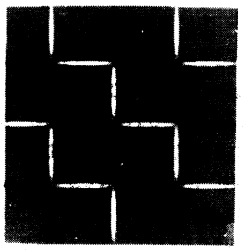


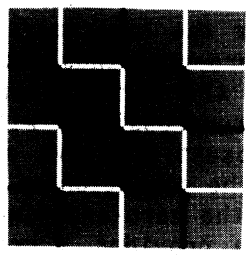


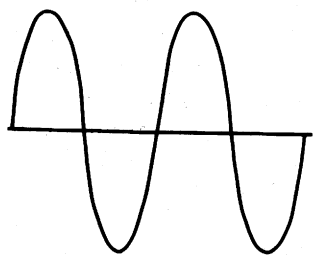


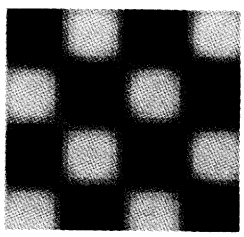


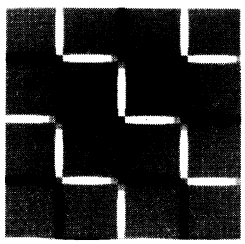


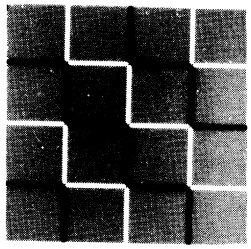


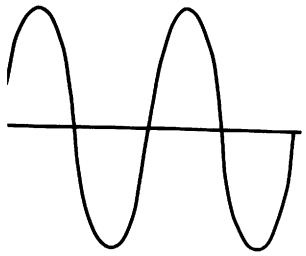




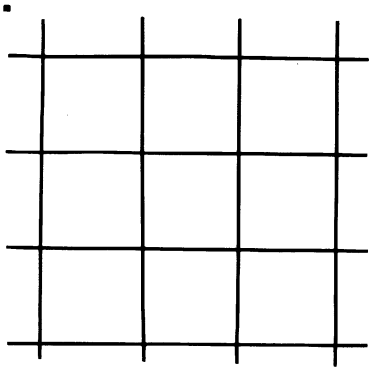




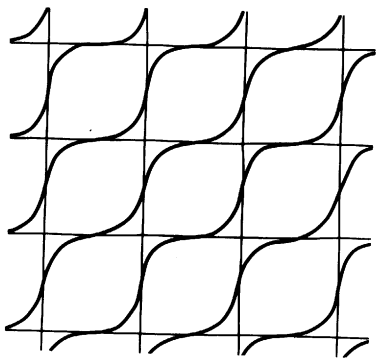




f.



g.



symbolically in Figure 24f. However, if we increase channel size with respect to block size, it is the detectors oriented along the orientation of the diagonals which would respond more strongly, even though the zero-crossing contours themselves still form a grid. The resulting description is shown symbolically in Figure 24g, for one diagonal direction. The fact that the difference in perception of the checkerboard pattern is related to what may be seen by different size channels, is, by itself, not so interesting. What is more interesting is the distance at which our perception changes from the pattern of squares, to the diagonals. This distance corresponds roughly to where the individual squares subtend about 3' of arc. This implies that there exist distances from which we may view the checkerboard, and see only the pattern of squares, while we have larger channels supplying information from which the pattern of diagonals would be stronger. It is analogous to the Lincoln picture, in which we have information from the larger channels which appears not to form a component of our eventual perception.

IV.3 From the Raw Primal Sketch

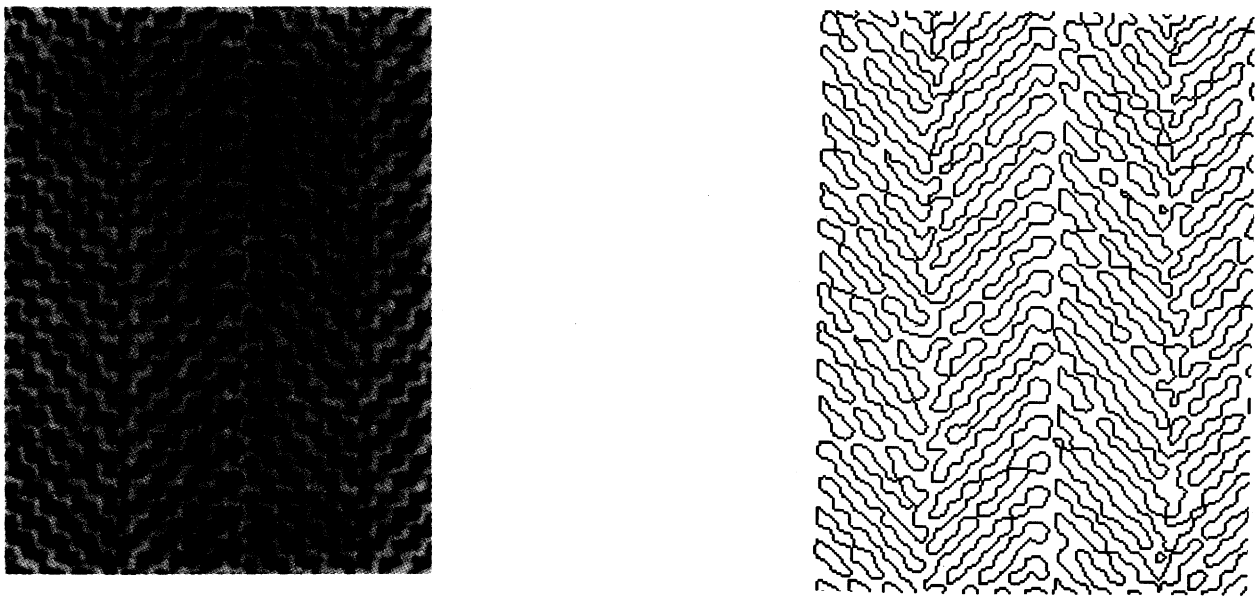
There are at least three classes of computations which will utilize either the raw primal sketch, or full primal sketch: elementary grouping operations, such as those described in Marr (1976b, 1980); light related computations, such as those dealing with reflectances, effects of illumination, and detection of light sources; and early motion analysis, in particular, motion correspondence (Ullman 1979). Further research in these areas could have important consequences for the integration of the channel descriptions.

There are two, seemingly different, types of grouping operations which may take place in early vision. The first is the summary of primitive primal sketch elements (segments, blobs, bars, and terminations) which are similar to one another along some dimension, such as orientation, contrast, or size. Two

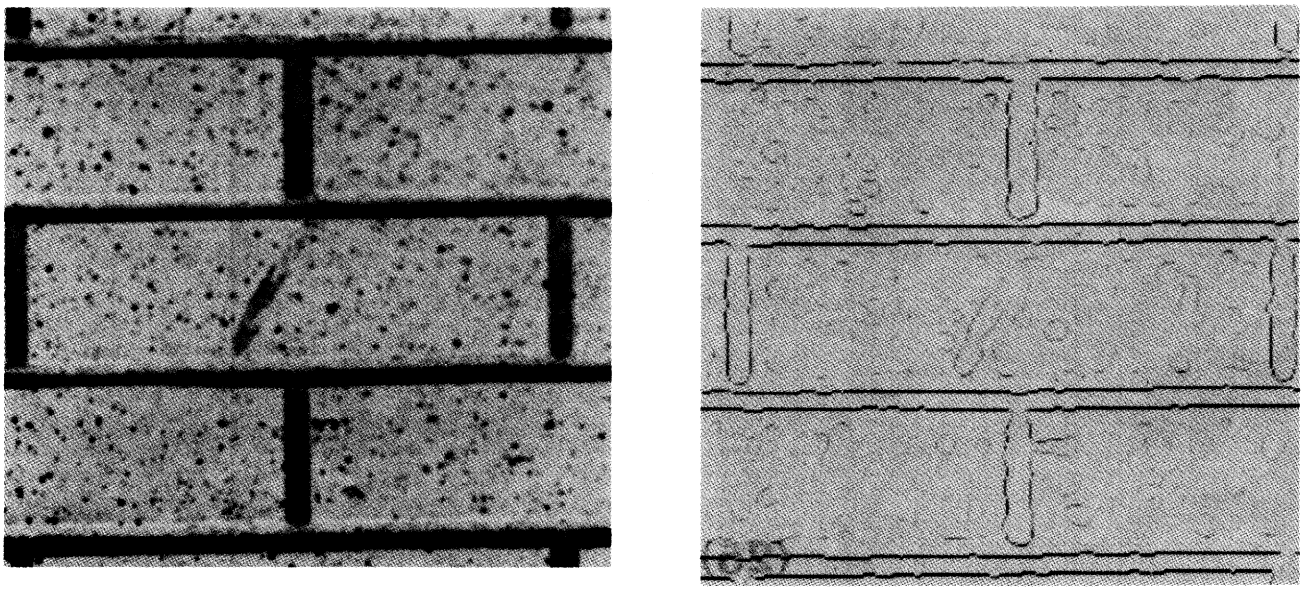
examples appear in Figure 25. In Figure 25a, we may group zero-crossing segments with similar orientations into columns; in Figure 25b, the similar property is contrast, which separates the textured faces of the bricks from their outline. The second type of grouping depends more on the spatial arrangement of the primitives, than properties of the primitives themselves, and is illustrated in Figure 25c. Here, the primitives are uniform: dots of the same size and contrast. However, there exists strong association between particular dots, which results in the formation of subjective contours. A strong motivation for the first type of grouping is that the summary of the local properties of orientation, contrast, and size over small areas, can be used for obtaining information about surface geometry, either local surface orientation or relative depth, or discontinuities in these properties (Stevens 1979, Marr 1980). The usefulness of the second type of grouping process appears more general; it may reflect a general ability to aggregate elements along contours (such as in the linear aggregation described in Marr (1976b)), where these contours may be the result of any of a number of physical phenomena, such as an incomplete structural boundary, illumination boundary, or boundary defining a discontinuity in surface orientation on a textured surface. The final description, containing both the primitive elements and their groupings has been called the full primal sketch.

Two remarks should be made concerning these early grouping operations. First, these operations may be quite rough. For example, concerning the use of the orientation of zero-crossing segments in recovering information about surface geometry, Marr (1980) has shown that the orientation of elements on a surface are not very sensitive to changes in the underlying surface orientation; that is, a significant change in surface orientation may only yield a small change in the orientation of an element on the surface. Riley (1977) provides evidence that the use of local orientation measures in texture discrimination are extremely coarse. In Figure 26a is the image of a piece of animal fur, in which it appears

a.



b.



25. Examples of elementary grouping. In (a), we have the image of a piece of tweed, and resultant zero-crossings. Using the property of orientation of the zero-crossing segments, we could group similar segments into columns. In (b), where we present the zero-crossings resulting from the image of a brick wall, the property of contrast allows us to summarize the fine texture on the face of the bricks, or the outline of the bricks. In (c) is an array of similar items: uniform dots of the same size and contast. The spatial arrangement of the dots yields subjective contours.

a.

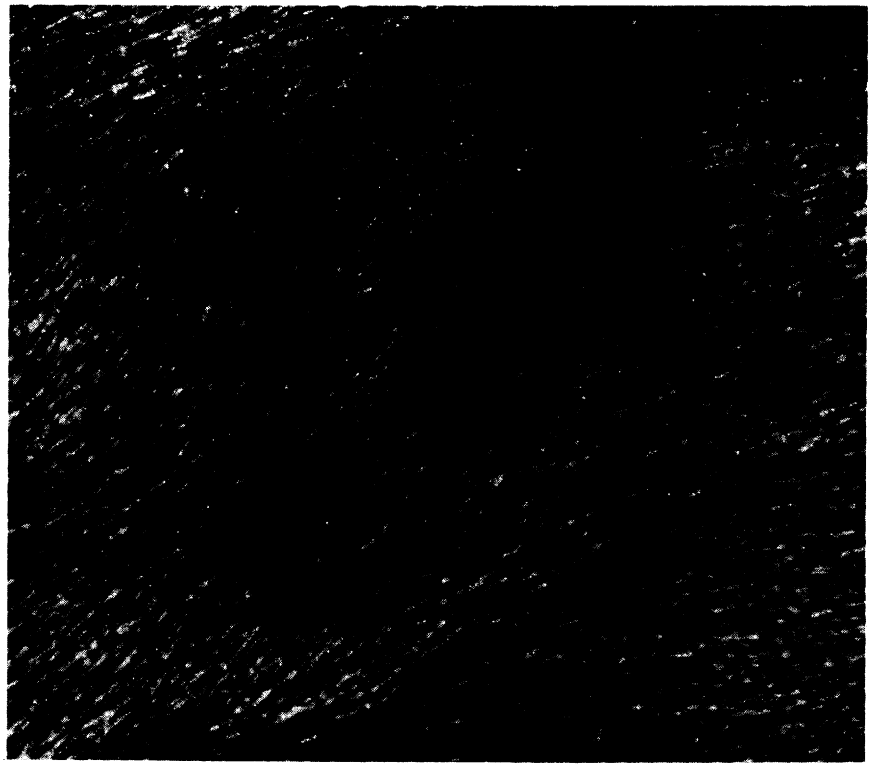
b.

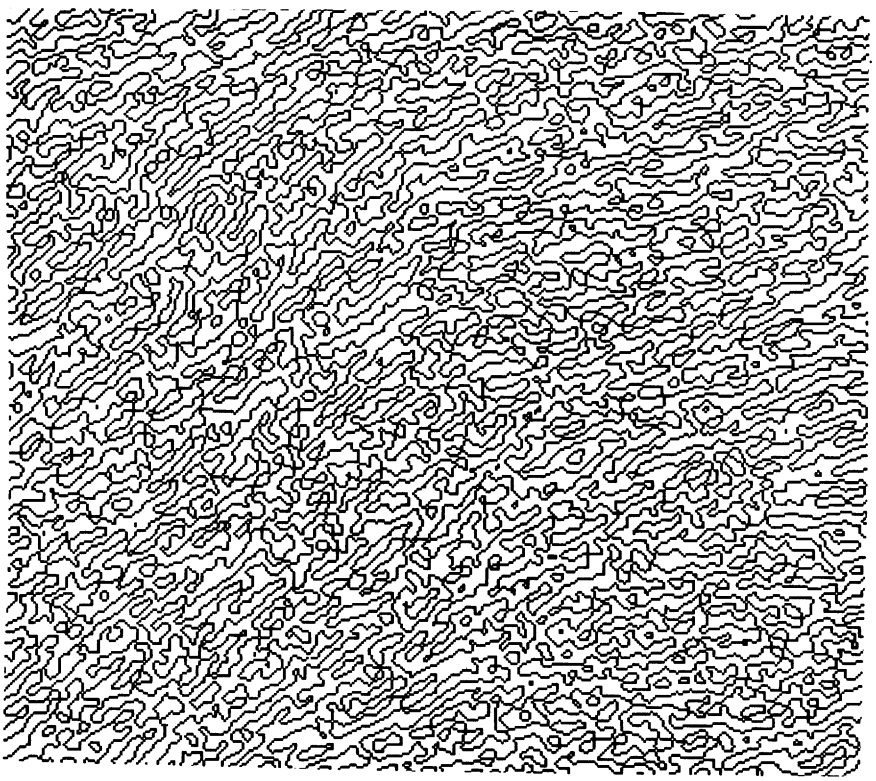
25. Examples of elementary grouping. In (a), we have the image of a piece of tweed, and resultant zero-crossings. Using the property of orientation of the zero-crossing segments, we could group similar segments into columns. In (b), where we present the zero-crossings resulting from the image of a brick wall, the property of contrast allows us to summarize the fine texture on the face of the bricks, or the outline of the bricks. In (c) is an array of similar items: uniform dots of the same size and contrast. The spatial arrangement of the dots yields subjective contours.

that the local orientation of the individual hairs provides information about the underlying surface. In Figure 26b are the zero-crossings in the output of the convolution of this image with a ∇^2G operator of size w = 6 pixels (the entire picture was 400 x 600 pixels). Average orientation of the zero-crossings changes slowly and smoothly as we move across the image, but locally, the zero-crossings are very rough. This will be true for any natural image of a fine texture such as this, so that any process utilizing this information for recovering surface geometry can not be too sensitive to local orientation measures. It should be noted here that the zero-crossings that we compute from the ∇^2G output do not necessarily give an accurate reflection of the analysis that takes place at the level of the simple cells in the human system. In the previous section, we saw in the example of the checkerboard, that although the zero-crossings theoretically remain on the grid as we increase the channel size, zero-crossing detectors, behaving in the manner of the simple cell model, would yield a stronger response if their optimal orientation were along the diagonals. When we draw parallels between the computations we are performing on images, and processing in the human visual system, some of the low-level differences (differences at the mechanism level described in Marr (1976a)) in the performance of these computations may be significant.

The second remark is that although grouping operations summarize a set of primitive elements, they may not necessarily replace those primitive elements in the final description. Ullman (1979) stresses that in the case of motion correspondence, we make use of group tokens *in addition* to the primitive elements they group together; that group tokens do not replace their constituents in the correspondence process. Riley (1980) has recently run some psychophysical experiments using groups of tokens in temporal sequence; in these experiments, a correspondence between the most primitive elements would result in a different perception of motion than correspondence between the group tokens; results

20. Zero-crossings in the output of a small $\nabla^2 G$ operator (w = 6 pixels) applied to an image of animal fur. The change in orientation of the individual hairs appears to provide information about the gradual changes in surface orientation. Obtaining this local orientation information from the image may be quite difficult, as the local orientation of the zero-crossing segments is very coarse.





indicate that the correspondence is formed between the group tokens.

In an attempt to understand better the nature of the final product of these grouping operations, Marr (1980), laid out 6 physical assumptions, on which the design of the representation for the full primal sketch may be based. The emphasis in these assumptions is on the relationship between the description of intensity changes, and the geometry of the surfaces from which they were generated, which we ultimately seek to characterize. I would like to repeat five of those assumptions here:

1. Existence of the Surface: the visible world can be regarded as being composed of smooth surfaces that have reflectance functions whose spatial structure may be elaborate.
2. Hierarchical Organization: the spatial organization of a surface's reflectance function is often generated by a number of different processes operating at different scales.
3. Similarity: the items, generated on a given surface by a reflectance-generating process acting at a given scale, tend to be more similar to one another in their size, local contrast, color, and spatial organization, than to other items on that surface.
4. Spatial Continuity of Spatial Markings: tokens often form smooth contours on a surface.
5. Continuity of Discontinuities: the loci of discontinuities in depth or in surface orientation are smooth almost everywhere.

An extensive discussion of these assumptions is given in Marr (1980); here, I would only like to provide a brief discussion indicating their relevance to the present work. Assumptions 4 and 5 form the key motivation for the grouping of elements along contours. In the imaging process, we may lose the continuity of

underlying physical phenomena, such as boundaries of discontinuity in surface orientation. By a loss of continuity, we mean that such a contour is not given explicitly by a continuous zero-crossing contour at some scale (see Riley (1980) for discussion on this point). Assumption 5 would justify the grouping of the points marking the discontinuity into some representation of a smooth, continuous contour.

The notion of scale used in assumptions 1 and 2 refers to different levels of spatial organization, and not necesarily to what we may see through different size differential operators. On one hand, we require a sensitivity to intensity changes occuring at a range of scales (different parts of the frequency spectrum), because the features we wish to group together may reflect intensity changes taking place in these different parts of the spectrum (for example, the summarization of the brick texture certainly requires sensitivity to high frequencies), but the organization itself is not explicitly given by the larger channels. In the brick wall example, the levels of organization we may want to make explicit are (1) the fine texture of the individual bricks, (2) the outline of the bricks, and (3) the organization of the bricks into rows and columns. Through the smallest operator, we can see the texture of the bricks, but we also see the outline of the bricks, and organization of the bricks into rows and columns. Through larger filters, we would see the outline of the bricks, and their organization into rows, but on the faces of the bricks, the contours would represent a smoothing of the fine texture. Assumption 3 implies that if we have a reflectance-generating process acting at a given scale, such as the texture on the bricks, then the items generated by this texture will tend to be more similar to one another (along some dimension) than to elements generated by processes acting at a different scale, such as the organization of the bricks into rows. In practice, this will be true both at the level of the individual scales, and in an integrated description.

These assumptions provide insight into what we want to compute for the full primal sketch, and the nature of the final representation. Two problems not addressed here are how this description is computed, and what is the form of the intermediate representation (raw primal sketch) from which these spatial organizations are computed. It is yet unclear what role (if any) the descriptions provided by the individual channels may have in the grouping process. In our brick wall example, the summary of the texture of the bricks (possibly based on similar contrast) could take place either from a single channel description, or a description obtained by combining the channels. In the case of the inner texture of the bricks, the corresponding output from the larger channels is not very meaningful for its summarization. A "chicken and egg" problem arises here: the ability to summarize what takes place at one scale may be useful for determining how to deal with information present at another scale, suggesting that grouping operations might be useful for integrating the channels. A different view is that very local operations are sufficient for this integration, which precedes the grouping operations.

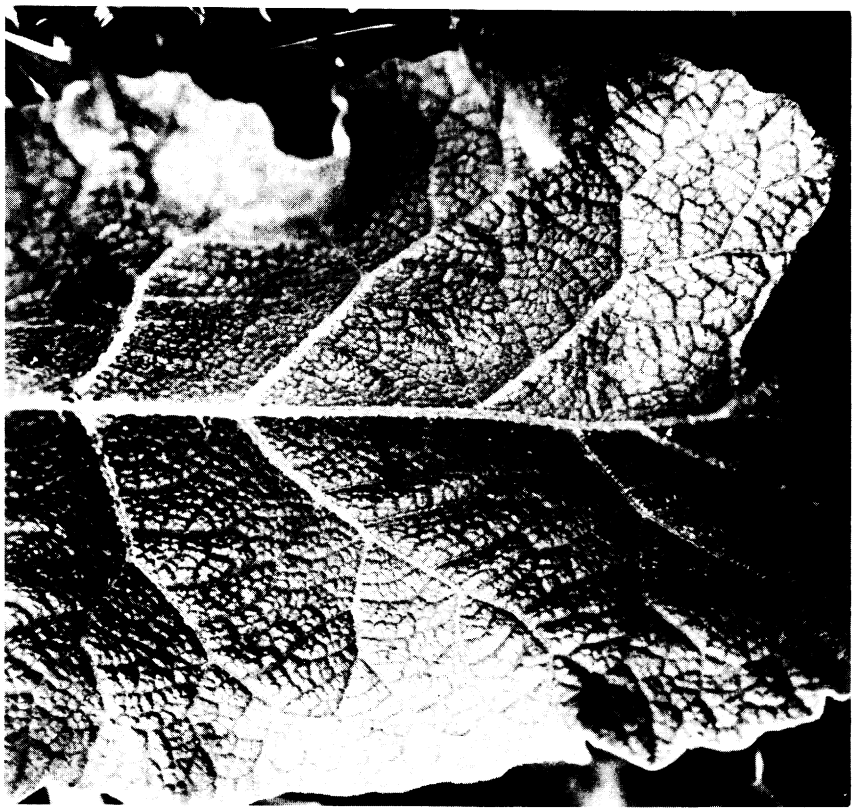
The significance of the above discussion for the present work is this: the theory presented by Marr and Hildreth (1979) proposes that the description of intensity changes at different scales is combined into a single representation of all intensity changes, called the raw primal sketch. It was found that the computation required for this integration is non-trivial. The difficulties encountered lead one to question again what we want to compute at this stage, and what should be the form of the representation. These questions are partially answered for the full primal sketch, but until we understand the grouping processes better, I feel that answers to these questions for the raw primal sketch can only be speculative.

Two other processes which may have bearing on the primal sketch are motion correspondence and lightness related computations. Motion correspondence

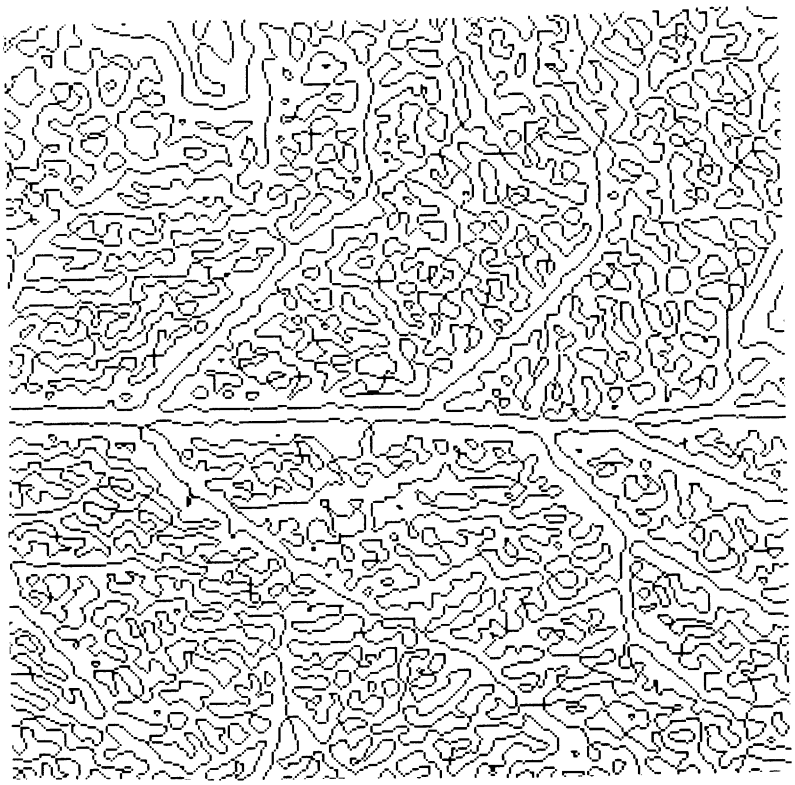
appears to utilize the grouping of elements (see Ullman 1979, p. 64), but work in this area has not addressed the role of scale in the correspondence computation; that is, whether correspondence is computed between single, hierarchical representations of intensity changes which also summarize the basic elements into groups reflecting some of the spatial structure of the image, or whether the computation might utilize descriptions which more closely resemble the channel outputs, with additional grouping operations performed on the separate channels. Again, I do not feel there is a clear answer to this question.

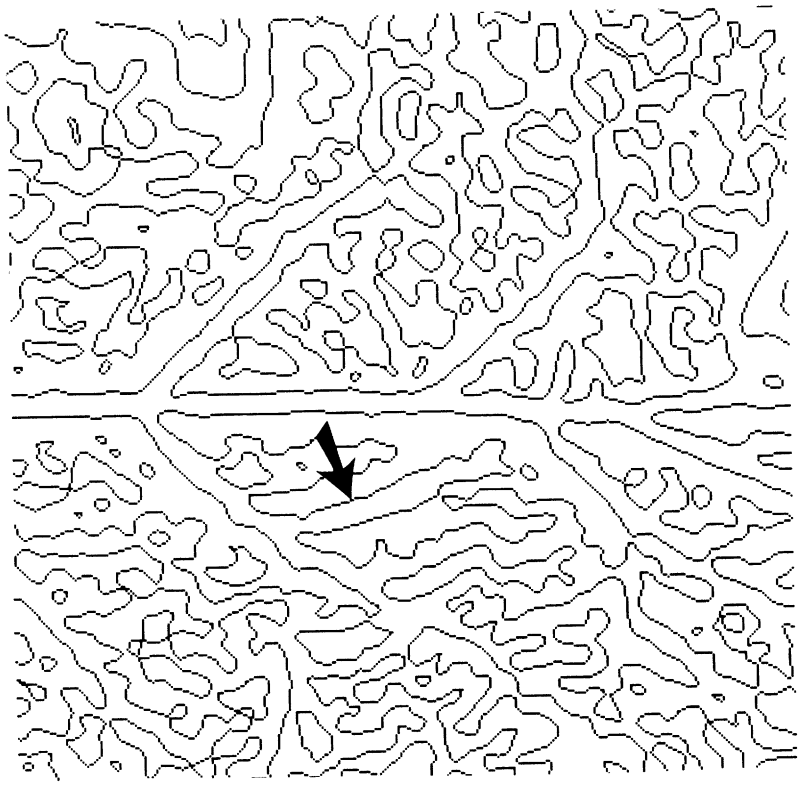
Another class of computations, whose time course we know little about, are the lightness related computations. In the case of these processes, the larger channels often make explicit information that could be valuable for them. Shading information, for example, is often conveyed by the low frequency channels. To illustrate this, the leaf image of Figure 3 is shown again here in Figure 27, with the zero-crossings from a larger channel. The zero-crossings from the smaller channel capture the bumpy texture, and through their slopes, indirectly carry information concerning overall changes in intensity due to the illumination on the leaf. In the larger channel, these changes in shading across the leaf are made explicit; they outline contours along which a change in surface orientation with respect to the direction of illumination, results in a significant change in intensity. Changes in intensity will result from changes in reflectance, illumination, depth, or surface orientation; the nature of a particular change might be conveyed somewhat by the relationship between information carried by the separate channels. If this were the case, it might be useful to retain a division of channel information for lightness related computations. These processes will therefor also be relevent for answering the questions of what is made explicit in the raw primal sketch, and how the information is represented.

In conclusion of this section, there have been significant advancements



27. Zero-crossings in the outputs of two $\nabla^2 G$ operators applied to the image of a leaf. The smaller operator primarily detects the bumpy texture on the surface of the leaf, whereas zero-crossing contours from the larger operator tend to outline the highlights on the leaf surface due to illumination.





over the past few years toward understanding some of the processes which [illegible] the primal sketch, at some stage in its development. This work, in turn, allowed significant advancement in the primal sketch theory. In particular, for the very low-level operations involved in stereopsis and directional selectivity, we can provide a set of channel descriptions which appear to contain all of the necessary information for their successful computation, in an efficient representation. There still, however, remains a gap between what we can provide now, and what is required by other computations, such as motion correspondence, and those pertaining to lightness. Further work in these other areas may be necessary to bridge this gap.

V. Other Approaches to Edge Detection

Other work in computer vision and image analysis has shaped the development of the primal sketch. Use of differential operators, in particular the Laplacian; a range of operator sizes, with selection criteria for deciding on the size operator whose response best reflects properties of an edge; and band-limited Gaussian filtering have received considerable attention, but few approaches have been able to integrate these ideas into a single coherent theory. As a result, the success of most edge detection schemes has been limited to fairly narrow application domains. In this section we present an overview of some of the other approaches which share ideas common to the theory proposed here.

Perhaps the greatest difficulty with differential operators proposed in the literature is that they utilize an extremely small support, generally covering an area of 10 to 20 picture elements (see, for example, reviews in Rosenfeld & Kak 1976, Davis 1975, and Pratt 1978). This contrasts drastically with the smallest operator used here, whose support covers roughly 900 pixels. It was shown in Section II that this large support is necessary, in order to reduce the operators' sensitivity to noise and quantization. A good demonstration of the sensitivity of these small differential operators to noise is given in Pratt (1978 p. 500). Larger differential operators are proposed in the literature (Davis 1975, Rosenfeld 1970, 1971, 1972, Kelly 1971), but as size increases, the constraints of localization in space and frequency, together with the property that it not be orientation dependent, become more critical.

The simplest such differential operators are the horizontal and vertical first-difference operators:

$$f_x(i,j) = f(i,j) - f(i-1,j)$$

$$f_y(i,j) = f(i,j) - f(i,j-1)$$

f(x,y) is the intensity function. Peaks are localized in the output of such an

operator, and a threshold function is then applied to remove edges due to noise and quantization (see,for example, Rosenfeld and Kak 1976, Pratt 1978). Choice of an appropriate threshold has been a critical issue for this type of operator.

Roberts (1965), recognizing that the sensitivity of a gradient operator to noise could be reduced by making the operator symmetric in x and y, proposed the following gradient operator:

$$G(i,j) = [\{f(i,j) - f(i+1,j+1)\}^2 + \{f(i+1,j) - f(i,j+1)\}^2]^{1/2}$$

which has a more commonly used approximation:

$$G(i,j) = \max (|f(i,j) - f(i+1,j+1)|, |f(i+1,j) - j(i,j+1)|)$$

Generalization of this type of operator to utilize larger neighborhood sizes has been suggested, and early in this effort, Rosenfeld (1971, 1972) proposed a straightforward extension of the first differences scheme to use a range of mask sizes, which will be described later in this section.

Herskovitz and Binford (1970) used an initial second derivative operation:

$$D(i,j) = -2f(i,j) + f(i+d,j) + f(i-d,j)$$

computed for a neighborhood about the point (i,j). Curiously, the subsequent operations performed on the output essentially isolated the zero-crossings of the initial second derivative operation. They first compute the function:

$$F_{step}(i,j) = \Sigma_k \text{sign}\{D(i+k,j)\} - \Sigma_k \text{sign}\{D(i-k,j)\}$$

where sign(i,j) = 1 if i > 0, -1 if i < 0, and 0 if i = 0. The output of this function yielded extrema both at the location of the original edge, and on the two sides of the edges, where the convolution output returns to zero. The next step in their algorithm removed the side 'satellites', leaving them with extrema whose position roughly corresponded to the position of the zero-crossings of the original second derivative output. A primary concern for their operator was that it not be sensitive to the particluar noise and distortion of their optical system, so parameters for each operator were carefully chosen to minimize these

effects.

The Sobel operator (Pratt 1978, p. 487) is a simple gradient operator, again covering a neighborhood of only 3 x 3 pixels. Let X and Y be the output of the following operators applied to the intensity array about a point:

vertical derivative:

$$\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$$

horizontal derivative:

$$\begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$$

The final operator is then: $G(i,j) = (X^2 + Y^2)^{1/2}$. Kirsch (1971) and Persoon (1976) propose extensions to the above scheme which measure the local derivatives in eight orientations, and choose the maximum derivative. Persoon utilizes a modified derivative operator, defined over a 5x5 patch, where neighborhoods on either side of a point are first fit to an ideal linear function, then the average intensities of the two regions, together with their deviation from the ideal model, are used in computing the derivative for each orientation. Again, these edge detection operators examine only limited neighborhood sizes.

The Laplacian has been used in a similar manner as the above operators (Rosenfeld 1976, Pratt 1978, Weschler and Fu 1978, Shanmugen, Dickey & Green 1979). In most applications, rather than isolating the zero-crossings in the output of the operator, the magnitude of the output is taken, followed by a thresholding operation. As a result, shallow edges appear as dark patches in the output, and are not very localized; sharp edges, which give rise to a steep slope in the convolution across zero, are lost. It is no wonder that researchers have been discouraged by exploring this operator further. Weschler and Fu (1978)

propose a modification to the Laplacian operation in which they scale the output, and add the gradient, but the resulting operator is still not satisfactory as a general edge detector:

$$G(i,j) = (D_x + D_y)^{-1/2} + 3*L(i,j)$$

where D_x and D_y are the derivatives in x and y, and L(i,j) is the Laplacian. Kelly (1971 p. 400) stated that the Laplacian, amoung other operators, was unsuccessful as an edge detector because of its tendency to amplify noise. Noise sensitivity is a function of the size and shape of the operator, and not a direct consequence of using the Laplacian per se.

An exception to these uses of the Laplacian can be found in the early work of Horn (1968) on edge detection. Horn utilized a local Laplacian operator, and recognized that the useful elements in the output of such an operator are the zero-crossings. Horn also emphasized the advantages of using Gaussians in the formation of an edge detection operator.

Most applications of the above edge detection schemes use small mask sizes, followed by subsequent thresholding and thinning operations, and resort to the addition of higher-level knowledge about the particular domain to locate desired edges (Roberts 1965, Kelly 1971, Shirai 1973, Persoon 1976, Weschler and Fu 1977, Abele and Wahl 1977). Kelly describes part of a vision system for recognizing human faces. His approach, termed "planning", is to first obtain a coarse 'plan' for the edges in the image by sampling the image at a very low resolution, and obtaining an edge description for the sampled image using a modified Roberts operator, and finally, use the position of the edges in the plan to locate edges in the original image. In this paper, Kelly was concerned primarily with obtaining the outline of the head, for which he could easily use particular top-down knowledge about the shape of the human head in rejecting many candidate edges, such as the fact that the outline will be a smoothly varying concave contour, roughly horizontal along the top, etc. Weschler & Fu, and

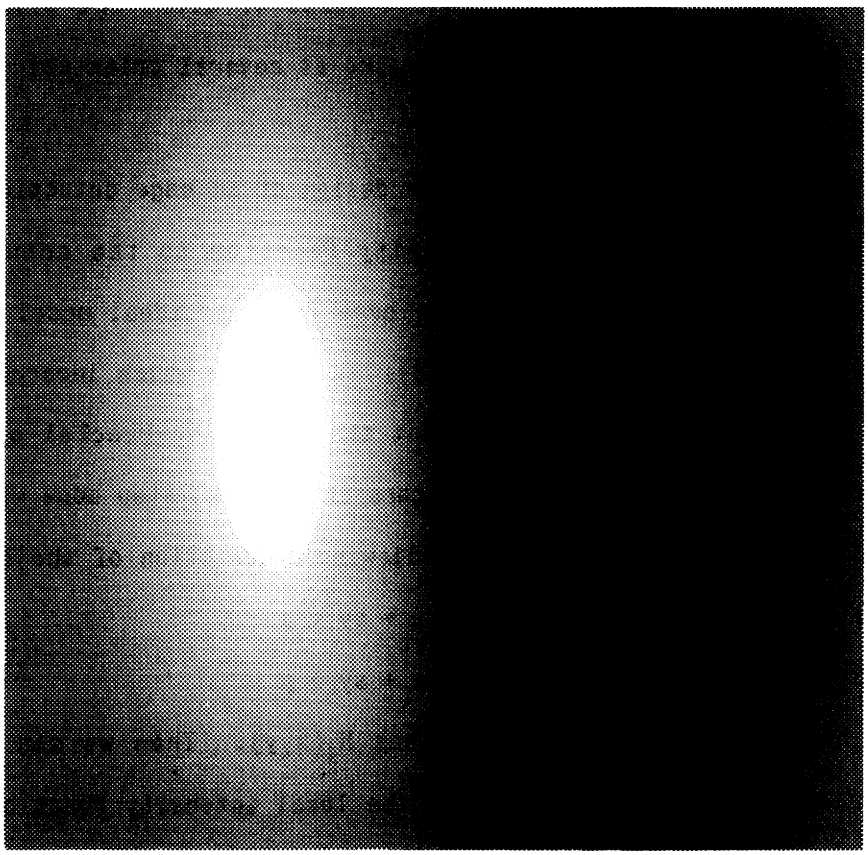
Persoon examine the problem of extracting the outline of ribs in chest X-rays. In this domain, they can take advantage of the expected orientation, contrast, and spatial arrangement of ribs to guide the low-level edge detection and contour building processes. Once two strong rib contours (one on each side) are located, the positions of the remaining contours are fairly well defined.

Binford & Horn (Horn 1973) and Shirai (1973) developed guided operators for locating the edges of polyhedra in the Blocks World. Shirai's scheme is to first find the most obvious edges in the scene, the contour lines which represent boundaries between objects and background. A set of heuristics then define the possible ways in which boundary lines (between two bodies) and internal lines (between two faces of the same object) can be extended from the present edge description. These heuristics restrict the locations at which the local operator is used to detect the presence of a possible edge, and allow the dynamic adjustment of the threshold used for detecting these edges. This reduces the chance of the operator in locating false edges. The Binford-Horn linefinder combines local edge detection and line-following techniques for locating edges. The initial edge-marking is performed by a non-linear, parallel operation across the image which detects local edge-points by correlating the local intensity function with the common step, roof, and peak profiles, and accumulates lists of edge-points which are connected. The lines are then segmented into lists of points likely to represent single edges, and finally, vertices and joint types are made explicit for subsequent scene analysis. This line-finder was designed to successfully locate edges in the particular world of convex polyhedra, with matte surfaces, and hence could take advantage of higher-level heuristics for integrating edge fragments into a full line-drawing, which are particular to these images.

Rosenfeld (1970, 1971, 1972) has proposed various schemes for utilizing a range of operator sizes. In his basic algorithm, one first computes, for each

point in the image, the average grey levels for a range of neighborhood sizes, typically from 2 to 32 pixels. For each size, a local difference is computed between adjacent neighborhoods, in four different orientations, and the orientation of maximum gradient chosen. At each point, a best size is then selected, which corresponds to the largest size for which the next larger neighborhood size does not yield a significantly larger response. (This is similar to the selection criterion described in Marr (1976b) for computing the width of an edge.) In his original proposal (1970), Rosenfeld suggested taking a product of the output of the different size difference operations at each point. It is simple to construct situations in which the product would not be the desirable measure; possibly this is the reason for proposing a second scheme. The final step in the algorithm is to extract maxima and minima from the output. Although the algorithm uses a range of sizes, it still maintains in the final output, edges which were detected by the smallest operator, so as a result, the final output still contains too many minor intensity changes.

It has also long been recognized that a desirable property for an edge operator is that it suppress very high spatial frequencies, containing noise and quantization effects, as well as very low frequencies, which do not capture the sharp edges which the operator is to detect. Some discussion of the design of an optimal filter appears in Duda & Hart (1973). Shanmugam, Dickey, and Green (1979), also approaching edge detection by optimizing frequency domain filters, propose an operator which yields maximum energy in the vicinity of an edge, within some specified resolution interval. The operator they propose is, in one dimension, the D^2G function, however, rather than localizing the zero-crossings in the output, they appear to take the magnitude, and threshold. A split Gaussian spatial filter, whose motivation was to design an operator which gave less weight to data values far from the location of an edge, was proposed by Argyle (1971). The operator is an oriented edge mask, with Gaussian weighting



28. The edge operator proposed by Macleod. The operator has Gaussian weighting functions both perpendicular to and along the orientation of the mask.

functions, but with a discontinuity along its central axis.

Perhaps the most successful edge detection operator has been proposed by Macleod (1972); the operator, illustrated in Figure 28, is an oriented edge mask with Gaussian weighting functions both perpendicular to and along the orientation of the mask. Macleod also observed that edges exist at different scales in the image, so that their detection requires masks of different size. Macleod's overall scheme combines the results of masks at several orientations, apparently with some success.

A very different approach to the design of an edge detection operator is to fit a particular idealized intensity function to the actual intensity function, and compute parameters of the edge from its ideal model (for example, Heuckel 1971, O'Gorman 1976, Brooks 1978). An example which incorporates some of the principles suggested in the present theory is the Hueckel operator. His basic approach is to compute the parameters of an ideal step edge which best fits the intensity function locally. The following description of the edge is given:

$$F(x,y,c,s,p,b,d) = b \qquad \text{if } cx + sy < p$$

$$F(x,y,c,s,p,b,d) = b + d \qquad \text{if } cx + sy > p$$

Assume we have a set of basis functions $H_i(x,y)$. Then we can describe the response of each of these functions to the local intensity function as follows:

$$a_i = \Sigma_d H_i(x,y) I(x,y)$$

The response of the ideal step edge would be:

$$f_i(c,s,p,d,b) = \Sigma_d H_i(x,y) F(x,y,c,s,p,d,b)$$

Minimizing the fit of the ideal edge to the intensity function is then equivalent to minimizing:

$$\Sigma_i [a_i - f_i(c,s,p,d,b)]^2$$

The particular basis operators were designed with the following criteria in mind:

(1) reduction of error in converting from the continuous to discrete formalization.

(2) decreased weight toward the periphery (to reduce the influence of nearby edges).

(3) attenuation of very high spatial frequencies; the H_i's can be described as spatial frequency filters, with frequency sensitivity increasing with i. By utilizing a finite set of these operators, Hueckel reduced the sensitivity of the overall operation to high frequency noise.

(4) minimization of computation time.

Hueckel solves this minimization problem; the particular basis operators to which it gave rise are described in Heuckel (1971). The Heuckel operator shares some of the basic principles on which our operators were based.

In summary, although particular segments of the present theory for edge detection have been given prior attention, the combination of these ideas into a single algorithm is essential for general edge detection. The use of zero-crossings in a second derivative Laplacian operator, a range of sizes, localization of the operator both in space and frequency, and constraint on the smallest size operator are all essential components of the theory. Most important, the present theory demonstrates that early processing of intensity information can successfully use simple, local operations; specific knowledge concerning the scene being viewed is not necessary.

VI. Summary

The intent of this work was to illustrate the importance of an implementation in refining the details of a computational theory, keeping in mind two goals which have motivated the primal sketch theory: the first was to contribute toward a more rigorous understanding of the general requirements of the edge detection stage in early visual processing; the second was to understand the underlying computations necessary to explain the early processing of information in the human visual system. In the case of the computation of the single channel descriptions, experimentation with different shape and size operators lead to the constraints of localization in space and frequency, no orientation dependence, and a lower bound on the smallest operator. Questions concerning the size of this smallest operator lead to a study of human acuity and hyperacuity, and the proposal of a 1.5' channel, and simple schemes for the interpolation of the output of this channel in order to localize zero-crossings in the hyperacuity range.

Concerning the integration of information from the separate channels, the computation of contrast and width, and role of the selection criterion was made more precise, for the case of simple, relatively isolated intensity changes. It was suggested that in more complex situations, this computation becomes non-trivial, and other sources of information which may contribute toward this recovery were suggested. The integration of the channels must be sensitive to the requirements of those further computations which will use the raw primal sketch. Examining the processes of grouping, motion correspondence, and lightness computations only briefly, opened up many questions concerning the representation of the raw primal sketch, and the grouping operations which lead to the full primal sketch. I feel that we do not yet have adequate answers for these questions.

In conclusion of this report, there are a few further points which I

would like to stress. First, the raw primal sketch is a modest goal for initial processing of an image. The implementation so far indicates that such a description is computable. Also, there is reason to believe that the single channel descriptions form a complete representation of the image. This conjecture first grew out of a result of Logan (1977), that a one-dimensional signal which is bandpass with width less than one octave, and has no free zeros, is completely reconstructible from its zero-crossings, up to a multiplicative constant. The relevance of this result to two-dimensional band-limited signals is discussed in Marr, Poggio, and Ullman (1978). Recently, progress has been made on the problem of reconstructing the two-dimensional output of the convolution of an image with $\nabla^2 G$ from its zero-crossings (Nishihara 1979). The implication of this result is that in the transformation of the original image to the single channel descriptions of the zero-crossing contours, no information is lost.

There are two additional implications of the primal sketch, which will be supported by the successful implementation of this and other stages in early visual processing. First, the primal sketch illustrates that the first, edge detection stage can be a low-level process, which does not require input from higher-level processes for the guidance of its operation. From the primal sketch, stereopsis, and early motion analysis alone, we can build a useful and sophisticated description of the visual scene. Second, it is the lowest level of description from which stereopsis and motion analysis may operate; these secondary processes no longer have access to the initial intensity values from which the primal sketch was derived.[11] These implications are critical, in light of other work in the area of early vision.

FOOTNOTES

1. The original Gaussian was defined as:

$$G(x,y) = \sigma^2 e^{-r^2/2\sigma^2}$$

where $r^2 = x^2 + y^2$. This gives:

$$\nabla^2 G = \partial^2 G/\partial x^2 + \partial^2 G/\partial y^2 = [2 - r^2/\sigma^2]e^{-r^2/2\sigma^2}$$

The constant factor is arbitrary, and not at all critical for the performance of the operator.

2. The more critical measure here is the positioning of the zero-crossings. The average deviation in position of the zero-crossings between the ideal and noisy convolution outputs was 0.15 pixels in this case.

3. In this case, the average deviation in position of the zero-crossings marking the bar edges (between the ideal and noisy outputs) was again roughly 0.15 pixels, but the total number of zero-crossings has tripled.

4. The psychophysically measured response function (Wilson & Bergen 1979) is:

$$G(x,y) = 1/\sigma_e{}^2 e^{-r^2/2\sigma_e{}^2} - 1/\sigma_i{}^2 e^{-r^2/2\sigma_i{}^2}$$

The ratio of space constants $\sigma_e:\sigma_i$ was calculated to be roughly 1:1.75 for the smaller, sustained channels, and 1:3.0 for the larger, transient channels.

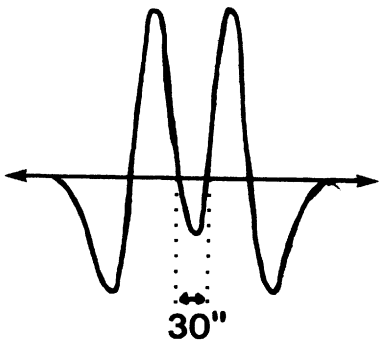
5. Hubel and Wiesel (1974a) observed that during tangential penetration of striate cortex in monkey, changes in the orientation to which the simple cells were most sensitive, seem to occur in small, relatively constant steps of about 10°. Schiller et. al. (1976b) studied quantitatively, the degree of orientation specificity for several classes of S-type cells (cells behaving roughly as Hubel and Wiesel's simple cells), and found a wide range of specificity, although the cells with one or two subfields appeared to be more finely tuned.

6. Acuity experiments generally use a forced-choice paradigm; one could therefor argue that there are a number of potential sources of information on which the discrimination is based. Using the basic configurations in Westheimer's bar experiment (Westheimer 1977), the thresholds which would be predicted by the use of other information were computed. In this experiment, the subject was shown a set of configurations consisting of two bars, each 3 units long and 1 unit wide, separated by a space which was also 3 units long, and 1 unit wide:

The size of the unit varied around 1', in 6" intervals, and the configuration was displayed either horizontally or vertically. The subjects' task was to decide whether the bars were horizontal or vertical. The threshold was defined as the

separation which yielded 75% correct responses. In what follows, I describe three potential sources of information for the discrimination, the criterion used in each case to define the threshold, and the threshold predicted, as a function of w, the size of the channel. Thresholds were computed for a one-dimensional profile, using the D^2G operator.

1. zero-crossings- The threshold is defined to be the separation between bars which would yield distinct zero-crossings in the convolution output, between the positions of the bars, and separated by at least 30":

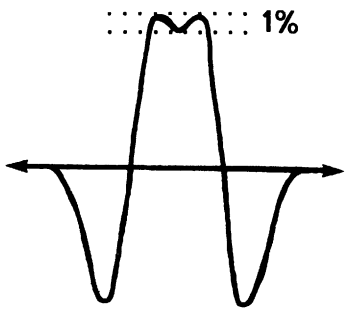


For small values of w, the threshold separation is an approximately linear function of w:

$$thresh = 0.875w - 0.375$$

A smallest channel of 3' would predict a threshold of 2.25', whereas the 1.5' channel would predict a separation of 0.94', which is roughly equivalent to the experimentally observed threshold.

2. peak detection- This threshold is defined to be the separaton between bars which would yield two distinct peaks in the convolution output, with a 1% dip in the magnitude of the output between the peaks. This was generally coupled with a separation between peaks of at least 1':



Again, for small values of w, the threshold separation is roughly linear:

$$thresh = 0.75w - 0.75$$

The 3' channel would predict a threshold of 1.5', the 1' channel would predict 0.125'.

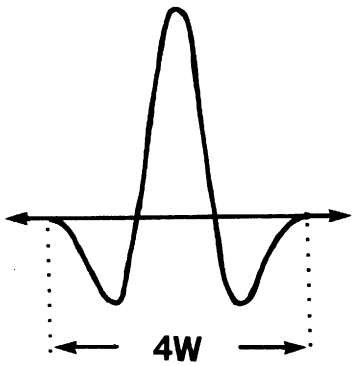
3. outer zero-crossings- In the particular experimental setup here, there would be a small shift in the outer zero-crossings between the horizontal and vertical presentations, which would be in the hyperacuity range, even for the larger, 3' channel. One could argue, therefor, that this information is used in making the discriminations, and that the discrimination could then be made with a smallest channel as large as 3'. However, it is possible to set up a single bar profile, and a profile with two bars, which would yield D^2G outputs (with a 3' channel) that differ by

very little (rms error is roughly 0.04). There would be no information from distinct peaks or inner zero-crossings in this case, so if the outer zero-crossings were the only information used to make the discriminations in Westheimer's experiment, then the two configurations should not be discriminable. Under rough experimental conditions, the two patterns are clearly discriminable; it would be interesting to know whether this remains to be the case under more careful experimental conditions.

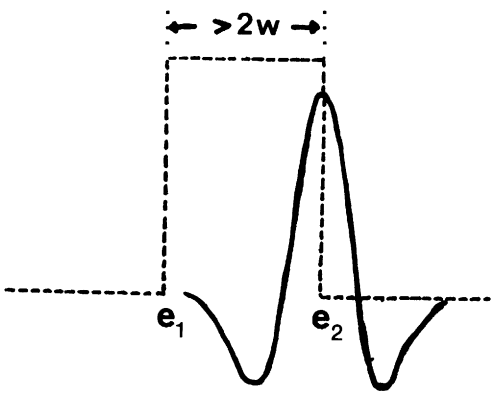
7. There are a number of parameters on which these statistics depend: the particular size and shape characteristics of the initial filter; the high frequency content of the signal (such as from noise); the sampling interval before reconstruction; and the resolution at which the signal is reconstructed. For example, using an initial filter whose central positive width corresponds to 3' of arc (vs. 1.5'), all reconstruction functions yield roughly zero error for sampling intervals of 20" and 40"; for a 60" interval, average shift is under 0.2 pixels. It was generally the case that much of the error in positioning the zero-crossings occurred within areas of low contrast, high frequency change.

8. Two assumptions were made in this computation. First, it assumes the extent of the mask to be large enough that its total area is roughly zero. Second, I assume the slope to be measured perpendicular to the orientation of the edge.

9. In one dimension, the total extent of the mask is roughly 4w:



This extent yields approximately zero area. In this case, an edge e_2 which is separated from a given edge e_1 by more than 2w will not have any influence on the zero-crossing at e_1:



An edge within a distance 2w can influence the zero-crossing's position and slope. The change in slope will vary linearly with the difference in contrast between the two edges, and fluctuates around zero as the difference in position of the two edges moves from 2w to zero.

10. We have the relations:

$$s_2/s_1 = ce^{-d^2/8(2\sigma_1)^2}/ce^{-d^2/8\sigma_1^2} = e^{d^2/16\sigma_1^2}$$

$$s_1 = ce^{-d^2/8\sigma_1^2}$$

Suppose, for example, we have two step changes in intensity, separated by some distance x, and we obtain zero-crossings from D^2G filters with $w_2 = 2x$ and $w_1 = x$. Here, $d = 0$, so ideally $s_2/s_1 = 1$ for each edge. However, due to a separation between the edges of $x = w_1 < w_2$, the larger channel will not measure the slope accurately. The error will be a function of the difference in contrast between the two edges. The shift in the zero-crossings is very small in this case. If we find a scheme for determining that the larger channel zero-crossing is not reliable (possibly based on the shift of the zero-crossing), and use the smaller channel slope, we need to estimate one of the edge parameters, most likely width, and then compute contrast using our estimated width. Unfortunately, the above formula is very sensitive to the estimation of width. If we used both slopes anyway, assuming they reflect an ideal situation, and use the above equations, our contrast could be in error by as much as a factor of two. This example was intended to emphasize the fact that local zero-crossing information may not be sufficient for a good estimation of the edge properties.

11. It is possible that in lightness related computations, other operations take place in parallel with the detection of edges, such as application of the $\Delta I/I$ operator proposed by Ullman (1976) for the detection of light sources.

REFERENCES

Abele, L. & Wahl, F. (1977) A Digital Procedure for Boundary Detection and Elimination of Background in Cytologic Images. in *Medinfo '77 (Proceedings of the Second World Conference on Medical Informatics)* (Shires, D. & Wolf, H., Eds.) Toronto.

Argyle, E. (1971) Techniques for Edge Detection. *Proceedings of the IEEE 59,2*, 285-287.

Barlow, H. B. (1979) Reconstructing the Visual Image in Space and Time. *Nature 279*, 189-190.

Beck, J. & Schwartz, T. (1978) Vernier Acuity With Test Dot Objects. *Vision Res. 19*, 313-319.

Blakemore, C. & Campbell, F. W. (1969) On the Existence of Neurones in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images. *J. Physiol. 203*, 237-260.

Brooks, M. J. (1978) Rationalizing Edge Detectors. *Computer Graphics and Image Processing 8*, 277-285.

Burr, D. C. (1979) Acuity for Apparent Vernier Offset. *Vision Res. 19*, 835-837.

Campbell, F. W. & Green, D. G. (1965) Optical and Retinal Factors Affecting Visual Resolution. *J. Physiol. 181*, 576-593.

Campbell, F. W. & Gubisch, R. W. (1966) Optical Quality of the Human Eye. *J. Physiol. 186*, 558-578.

Campbell, F. W. & Robson, J. G. (1968) Applications of Fourier Analysis to the Visibility of Gratings. *J. Physiol. (Lond.) 197*, 551-556.

Cleland, B. G., Dubin, M. W. & Levick, W. R. (1971a) Sustained and Transient Neurones in the Cat's Retina and Lateral Geniculate Nucleus. *J. Physiol. (Lond.) 217*, 473-496.

Cleland, B. G., Dubin, M. W. & Levick, W. R. (1971b) Simultaneous Recording of Input and Output of Lateral Geniculate Neurones. *Nature, New Biology 231*, 191-192.

Cowan, J. D. (1977) Some Remarks on Channel Bandwidths for Visual Contrast Detection. *Neurosciences Res. Prog. Bull. 15*, 492-517.

Crick, F. H. C., Marr, D. & Poggio, T. (1980) An Information Processing Approach to Understanding the Visual Cortex. To appear in the N.R.P. Symposium *The Cerebral Cortex* (Schmidt, F. O. & Worden, F.G., Eds.).

Davis, L. (1975) A Survey of Edge Detection Techniques. *Computer Graphics and Image Processing 4*, 248-270.

deMonasterio, F. M. (1978a) Properties of Concentrically Organized X and Y Ganglion Cells of Macaque Retina. *J. Neurophysiol. 41*, 1394-1417.

deMonasterio, F. M. (1978b) Center and Surround Mechanisms of Opponent-Color X and Y Ganglion Cells of Retina of Macaques. *J. Neurophysiol. 41*, 1418-1434.

deMonasterio, F. M. & Gouras, P. (1975) Functional Properties of Ganglion Cells of the Rhesus Monkey Retina. *J. Physiol. 251*, 167-195.

Dowling, J. E. & Werblin, F. S. (1969) Organization of the Retina of the Mudpuppy, Necturus Maculosus: I. Synaptic Structure. *J. Neurophysiol. 32*, 315-338.

Duda, R. O. & Hart, P. E. (1973) *Pattern Recognition and Scene Analysis.* John Wiley, N.Y.

Enroth-Cugell, C. & Robson, J. G. (1966) The Contrast Sensitivity of Retinal Ganglion Cells of the Cat. *J. Physiol. (Lond.) 187*, 517-552.

Frisby, J. P. & Julesz, B. (1976) The Effect of Length Differences Between Corresponding Lines on Stereopsis From Single and Multi-line Stimuli. *Vision Res. 16*, 83-87.

Griffith, A. K. (1973) Edge Detection in Simple Scenes Using A Priori Information. *IEEE Transactions on Computers, vol C-22*, 371-181.

Grimson, W. E. L. (1980) *Aspects of a Computational Theory of Stereo Vision.* (MIT Ph. D. Thesis in preparation).

Grimson, W. E. L. & Marr, D. (1979) A Computer Implementation of a Theory of Human Stereo Vision. In: *Proceedings of ARPA Image Understanding Workshop*, SRI, pp. 41-45.

Hayes, K. C. & Rosenfeld, A. (1972) Efficient Edge Detectors and Applications. *U. Maryland TR-207* November.

Herskovitz, A. & Binford, T. O. (1970) On Boundary Detection. *MIT AI Lab Memo 183.*

Horn, B. K. P. (1968) *The Application of Fourier Transform Methods to Image Processing*, MIT M.S. Thesis.

Horn, B. K. P. (1973) The Binford-Horn Line-Finder. *MIT AI Lab Memo 285.*

Hubel, D. H. & Wiesel, T. N. (1961) Integrative Action in the Cat's Lateral Geniculate Body. *J. Physiol. (Lond.) 155*, 385-398.

Hubel, D. H. & Wiesel, T. N. (1962) Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *J. Physiol. (Lond.) 160*, 106-154.

Hubel, D. H. & Wiesel, T. N. (1968) Receptive Fields and Functional Architecture of Monkey Striate Cortex. *J. Physiol. (Lond.) 195*, 215-243.

Hubel, D. H. & Wiesel, T. N. (1974a) Sequence Regularity and Geometry of Orientation Columns in Monkey Striate Cortex. *J. Comp. Neurol. 158*, 267-293.

Hubel, D. H. & Wiesel, T. N. (1974b) Uniformity of Monkey Striate Cortex: a Parallel Relationship Between Field Size, Scatter, and Magnification Factor. *J.*

Comp. Neurol. 158, 295-306.

Hueckel, M. H. (1971) An Operator Which Locates Edges in Digital Pictures, *J. ACM 18*, 113-125.

Julesz, B. (1971) *Foundations of Cyclopian Perception*. The University of Chicago Press. Chicago.

Kelly, M. D. (1971) Edge Detection in Pictures By Computer Using Planning. in *Machine Intelligence VI*, Edinburgh U. Press, 397-409.

Kirsh, R. (1971) Computer Determination of the Constituent Structure of Biological Images. *Computers and Biomedical Research, 4, 3*, 315-328.

Kuffler, S. W. (1953) Discharge Patterns and Functional Organization of the Mammalian Retina. *J. Neurophysiol. 16*, 37-68.

Leipnik, R. (1960) The Extended Entropy Uncertainty Principle. *Inform. and Control 3*, 18-25.

Logan, B. F. (1977) Information in the Zero-crossings of Bandpass Signals. *Bell System Technical Journal 56*, 487-510.

Maffei, L. & Fiorentini, A. (1972) Retinogeniculate Convergence and Analysis of Contrast. *J. Neurophysiol. 35*, 65-72.

Macleod, I. D. G. (1972) Comments on 'Techniques for Edge Detection'. *Proceedings of the IEEE 60,3*, 344.

Marr, D. (1976a) Artificial Intelligence-- A Personal View. *MIT AI Lab Memo 355.*

Marr, D (1976b) Early Processing of Visual Information. *Phil. Trans. R. Soc. B. 275*, 483-524.

Marr, D. (1978) Representing Visual Information. AAAS 143rd Annual Meeting, Symposium on *Some Mathematical Questions in Biology*, February 1977. Published in *Lectures in the Life Sciences 10*, 101-180. Also available as MIT AI Lab Memo 415.

Marr, D. (1980) *VISION*. W. H. Freeman Co. San Francisco.

Marr, D. & Hildreth, E. C. (1979) Theory of Edge Detection. *MIT AI Lab Memo 518.*

Marr, D. & Nishihara, H. K. (1977) Representation and Recognition of the Spatial Organization of Three Dimensional Shapes. *MIT AI Lab Memo 416.*

Marr, D & Poggio, T. (1979) A Computational Theory of Human Stereo Vision. *Proc. R. Soc. Lond B. 204*, 301-328.

Marr, D., Poggio, T., & Hildreth, E. (1979) The Smallest Channel in Early Human Vision. (in press).

Marr, D., Poggio, T. & Ullman, S. (1979) Bandpass Channels, Zero-crossings, and Early Visual Information Processing. *J. Opt. Soc. Am.*, (in press).

Marr, D. & Ullman, S. (1979) Directional Selectivity and its Use in Early Visual Processing. *MIT AI Lab Memo 524.*

Nakayama, K. & Roberts, D. J. (1972) Line-length Detectors in the Human Visual System: Evidence From Selective Adaption. *Vision Res. 12*, 1709-1713.

Nishihara, H. K. (1979) Reconstruction of Bandwidth Limited Surfaces From Gradients at Zero-crossings. (in preparation)

Persoon, E. (1976) A New Edge Detection Algorithm and Its Applications. *Computer Graphics and Image Processing*, *5*, 425-446.

Poggio, G. F. & Fischer, B. (1978) Binocular Interaction and Depth Sensitivity of Striate and Prestriate Neurons of the Behaving Rhesus Monkey. *J. Neurophysiol. 40*, 1392-1405.

Polyak, S. L. (1941) *The Retina*. University of Chicago Press, Chicago.

Pratt, W. (1978) *Digital Image Processing*. John Wiley & Sons, N.Y.

Riley, M. (1977) *Descrimination of Bar Textures With Differing Orientation and Length Distributions*. MIT B.S. Thesis.

Riley, M. (1980) *Representing the Image* (MIT M.S. Thesis in preparation).

Roberts, L. G. (1965) Machine Perception of Three Dimensional Solids. in *Optical and Electro-Optical Information Processing* (J. Tippett, D. Berkowitz, L. Clapp, C. Koester, A. Vanderbergh, Eds.), MIT Press, 159-197.

Rodieck, R. W. & Stone, J. (1965) Analysis of Receptive Fields of Cat Retinal Ganglion Cells. *J. Neurophysiol. 28*, 833-849.

Rosenfeld, A. & Kak, A. (1976) *Digital Picture Processing*, Academic Press, New York.

Rosenfeld, A. & Thurston, M. (1971) Edge and Curve Detection for Visual Scene Analysis. *IEEE Transactions on Computers, vol C-20*, 562-569.

Rosenfeld, A. & Thurston, M. & Lee, Y. (1971) Edge and Curve Detection: Further Experiments. *IEEE Transactions on Computers, vol C-21*, 677-715.

Sachs, M. B., Nachmias, J. & Robson, J. G. (1971) Spatial-frequency Channels in Human Vision. *J. Opt. Soc. Am. 61*, 1176-1186.

Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976a) Quantitative Studies of Single-cell Properties in Monkey Striate Cortex. I. Spatiotemporal Organization of Receptive Fields. *J. Neurophysiol. 39*, 1288-1319.

Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976b) Quantitative Studies of Single-cell Properties in Monkey Striate Cortex. II. Orientation Specificity and Ocular Dominance. *J. Neurophysiol. 39*, 1320-1333.

Shanmugen, K. S., Dickey, F. M. & Green, J. A. (1979) On Optimal Frequency Domain Filter for Edge Detection in Digital Pictures. *IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-1*, 37-49.

Shirai, Y. A. (1975) A Context Sensitive Line-finder for Recognition of Polyhedra. *Artificial Intelligence 4*, 95-113.

Snyder, A. W. & Miller, M. H. (1977) Photoreceptor Diameter and Spacing for Highest Resolving Power. *J. Opt. Soc. Am. 76*. 696-698.

Stevens, K. (1979) Surface Perception From Local Analysis of Texture and Contour. *MIT PhD Thesis*.

Tolhurst, D. J. (1973) Separate Channels for the Analysis of the Shape and the Movement of a Moving Visual Stimulus. *J. Physiol. Lond. 231*, 385-402.

Ullman, S. (1976) On Visual Detection of Light Sources. *Biol. Cybernetics 21*,205-212.

Ullman, S. (1979) *The Interpretation of Visual Motion*. MIT Press, Cambridge, Ma.

Werblin, F. S. & Dowling, J. E. (1969) Organization of the Retina of the Mudpuppy, Necturus Maculosus: II. Intra-cellular Recording. *J. Neurophysiol. 32*, 339-355.

Weschler, H & Fu, K. S. (1977) Automatic Analysis of Chest Radiographs. in *Medinfo '77 (Proceedings of the Second World Conference on Medical Informatics)* (Shires, D. & Wolf, H., Eds.) Toronto.

Weschler, H & Fu, K. S. (1978) Image Processing Algorithms Applied to Rib Boundary Detection in Chest Radiographs. *Computer Graphics and Image Processing, 7*, 375-390.

Westheimer, G. (1976) Diffraction Theory and Visual Hyperacuity. *Am. J. Optometry and Physiol. Optics 53*, 362-364.

Westheimer, G. (1977) Spatial Frequency and Light-spread Descriptions of Visual Acuity and Hyperacuity. *J. Opt. Soc. Am. 67*, 207-212.

Westheimer, G. & McKee, S. (1975) Visual Acuity in the Presence of Retinal Image Motion. *J. Opt. Soc. Am. 65*, 847-850.

Westheimer, G. & McKee, S. (1976) Spatial Configurations for Visual Hyperacuity. *Vision Res. 17*, 941-947.

Westheimer, G. & McKee, S. (1977) Integration Regions for Visual Hyperacuity. *Vision Res. 17*, 89-93.

Wilson, H. R. & Bergen, J. R. (1979) A Four Mechanism Model for Threshold Spatial Vision. *Vision Res. 19*, 19-32.

Wilson, H. R. & Giese, S. C. (1977) Threshold Visibility of Frequency Grating Patterns. *Vision Res. 17*, 1177-1190.

CS-TR Scanning Project
Document Control Form

Date : 2/29/96

Report # AI-TR-579

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☒ Technical Report (TR) ☐ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 122(129-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☐ Single-sided or
☒ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☒ Offset Press ☐ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:
☒ DOD Form ☒ Funding Agent Form ☒ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages(by page number):_____

Photographs/Tonal Material (by page number): 11,14,16,20,24,32-33,35,38,45,51-52,85,88,91,95,100,109,[redacted].

Other (note description/page number):

Description :	Page Number:
IMAGE MAP: (1-122) UNNUMBERED TITLE PAGE, 2-122	
(123-129) SCANCONTROL, COVER, FUNDINGAGENT, DOD, TRGT'S (3).	

Scanning Agent Signoff:
Date Received: 2/29/96 Date Scanned: 3/6/96 Date Returned: 3/7/96

Scanning Agent Signature: Michael W. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AI-TR-579

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Implementation of a Theory of Edge Detection

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Ellen C. Hildreth

8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0643

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: April 1980

13. NUMBER OF PAGES: 124

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):

Edge Detection
Computer Vision
Image Analysis
Image Understanding
Zero-Crossings
Primal Sketch
Human Vision

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

This report describes the implementation of a theory of edge detection, proposed by Marr and Hildreth (1979). According to this theory, the image is first processed independently through a set of different size filters, whose shape is the Laplacian of a Gaussian, $G(x,y)$. Zero-crossings in the output of these filters mark the positions of intensity changes at different resolutions. Information about these zero-crossings is then used for deriving a full symbolic description of changes in intensity in the image, called the raw primal sketch.

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)

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 3/6/1996

M.I.T. Libraries
Document Services