PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Theory of Edge Detection

D. Marr and E. Hildreth

Proc. R. Soc. Lond. B 1980 **207**, 187-217
doi: 10.1098/rspb.1980.0020

Proc. R. Soc. Lond. B **207,** 187–217 (1980)
Printed in Great Britain

Theory of edge detection

By D. Marr and E. Hildreth

M.I.T. Psychology Department and Artificial Intelligence Laboratory, 79 Amherst Street, Cambridge, Massachusetts 02139, *U.S.A.*

(*Communicated by S. Brenner, F.R.S. – Received* 22 *February* 1979)

A theory of edge detection is presented. The analysis proceeds in two parts. (1) Intensity changes, which occur in a natural image over a wide range of scales, are detected separately at different scales. An appropriate filter for this purpose at a given scale is found to be the second derivative of a Gaussian, and it is shown that, provided some simple conditions are satisfied, these primary filters need not be orientation-dependent. Thus, intensity changes at a given scale are best detected by finding the zero values of $\nabla^2 G(x, y) * I(x, y)$ for image I, where $G(x, y)$ is a two-dimensional Gaussian distribution and ∇^2 is the Laplacian. The intensity changes thus discovered in each of the channels are then represented by oriented primitives called zero-crossing segments, and evidence is given that this representation is complete. (2) Intensity changes in images arise from surface discontinuities or from reflectance or illumination boundaries, and these all have the property that they are spatially localized. Because of this, the zero-crossing segments from the different channels are not independent, and rules are deduced for combining them into a description of the image. This description is called the raw primal sketch. The theory explains several basic psychophysical findings, and the operation of forming oriented zero-crossing segments from the output of centre–surround $\nabla^2 G$ filters acting on the image forms the basis for a physiological model of simple cells (see Marr & Ullman 1979).

Introduction

The experiments of Hubel & Wiesel (1962) and of Campbell & Robson (1968) introduced two rather distinct notions of the function of early information processing in higher visual systems. Hubel & Wiesel's description of simple cells as linear with bar- or edge-shaped receptive fields led to a view of the cortex as containing a population of feature detectors (Barlow 1969, p. 881) tuned to edges and bars of various widths and orientations. Campbell & Robson's experiments, showing that visual information is processed in parallel by a number of independent orientation and spatial-frequency-tuned channels, suggested a rather different view, which, in its extreme form, would describe the visual cortex as a kind of spatial Fourier analyser (Pollen *et al.* 1971; Maffei & Fiorentini 1977).

Protagonists of each of these views are able to make substantial criticisms of the other. The main points against a Fourier interpretation are: (1) The bandwidth of the channels is not narrow (1.6 octaves, Wilson & Bergen 1979). The corresponding receptive fields have a definite spatial localization. (2) As Campbell & Robson found, early visual information processing is not linear (e.g. probability summation (Graham 1977; Wilson & Giese 1977), and failure of superposition (Maffei & Fiorentini 1972*a*)). (3) Only rudimentary phase information is apparently encoded (Atkinson & Campbell 1974).

The main point against the linear feature-detector idea is that if a simple cell truly signals either the positive or the negative part of the linear convolution of its bar-shaped receptive field with the image intensity, it can hardly be thought of as making some symbolic assertion about the presence of a bar in the image (Marr 1976*a*, p. 648). Such a cell would necessarily respond to many stimuli other than a bar, more vigorously, for example, to a bright edge than to a dim bar, and thus would not be specific enough in its response to warrant being called a feature detector.

Perhaps the greatest difficulty faced by both camps is that neither approach can give direct information about the goals of the early analysis of an image. This motivated a new approach to vision, which enquired directly about the information processing problems inherent in the task of vision itself (Marr 1976*a*, *b*; and see Marr 1978 for the overall scheme). According to this scheme, the purpose of early visual processing is to construct a primitive but rich description of the image that is to be used to determine the reflectance and illumination of the visible surfaces, and their orientation and distance relative to the viewer. The first primitive description of the image was called the primal sketch (*Marr* 1976*b*) and it is formed in two parts. First, a description is constructed of the intensity changes in an image, using a primitive language of edge-segments, bars, blobs and terminations. This description was called the raw primal sketch (Marr 1976*b*, p. 497). Secondly, geometrical relations are made explicit (using virtual lines), and larger, more abstract tokens are constructed by selecting, grouping and summarizing the raw primitives in various ways. The resulting hierarchy of descriptions covers a range of scales, and is called the full primal sketch of an image.

Although the primal sketch was inspired by findings about mammalian visual systems, we were until recently unable to make it the basis of a detailed theory of human early vision. Three developments have made this possible now: (*a*) the emergence of quantitative information about the channels present in early human vision (Cowan 1977; Graham 1977; Wilson & Giese 1977; Wilson & Bergen 1979); (*b*) Marr & Poggio's (1979) theory of human vision (especially the framework within which it was written); and (*c*) the related observations of Marr *et al.* (1979) about the relevance of a result like Logan's (1977) theorem to early vision.

These advances have made possible the formulation of a satisfactory computational theory. This article deals with the first part, the derivation of the raw primal sketch. The theory itself is given in two sections, the first dealing with the

analysis within each channel, and the second, with combining information from different channels. Each computational section discusses algorithms for implementing the theory, and gives examples.

The second half of the article examines the implications for biology. The behaviour of the algorithms is shown to account for a range of basic psychophysical findings, and a specific neural implementation is presented. Our model is not intended as a complete proposal for a physiological mechanism, because it ignores the attribute of directional selectivity that so pervades cortical simple cells. The model does, however, make explicit certain nonlinear features that we regard as critical, and it forms the starting point for the more complete proposal of Marr & Ullman (1979), which incorporates directional selectivity.

Detecting and representing intensity changes in an image

A major difficulty with natural images is that changes can and do occur over a wide range of scales (Marr 1976*a*, *b*). No single filter can be optimal simultaneously at all scales, so it follows that one should seek a way of dealing separately with the changes occurring at different scales. This requirement, together with the findings of Campbell & Robson (1968), leads to the basic idea, illustrated in figure 1, in which one first takes local averages of the image at various resolutions and then detects the changes in intensity that occur at each one. To realize this idea, we need to determine (*a*) the nature of the optimal smoothing filter, and (*b*) how to detect intensity changes at a given scale.

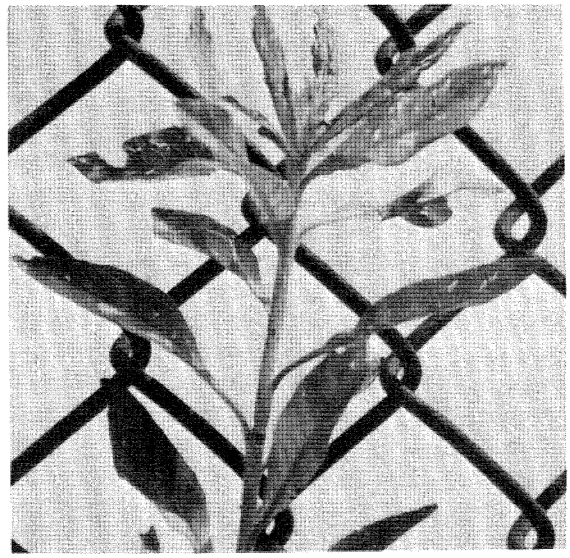
The optimal smoothing filter

There are two physical considerations that combine to determine the appropriate smoothing filter. The first is that the motivation for filtering the image is to reduce the range of scales over which intensity changes take place. The filter's spectrum should therefore be smooth and roughly band-limited in the frequency domain. We may express this condition by requiring that its variance there, $\Delta\omega$, should be small.

The second consideration is best expressed as a constraint in the spatial domain, and we call it the constraint of spatial localization. The things in the world that give rise to intensity changes in the image are: (1) illumination changes, which include shadows, visible light sources and illumination gradients; (2) changes in the orientation or distance from the viewer of the visible surfaces; and (3) changes in surface reflectance. The critical observation here is that, at their own scale, these things can all be thought of as spatially localized. Apart from the occasional diffraction pattern, the visual world is not constructed of ripply, wave-like primitives that extend and add together over an area (c.f. Marr 1970, p. 169), but of contours, creases, scratches, marks, shadows and shading.

The consequence for us of this constraint is that the contributions to each

(*a*)



(*b*)

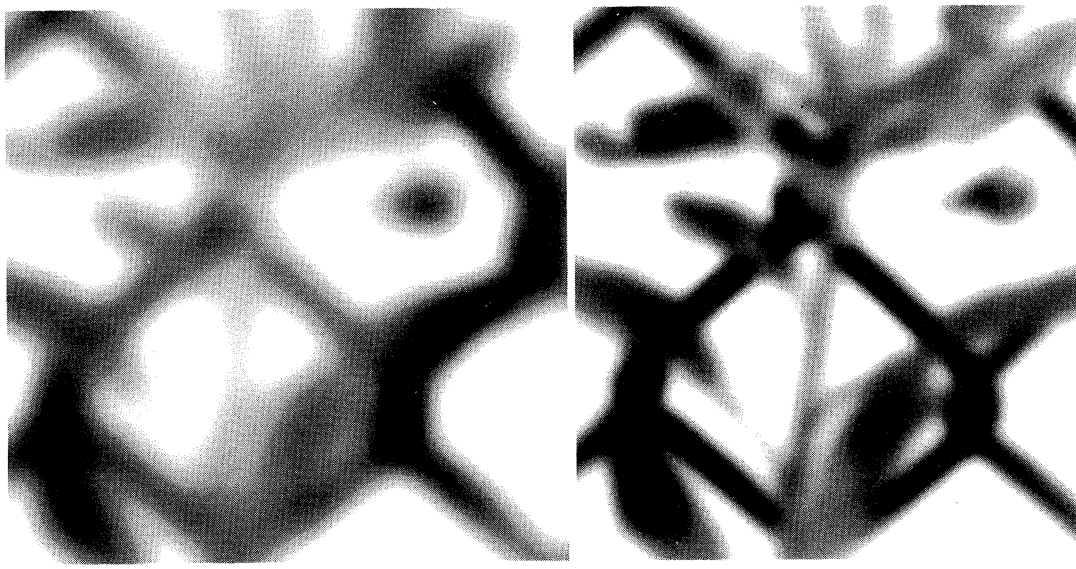


FIGURE 1. A local-average filtered image. In the original image (*a*), intensity changes can take place over a wide range of scales and no single operator will be very efficient at detecting all of them. The problem is much simplified in a Gaussian band-limited filtered image because there is effectively an upper limit to the rate at which changes can take place. The first part of our scheme can be thought of as decomposing the original image into a set of copies, each filtered like this, and detecting the intensity changes separately in each. In (*b*) the image is filtered with a Gaussian having $\sigma = 8$ picture elements, and, in (*c*), $\sigma = 4$. The image is 320×320 picture elements.

point in the filtered image should arise from a smooth average of nearby points, rather than any kind of average of widely scattered points. Hence the filter that we seek should also be smooth and localized in the spatial domain, and in particular its spatial variance, Δx, should also be small.

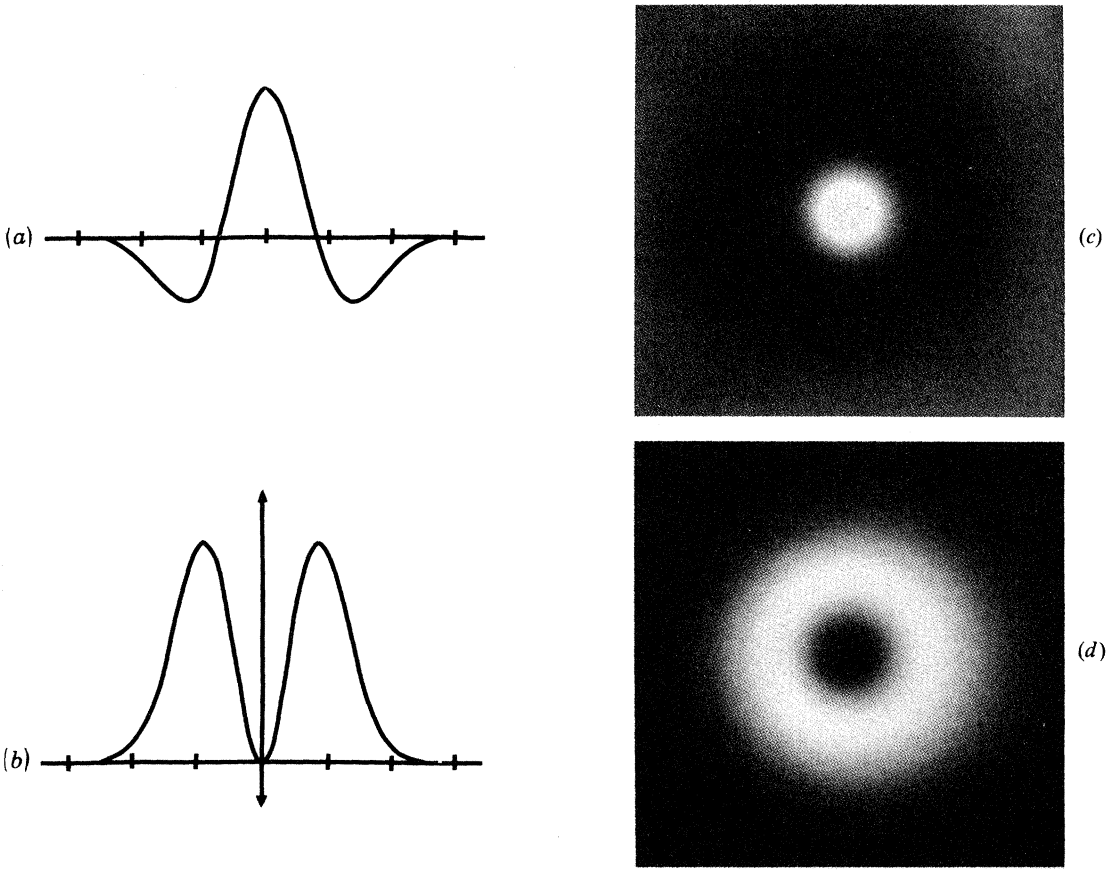


FIGURE 2. The operators G'' (equation 5) and $\nabla^2 G$: (*a*) shows G'', the second derivative of the one-dimensional Gaussian distribution; (*c*) shows $\nabla^2 G$, its rotationally symmetric two-dimensional counterpart; (*b*) and (*d*) exhibit their Fourier transforms.

Unfortunately, these two localization requirements, the one in the spatial and the other in the frequency domain, are conflicting. They are, in fact, related by the uncertainty principle, which states that $\Delta x\, \Delta\omega \geqslant \frac{1}{4}\pi$ (see, for example, Bracewell 1965, pp. 160–163). There is, moreover, only one distribution that optimizes this relation (Leipnik 1960), namely the Gaussian

$$G(x) = [1/\sigma(2\pi)^{\frac{1}{2}}] \exp(-x^2/2\sigma^2), \text{ with Fourier transform} \quad (1)$$

$$\tilde{G}(\omega) = \exp(-\tfrac{1}{2}\sigma^2\omega^2). \quad (2)$$

In two dimensions, $G(r) = (\frac{1}{2}\pi\sigma^2)\exp(-r^2/2\sigma^2)$.

The filter G thus provides the optimal trade-off between our conflicting requirements.

Detecting intensity changes

Wherever an intensity change occurs, there will be a corresponding peak in the first directional derivative, or equivalently, a zero-crossing in the second directional derivative of intensity (Marr 1976*b*; Marr & Poggio 1979). In fact, we may define an intensity change in this way, so that the task of detecting these changes

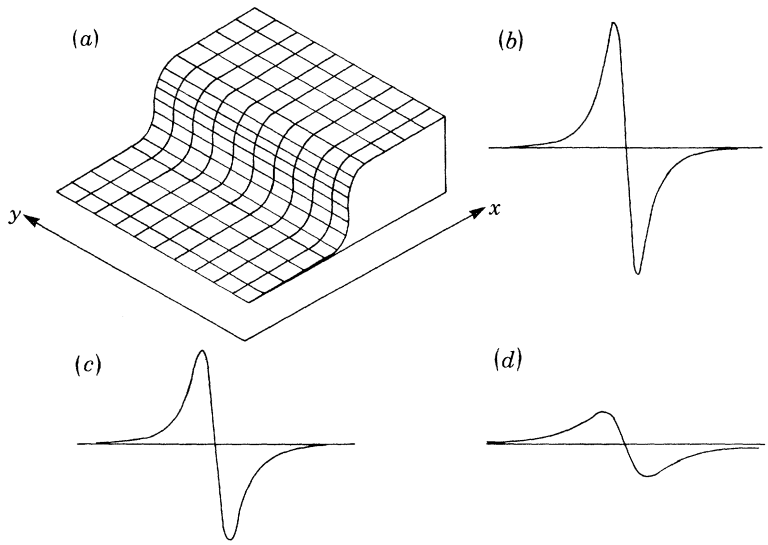


FIGURE 3. Spatial and directional factors interact in the definition of a zero-crossing segment; (*a*) shows an intensity change, and (*b*), (*c*) and (*d*) show values of the second directional derivative near the origin at various orientations across the change. In (*b*), the derivative is taken parallel to the x-axis, and in (*c*) and (*d*), at 30° and 60° to it. There is a zero-crossing at every orientation except for $\partial^2 I/\partial y^2$, which is identically zero. Since the zero-crossings line up along the y-axis, this is the direction that is chosen. In this example, it is also the direction that maximizes the slope of the second derivative.

can be reduced to that of finding the zero-crossings of the second derivative D^2 of intensity, in the appropriate direction. That is to say, we seek the zero-crossings in

$$f(x, y) = D^2[G(r)*I(x, y)], \tag{3}$$

where $I(x, y)$ is the image, and $*$ is the convolution operator. By the derivative rule for convolutions,

$$f(x, y) = D^2G * I(x, y). \tag{4}$$

We can write the operator D^2G as G'', and in one dimension

$$G''(x) = [-1/\sigma^3(2\pi)^{\frac{1}{2}}]\,(1-x^2/\sigma^2)\exp(-x^2/2\sigma^2). \tag{5}$$

$G''(x)$ looks like a Mexican hat operator (see figure 2), it closely resembles Wilson & Giese's (1977) difference of two Gaussians (DOG), and it is, in fact, the limit of the DOG function as the sizes of the two Gaussians tend to one another (see figure 11 and appendix B). It is an approximately bandpass operator, with a half-power bandwidth of about 1.2 octaves, and so it can be thought of as looking at the information contained in one particular part of the spectrum of the image.

These arguments establish that intensity changes at one scale may, in principle, be detected by convolving the image with the operator D^2G and looking for zero-crossings in its output. Only one issue is still unresolved, and it concerns the orientation associated with D^2. It is not enough to choose zero-crossings of the second derivative in *any* direction. To understand this, imagine a uniform intensity change running down the y-axis, as shown in figure 3. At the origin, the second directional derivative is zero in every direction, but it is non-zero nearby in every direction except along the y-axis.

In which direction should the derivative be taken?

To choose which directional derivative to use, we observe that the underlying motivation for detecting changes in intensity is that they will correspond to useful properties of the physical world, like changes in reflectance, illumination, surface orientation, or distance from the viewer. Such properties are spatially continuous and can almost everywhere be associated with a direction that projects to an orientation in the image. The orientation of the directional derivative that we choose to use is therefore that which coincides with the orientation formed locally by its zero-crossings. In figure 3, this orientation is the y-axis, and so the directional derivative we would choose there is $\partial^2 I/\partial x^2$.

Under what conditions does this direction coincide with that in which the zero-crossing has maximum slope? The answer to this is given by theorem 1 (see appendix A), and we call it the *condition of linear variation*:

> the intensity variation near and parallel to the line of zero-crossings should locally be linear.

This condition will be approximately true in smoothed images, and in the rest of this article we shall assume that the condition of linear variation holds.

This direction can be found by means of the Laplacian

There are three main steps in the detection of zero-crossings. They are: (1) a convolution with D^2G, where D^2 stands for a second directional derivative operator; (2) the localization of zero-crossings; and (3) checking of the alignment and orientation of a local segment of zero-crossings. Although it is possible to implement this scheme directly (Marr 1976*b*, p. 494), one immediate question that can be asked is, are directional derivatives of critical importance here? Convolutions are relatively expensive, and it would much lessen the computational burden if

their number could be reduced, for example, by using just one orientation-independent operator.

The only orientation-independent second-order differential operator is the Laplacian ∇^2, and theorem 2 (see appendix A) makes explicit the conditions under which it can be used. They are weaker than the condition of linear variation, which we met in theorem 1, and they state that provided the intensity variation in $(G * I)$ is linear along but not necessarily near to a line of zero-crossings, then the zero-crossings will be detected and accurately located by the zero values of the Laplacian. Again, because in our application the condition of linear variation is approximately satisfied, so will be this condition. It follows that the detection of intensity changes can be based on the filter $\nabla^2 G$, illustrated in figure 2. It is, however, worth remembering that in principle, if intensity varies along a segment in a very non-linear way, the Laplacian, and hence the operator $\nabla^2 G$ will see the zero-crossing displaced to one side.

Summary of the argument

The main steps in the argument so far are, therefore, these.

(1) To limit the rate at which intensities can change, we first convolve the image I with a two-dimensional Gaussian operator G.

(2) Intensity changes in $G * I$ are then characterized by the zero-crossings in the second directional derivative $\mathrm{D}^2(G * I)$. This operator is roughly bandpass, and so it examines only a portion of the spectrum of the image.

(3) The orientation of the directional derivative should be chosen to coincide with the local orientation of the underlying line of zero-crossings.

(4) Provided that the condition of linear variation holds, this orientation is also the one at which the zero-crossing has maximum slope (measured perpendicular to the orientation of the zero-crossing).

(5) By theorem 1 of appendix A, if the condition of linear variation holds, the lines of zero-crossings defined by (3) are precisely the zero-crossings of the orientation-independent differential operator, the Laplacian ∇^2.

(6) The loci of zero-crossings defined by (3) may therefore be detected economically in the image at each given scale by searching for the zero values of the convolution $\nabla^2 G * I$. In two dimensions,

$$\nabla^2 G(r) = -1/\pi\sigma^4[1 - r^2/2\sigma^2] \exp(-r^2/2\sigma^2).$$

We turn now to the question of how to represent the intensity changes thus detected.

Representing the intensity changes

In a band-limited image, changes take place smoothly, so it is always possible to divide a line of zero-crossings into small segments, each of which approximately obeys the condition of linear variation. This fact allows us to make the following definitions.

(1) *A zero-crossing segment* in a Gaussian filtered image consists of a linear segment l of zero-crossings in the second directional derivative operator whose direction lies perpendicular to l.

(2) We can also define an *amplitude* ν associated with a zero-crossing segment, as the slope of the directional derivative taken perpendicular to the segment. To see why this is an appropriate measure, observe that a narrow bandpass channel near a zero-crossing at the origin can be described approximately by $\nu \sin \omega x$, which has slope $\nu\omega$ at the origin. Hence, if s is the measured slope of the zero-crossing, $\nu = s/\omega$. The factor $1/\omega$ is a space constant, and scales linearly with the sampling interval required.

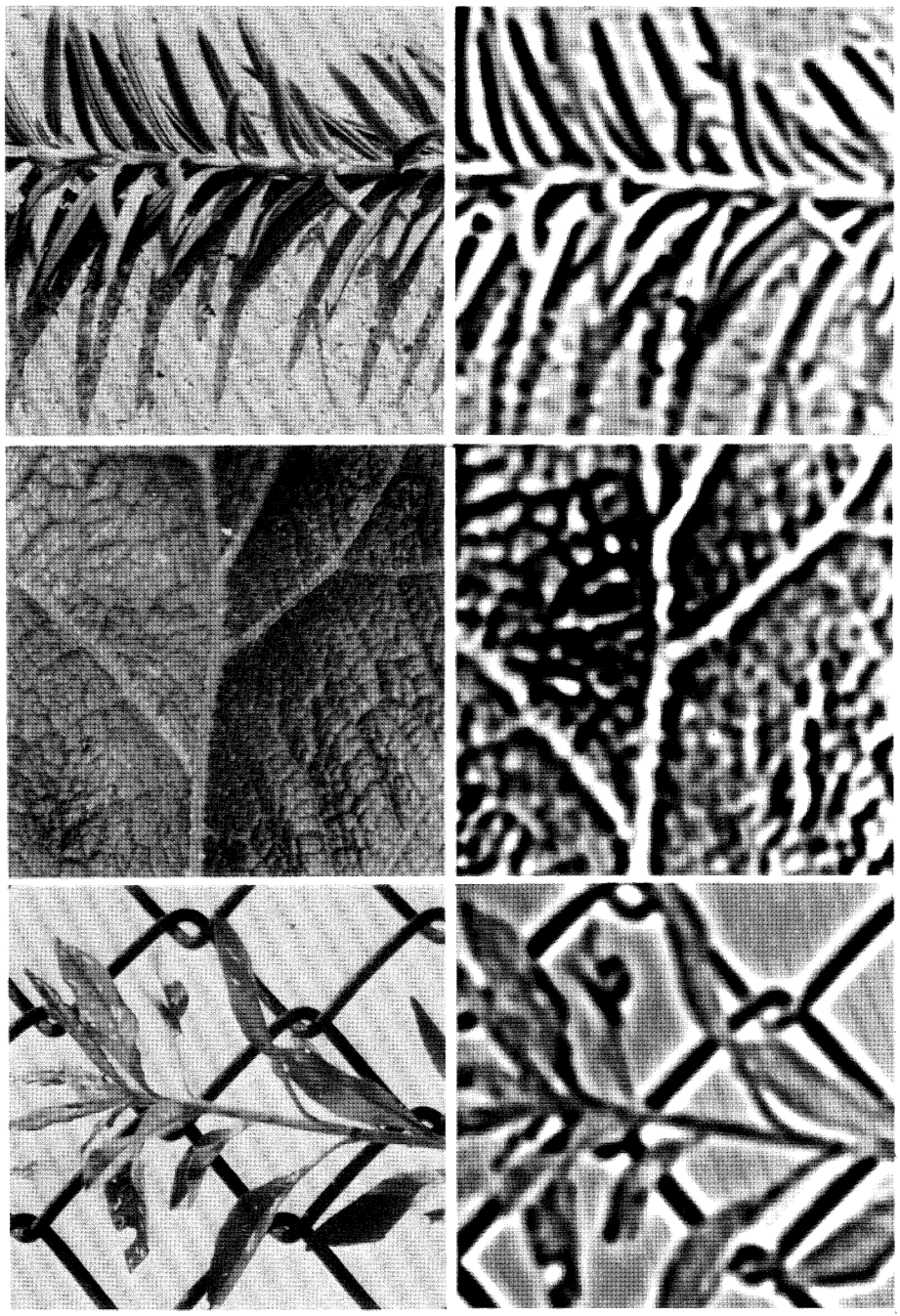
The set of zero-crossing segments together with their amplitudes, constitutes a primitive symbolic representation of the changes taking place within one region of the spectrum of an image. Full coverage of the spectrum can now be had simply by applying the analysis over a sufficient number of channels simultaneously.

Finally, there are grounds for believing that this representation of the image is complete. Marr *et al.* (1979) noted that Logan's (1977) recent theorem, about the zero-crossings of one-octave bandpass signals, shows that the set of such zero-crossing segments is extremely rich in information. If the filters had bandwidth of an octave or less, they would in fact contain complete information about the filtered image. In practice, the $\nabla^2 G$ filter has a half-sensitivity bandwidth of about 1.75 octaves, which puts it outside the range in which Logan's theorem applies. On the other hand, if we add information about the slopes of the zero-crossings, the situation may be more congenial. In the standard sampling theorem, if the first derivative, as well as the value, is given, the sampling density can be halved (see, for example, Bracewell 1978, pp. 198–200). It seems likely than an analogous extension holds for Logan's (1977) theorem. If this were true, the zero-crossing segments, whose underlying motivation is physical, would in fact provide a sufficient basis for the recovery of arbitrary intensity profiles.

In summary, then, we have shown how intensity changes at one scale may be detected by means of the $\nabla^2 G$ operator and that they may be represented, probably completely, by oriented zero-crossing segments and their amplitudes. To detect changes at all scales, it is necessary only to add other channels, like the one described above, and to carry out the same computation in each. These representations are precursors of the descriptive primitives in the raw primal sketch, and mark the transition from the 'analytic' to the 'symbolic' analysis of an image. The remaining step is to combine the zero-crossings from the different channels into primitive 'edge' elements, and this task is addressed later in the article.

Examples and comments

Figure 4 shows some examples of zero-crossings. The top row shows images and the second shows their convolutions with the operator $\nabla^2 G$, exhibited in figure 2. Zero is represented here by an intermediate grey, so that very positive values



(a) (b)

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

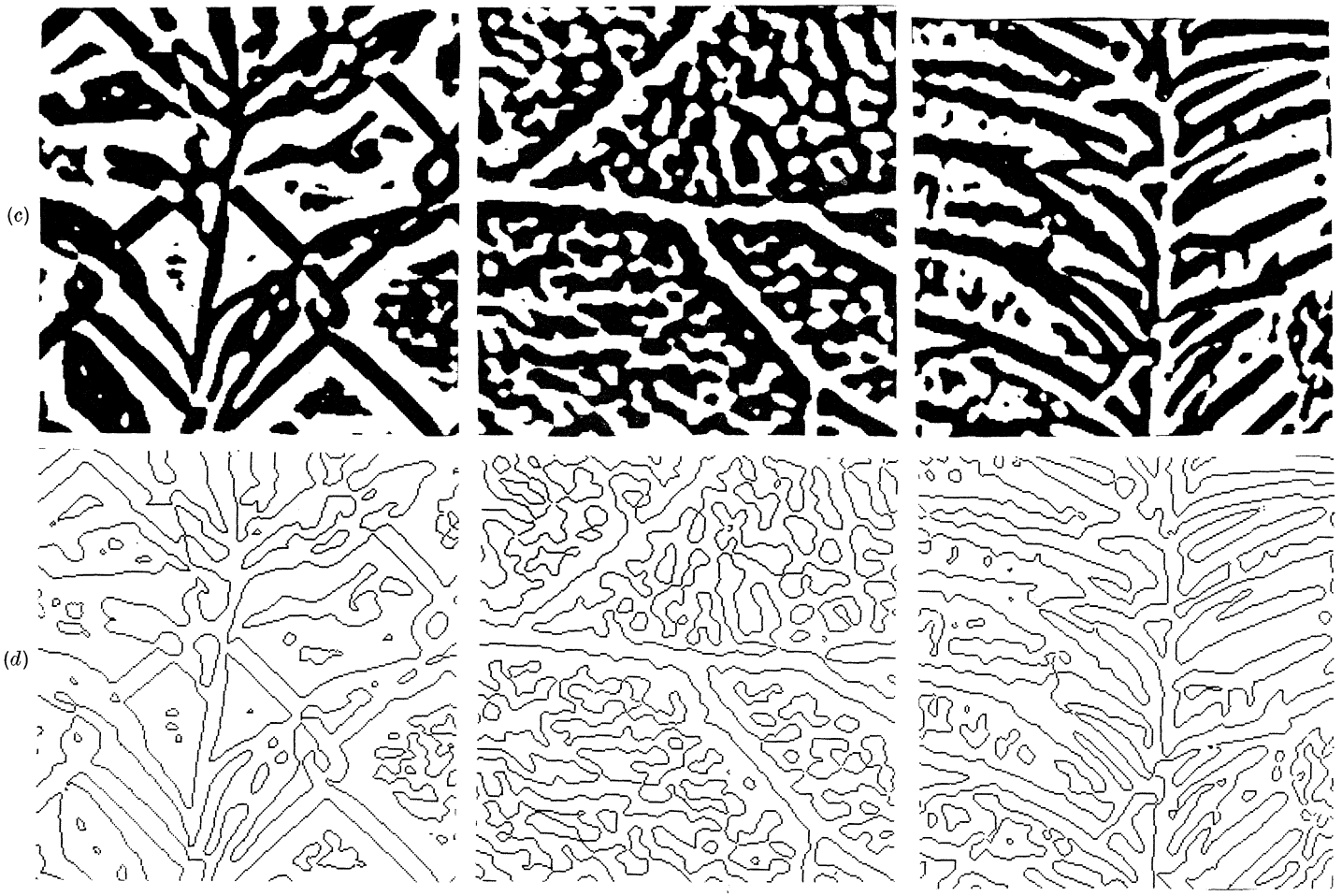


FIGURE 4. Examples of zero-crossing detection by means of $\nabla^2 G$. Row (*a*) shows three images and row (*b*) shows their convolutions with the $\nabla^2 G$ filter of figure 2 ($w = 2\sigma = 8$), zero being represented by an intermediate grey. In row (*c*), positive values are shown white, and negative, black; and in row (*d*) only the zero-crossings appear.

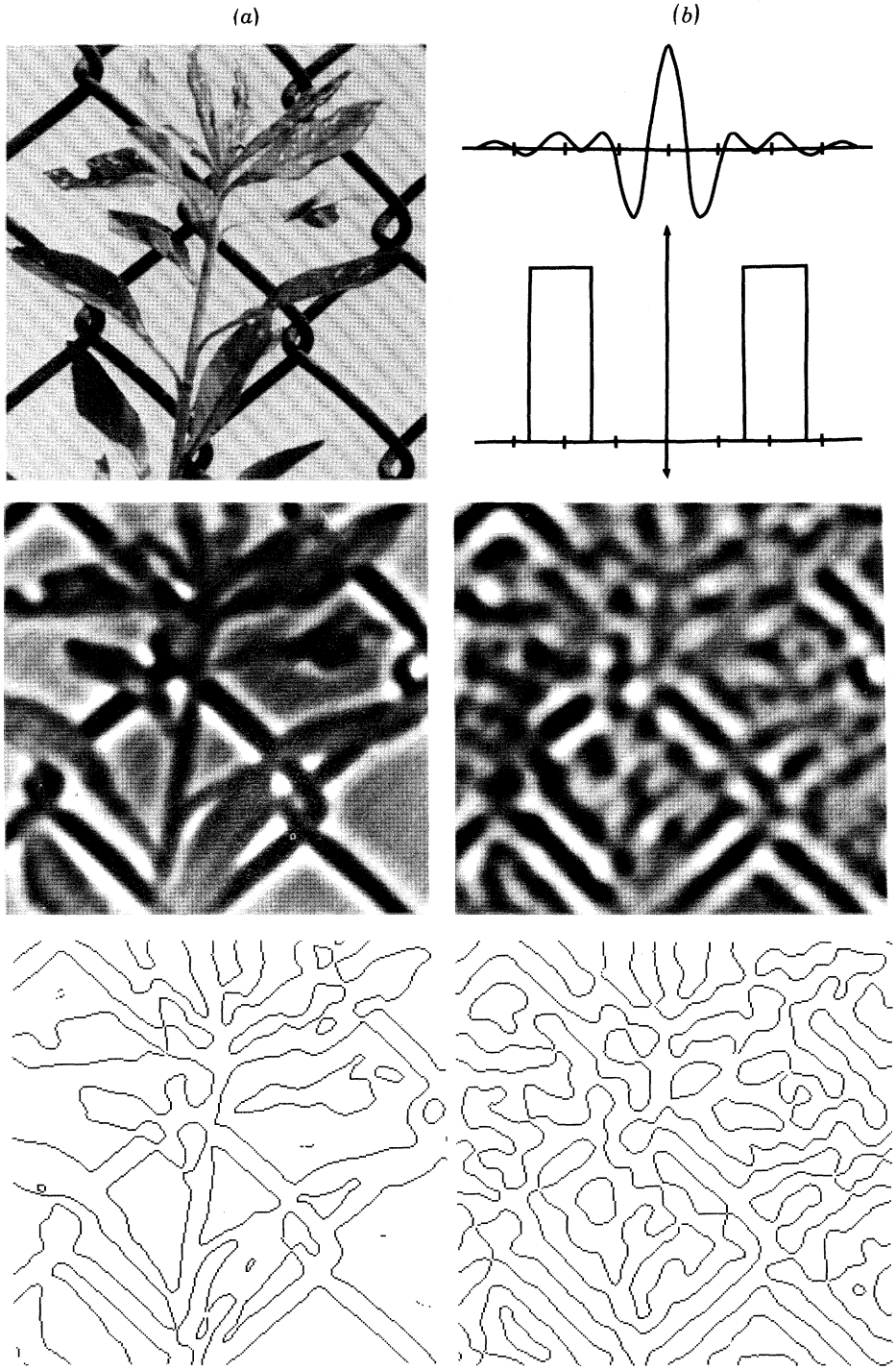
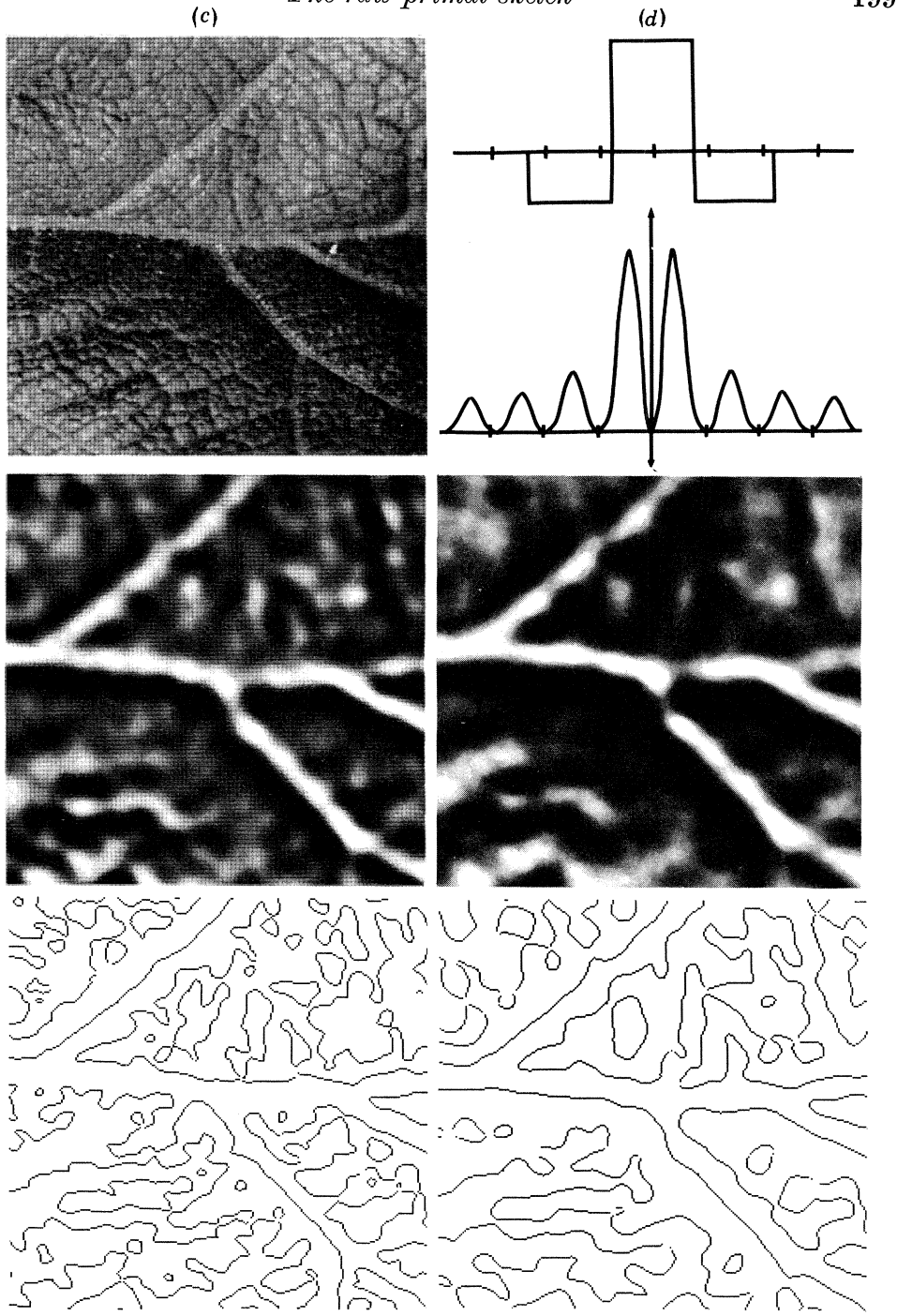


FIGURE 5. Comparison of the performance of $\nabla^2 G$ with that of similar filters. Column (*a*) shows an image, its convolution with $\nabla^2 G$ and the resulting signed zero-crossings. Column (*b*) contains the same sequence, but for the pure one-octave bandpass filter shown, with its Fourier transform, at the top of the column. The zero-crossing array contains echoes of the strong edges in the image. Columns (*c*) and (*d*) exhibit the same



analysis of another image, except that here, $\nabla^2 G$ is compared with a square-wave approximation to the second derivative. The widths of the central excitatory regions of the filters are the same for each comparison pair, being 12 for (*a*) and (*b*), and 18 for (*c*) and (*d*). The square-wave filter sees relatively few zero-crossings.

appear white, and very negative ones, black. In the third row, all positive values appear completely white, and all negative ones are black, and the fourth row shows just the loci of zero values. It will be observed that these delineate well the visible edges in the images. (See the legend for more details.) It remains only to break the zero value loci into oriented line segments.

It is interesting to compare the zero-crossings found by means of $\nabla^2 G$ with those found by means of similar operators that, according to our arguments, are not optimal. Our choice of the Gaussian filter was based on the requirements of simultaneous localization in the frequency and spatial domains. We therefore show examples in which each of these requirements is severely violated. An ideal one-octave bandpass filter satisfies the localization requirement in the frequency domain, but violates it in the spatial domain. The reason is that strict band-limiting gives rise to sidelobes in the spatial filter, and the consequence of these is that, in the zero-crossing image, strong intensity changes give rise to echoes as well as to the directly corresponding zero-crossings (see figure 5). These echoes have no direct physical correlate, and are therefore undesirable for early visual processing.

On the other hand, if one cuts off the filter in the spatial domain, one acquires sidelobes in the frequency domain. Figure 5 also shows a square-wave approximation to the second derivative operator, together with an example of the zero-crossings to which it gives rise. This operator sees fewer zero-crossings, essentially because it is averaging out the changes that occur over a wider range of scales.

Interestingly, Rosenfeld & Kak (1976, pp. 281–4) discuss the Laplacian in relation to 'edge' detection, but they do not report its having been used very effectively. One reason for this is that it is not very effective unless it is used in a band-limited situation and one uses its zero-crossings, and these ideas do not appear in the computer vision literature (see, for example, Rosenfeld & Kak 1976, fig. 10, for how the Laplacian has previously been used). In fact, the idea of using narrow bandpass differential operators did not appear until the human stereo theory of Marr & Poggio (1979), which was also the first theory to depend primarily on zero-crossings.

Another, more practical, reason why 'edge-detecting' operators have previously been less than optimally successful in computer vision is that most current operators examine only a very small part of the image, their 'receptive fields' are of the order of 10 to 20 image points at most. This contrasts sharply with the smallest of Wilson's four psychophysical channels, the receptive field of which must cover over 500 foveal cones (see figure 4).

Finally, notice that G'', and hence $\nabla^2 G$, is approximately a second derivative operator, because its Fourier transform is $-4\pi^2\omega^2 \exp(-\sigma^2\omega^2)$, which behaves like $-\omega^2$ near the origin.

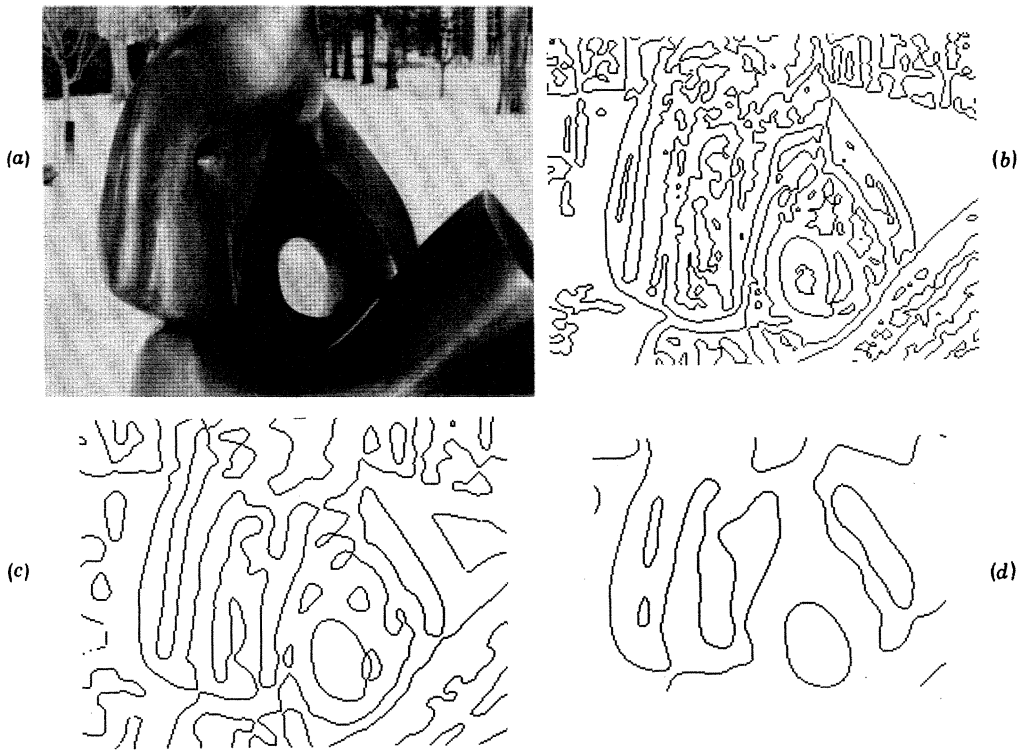


FIGURE 6. The image (*a*) has been convolved with $\nabla^2 G$ having $w = 2\sigma = 6$, 12 and 24 pixels. These filters span approximately the range of filters that operate in the human fovea. In (*b*), (*c*) and (*d*) are shown the zero-crossings thus obtained. Notice the fine detail picked up by the smallest. This set of figures neatly poses our next problem: how does one combine all this information into a single description?

COMBINING INFORMATION FROM DIFFERENT CHANNELS

The signals transmitted through channels that do not overlap in the Fourier domain will be generally unrelated unless the underlying signal is constrained. The critical question for us here is, therefore (and we are indebted to T. Poggio for conversations on this point), what additional information needs to be taken into account when we consider how to combine information from the different channels to form a primitive description of the image? In other words, are there any general physical constraints on the structure of the visual world that allow us to place valid restrictions on the way in which information from the different channels may be combined? Figure 6 illustrates the problem that we have to solve.

The spatial coincidence assumption

The additional information that we need here comes from the constraint of spatial localization, which we defined in the previous section. It states that the physical phenomena that give rise to intensity changes in the image are spatially

localized. Since it is these changes that produce zero-crossings in the filtered images, it follows that if a discernible zero-crossing is present in a channel centred on wavelength λ_0, there should be a corresponding zero-crossing at the same spatial location in channels for wavelengths $\lambda > \lambda_0$. If this ceases to be true at some wavelength $\lambda_1 > \lambda_0$, it will be for one of two reasons: either (*a*) two or more local intensity changes are being averaged together in the larger channel; or (*b*) two independent physical phenomena are operating to produce intensity changes in the same region of the image but at different scales. An example of situation (*a*) would be a thin bar, whose edges will be accurately located by small channels but not by large ones. Situations of this kind can be recognized by the presence of two nearby zero-crossings in the smaller channels. An example of situation (*b*) would be a shadow superimposed on a sharp reflectance change, and it can be recognized if the zero-crossings in the larger channels are displaced relative to those in the smaller. If the shadow has exactly the correct position and orientation, the locations of the zero-crossings may not contain enough information to separate the two physical phenomena, but, in practice, this situation will be rare.

We can therefore base the parsing of sets of zero-crossing segments from different $\nabla^2 G$ channels on the following assumption, which we call the *spatial coincidence assumption*:

> If a zero-crossing segment is present in a set of independent $\nabla^2 G$ 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 may be taken to indicate the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, illumination, depth or surface orientation).

In other words, provided that the zero-crossings from independent channels of adjacent sizes coincide, they can be taken together. If they do not, they probably arise from distinct surfaces or physical phenomena. It follows that the minimum number of channels required is two, and that provided the two channels are reasonably separated in the frequency domain, and their zero-crossings agree, the combined zero-crossings can be taken to indicate the presence of an edge in the image.

The parsing of sets of zero-crossing segments

Figure 6 shows the zero-crossings obtained from two channels whose dimensions are approximately the same as the two sustained channels present at the fovea in the human visual system (Wilson & Bergen 1979). We now derive the parsing rules needed for combining zero-crossings from the different channels.

Case (1): *isolated edges*

For an isolated, linearly disposed intensity change, there is a single zero-crossing present at the same orientation in all channels above some size that depends upon the channel sensitivity and the spatial extent of the edge. This set of zero-

crossings may, therefore, be combined into a symbol that we shall call an edge-segment, with the attributes of edge-amplitude and width, which we may obtain as follows.

Calculation of edge-amplitude. Because the assumptions that we have made mean that the type of intensity change involved is a simple one, we can, in fact, use what Marr (1976 figure 1) called the selection criterion, according to which one

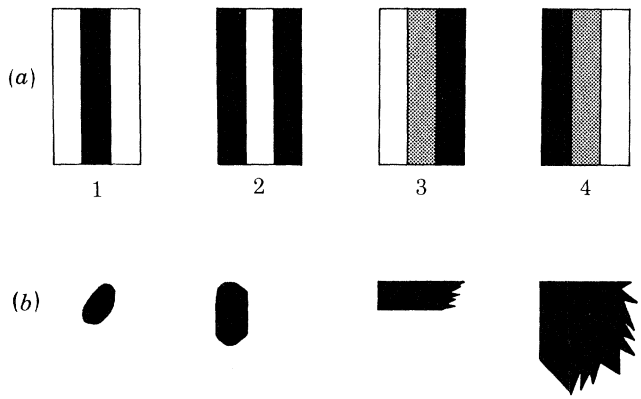


FIGURE 7. Parsing of sets of zero-crossing segments. (*a*) If zero-crossing segments lie close and roughly parallel (as in profile (*a*) of column 3 above), larger masks cannot be used, only the smaller masks. There are four possible configurations, shown in (1)–(4), and the figure represents the way in which the contrast changes across the edge. Each of these cases needs to be detected separately. (*b*) If the bar- or edge-segments are terminated, special descriptors are required. Doubly terminated bars, with $l \leqslant 3w$, are called *blobs* and the other assertions are labelled *terminations*. These are illustrated here for one contrast sign. Termination assertions may mark only a discontinuity in edge orientation, but it is often useful later on to have such positions explicitly available.

selects the smallest channel to which the intensity change is essentially indistinguishable from a step function, and uses that channel alone to estimate the contrast by means of the amplitude ν derived above. If one has just two independent channels with amplitudes ν_1 and ν_2, an approximation to the edge amplitude is $\sqrt{(\nu_1^2+\nu_2^2)}$.

Calculation of width. The *width* of the edge in this case can also be estimated from the channel selected according to the selection criterion. For a narrow channel with central wavelength λ, the physical notion of width corresponds to the distance over which intensity increases. This distance is $\frac{1}{2}\lambda$, which is approximately w, the width of the central excitatory region of the receptive field associated with the most excited channel (in fact, $\lambda = \pi w$).

Case (2): *bars*

If two parallel edges with opposite contrast lie only a small distance d apart in the image, zero-crossings from channels with associated wavelength that exceeds about $2d$ cannot be relied upon to provide accurate information about the positions

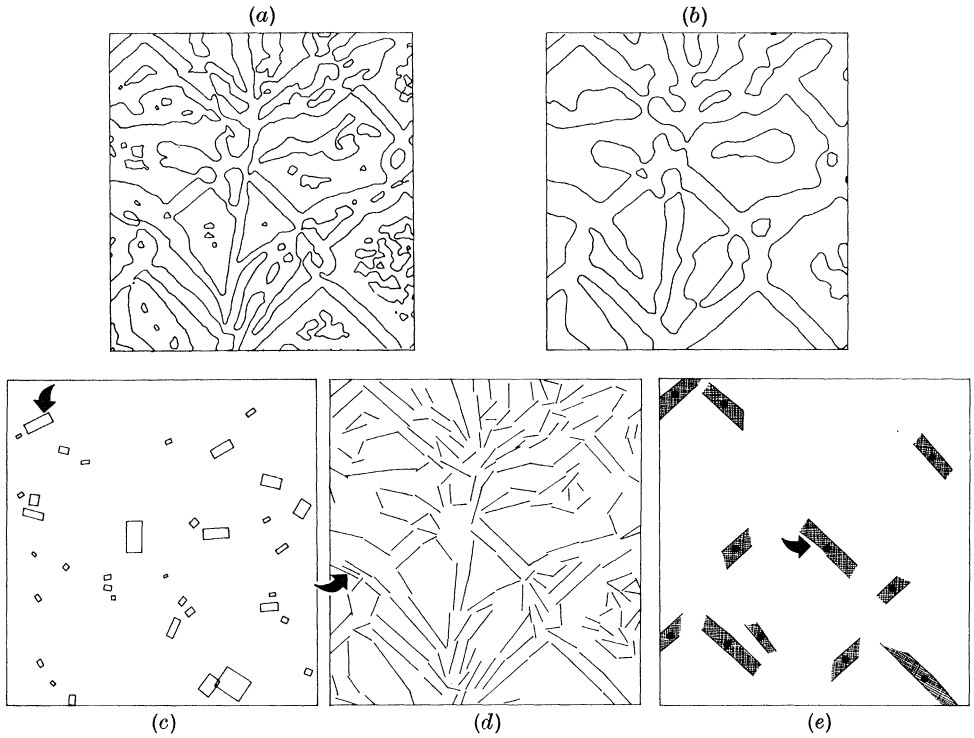


FIGURE 8. Combination of information from two channels. In (*a*) and (*b*) are shown the zero-crossings obtained from one of the images of figure 4, by means of masks with $w = 9$ and 18. Because there are no zero-crossings in the larger channel that do not correspond to zero-crossings in the smaller channel, the locations of the edges in the combined description also correspond to (*a*). In (*c*), (*d*) and (*e*) are shown symbolic representations of the descriptors attached to the locations marked in (*a*): (*c*) shows the blobs; (*d*), the local orientations assigned to the edge segments; and (*e*), the bars. These diagrams show only the spatial information contained in the descriptors. Typical examples of the full descriptors are as follows.

(BLOB (POSITION 146 21)	(EDGE (POSITION 104 23)	(BAR (POSITION 118 134)
(ORIENTATION 105)	(ORIENTATION 120)	(ORIENTATION 120)
(CONTRAST 76)	(CONTRAST −25)	(CONTRAST −25)
(LENGTH 16)	(LENGTH 25)	(LENGTH 25)
(WIDTH 6))	(WIDTH 4))	(WIDTH 4))

The descriptors to which these correspond are marked with arrows. The resolution of this analysis of the image of figure 4 roughly corresponds to what a human would see when viewing it from a distance of about 6 ft (1.83 m).

or contrasts of the edges. In these circumstances, the larger channels must be ignored, and the description formed solely from small channels of which the zero-crossing segments do superimpose. An edge can have either positive or negative contrast, and so two together give us the four situations shown in figure 7*a*. There is, of course, no reason why the two edges should have the same contrast, and the contrast of each edge must be obtained individually from the smallest channels

($w < d$). Two other parameters are useful; one is the average orientation of the two zero-crossing segments, and the other is their average separation.

Our case (2) applies only to situations in which neither zero-crossing segment terminates and they both remain approximately parallel (w or less apart). When the two edges are closer together than w for the smallest available channel, the zero-crossings associated with even the smallest channel will not accurately reflect the positions of the two edges, they will over estimate the distance between them. If the two edges have opposite contrasts that are not too different in absolute magnitude, the position of the centre of the 'line segment' so formed in the image will be the midpoint of the two corresponding zero-crossings. In these circumstances, the parameters associated with the line segment will be more reliable than those associated with each individual edge.

Case (3): *blobs and terminations*

It frequently happens that the zero-crossing segments do not continue very far across the image. Two parallel segments can merge, or be joined by a third segment, and in textured images they often form small closed curves (see figure 6), which are quite small compared to the underlying field size. Both situations can give rise to anomalous effects at larger channel sizes, and so are best made explicit early on. Following Marr (1976*b*), the closed contours we call BLOBS, and assign to them a length, width, orientation and (average) contrast; and the terminations are assigned a position and orientation (see figure 7*c*).

Remarks

Two interesting practical details have emerged from our implementation. First, the intensity changes at each edge of a bar are, in practice, rarely the same, so it is perhaps more proper to think of the BAR descriptor as a primitive grouping predicate that combines two edges the contrasts of which are specified precisely by the smallest channel. Brightness within the area of the bar will, of course, be constant. Secondly, it is often the case that the zero-crossings from the small and from the large masks roughly coincide, but those from the small mask weave around much more, partly because of the image structure and partly because of noise and the image tesselation. Local orientation has little meaning over distances shorter than the width w of the central excitatory region of the $\nabla^2 G$ filter, so if the zero-crossings from the smaller filter are changing direction rapidly locally, the orientation derived from the larger mask can provide a more stable and more reliable measure.

IMPLICATIONS FOR BIOLOGY

We have presented specific algorithms for the construction of the raw primal sketch, and we now ask whether the human visual system implements these algorithms or something close to them. There are two empirically accessible

characteristics of our scheme. The first concerns the underlying convolutions and zero-crossing segments, and the second, whether zero-crossing segments from the different channels are combined in the way that we have described.

Detection of zero-crossing segments

According to our theory, the most economical way of detecting zero-crossing segments requires that the image first be filtered through at least two independent $\nabla^2 G$ channels, and that the zero-crossings then be found in the filtered outputs. These zero-crossings may be divided into short, oriented zero-crossing segments.

The empirical data

Recent psychophysical work by Wilson & Giese (1977), Wilson & Bergen (1979) (see also Macleod & Rosenfeld 1974), has led to a precise quantitative model of the orientation-dependent spatial-frequency-tuned channels discovered by Campbell & Robson (1968). At each point in the visual field, there are four such channels spanning about three octaves, and their peak sensitivity wavelength increases linearly with retinal eccentricity. The larger two channels at each point are transient and the smaller two are sustained. These channels can be realized by linear units with bar-shaped receptive fields made of the difference of two Gaussian distributions, with excitatory to inhibitory space constants in the ratio of 1:1.75 for the sustained, and 1:3.0 for the transient, channels (Wilson & Bergen 1979). The largest receptive field at each point is about four times the smallest.

This state of affairs is consistent with the neurophysiology since Hubel & Wiesel (1962) originally defined simple cells by the linearity of their response, and they reported many bar-shaped receptive fields. In addition, simple cell receptive field sizes increase linearly with eccentricity (Hubel & Wiesel 1974, fig. 6*a*), and the scatter in size at each location seems to be about 4:1 (Hubel & Wiesel 1974, fig. 7). It is therefore tempting to identify at least some of the simple cells with the psychophysical channels. If so, the first obvious way of making the identification is to propose that the simple cells measure the second directional derivatives, thus perhaps providing the convolution values from which zero-crossing segments are subsequently detected.

There are, however, various reasons why this proposal can probably be excluded. They are:

(1) If the simple cells are essentially performing a linear convolution that approximates the second directional derivative, why are they so orientation sensitive? Three measurements, in principle, suffice to characterize the second derivative completely and, in practice, the directional derivatives measured along four orientations are apparently enough for this stage (see Marr 1976*b*; Hildreth, in preparation), and yet simple cells divide the domain into about 12 orientations.

(2) Schiller *et al.* (1976*b*, pp. 1324–5) found that the orientation sensitivity of simple cells is relatively independent of the strength of flanking inhibition, and

of the separation and lengths of the positive and negative subfields of the receptive field of the cell. In addition, tripartite receptive fields did not appear to be more orientation sensitive than bipartite ones. These points provide good evidence that simple cells are not linear devices.

(3) If the simple cells perform the convolution, what elements find the zero-crossings and implement the spatial part of the computation, lining the zero-crossings up with the convolution orientations, for example?

Wilson's channel data is consistent with $\nabla^2 G$

Wilson's DOG functions are very similar to $\nabla^2 G$, and probably indistinguishable by means of his experimental technique, which yields about 10% accuracy (H. G. Wilson, personal communication). In appendix B, we show: (*a*) that $\nabla^2 G$ is the limit of the DOG function as σ_i/σ_e, the ratio of the inhibitory to excitatory space constants, tends to unity; and (*b*) that if an approximation to $\nabla^2 G$ is to be constructed out of the difference of two Gaussian distributions, one excitatory and the other inhibitory, the optimal choice on engineering grounds for σ_i/σ_e is about 1.6.

A specific proposal: lateral geniculate X*-cells carry* $\nabla^2 G * I$*, and some simple cells detect and represent zero-crossing segments*

It is known that retinal ganglion X-cells have receptive fields that are accurately described by the difference of two Gaussian distributions (Rodieck & Stone 1965; Ratliff 1965; Enroth-Cugell & Robson 1966). The positive and negative parts are not quite balanced (there is a response to diffuse illumination and it increases with intensity), and since the ganglion cells have a spontaneous resting discharge, they signal somewhat more than just the positive or just the negative part of such a convolution. Interestingly, there is little scatter in receptive field sizes of X-cells at a given location in the retina (Peichl & Wässle 1979).

There is some controversy about the way in which lateral geniculate receptive fields are constructed (cf. Maffei & Fiorentini 1972*b*), but it seems most likely that the on-centre geniculate X-cell fields are formed by combining a small number of on-centre retinal ganglion X-cell fields of which the centres approximately coincide (Cleland *et al.* 1971). It seems likely that the scatter in receptive field size arises in this way, since the amount of scatter required to account for the psychophysical findings is only a factor of two in both the X and the Y channels. Finally, lateral geniculate cells give a smaller response to diffuse illumination than do retinal ganglion cells, sometimes giving no response at all (Hubel & Wiesel 1961).

These facts lead us to a particularly attractive scheme, which, for simplicity, we present in idealized form.

(1) *Measurement of* $\nabla^2 G$. The sustained, or X-cell, geniculate fibres can be thought of as carrying either the positive or the negative part of $\nabla^2 G * I$, where the filter $\nabla^2 G$ of figure 2 is, in practice, approximated by a difference

of Gaussian convolution operator with centre-to-surround space constants in the ratio 1:1.75. (One should probably think of this as being a convolution on linear intensity values, rather than on their logarithms. The reason for this is that although the nerve signal in the retina is an adaptation term multiplied by $I/(I+K)$, where I is the incident illumination and $K = 800$ quanta per receptor per second (Alpern *et al.* 1970), in any given image the ratio of the darkest to the brightest portion rarely exceeds 25 (a local ratio of around

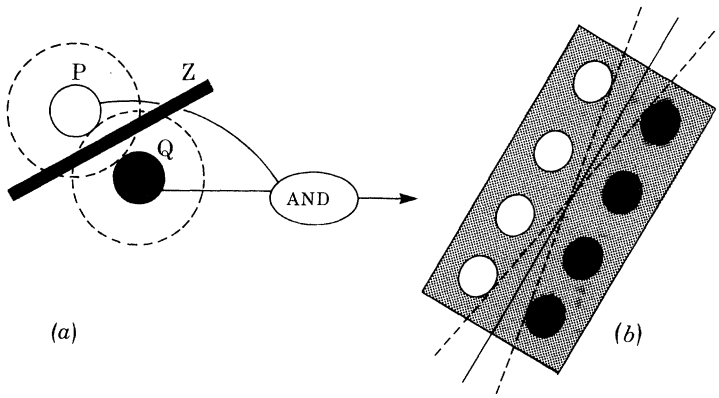


FIGURE 9. Proposed mechanism whereby some simple cells detect zero-crossing segments. In (*a*), if P represents an on-centre geniculate X-cell receptive field, and Q, an off-centre one, then if both are active, a zero-crossing Z in the Laplacian passes between them. If they are connected to a logical AND gate, as shown, then the gate will 'detect' the presence of the zero-crossing. If several are arranged in tandem, as in (*b*), and also connected by logical ANDs, the resulting operation detects an oriented zero-crossing segment within the orientation bounds given roughly by the dotted lines. This gives our most primitive model for simple cells. Ideally, one would like gates such that there is a response only if all (P, Q) inputs are active, and the magnitude of the response then varies with their sum. Marr & Ullman (1979) extend this model to include directional selectivity.

30 is seen as a light source (Ullman 1976)), and over such ranges this function does not depart far from linearity.) At each point in the visual field, there are two sizes of filter (the minimum required for combining zero-crossings between channels), and these correspond to Wilson & Bergen's (1979) N and S channels. The one-dimensional projection of the widths w of the central excitatory regions of these two channels scales linearly with eccentricity from 3.1′ and 6.2′ at the central fovea.

The basic idea behind our model for the detection of zero-crossings rests on the following observations: if an on-centre geniculate cell is active at location P and an off-centre cell is active at nearby location Q, then the value of $\nabla^2 G * I$ passes through zero between P and Q (see figure 9*a*). Hence, by combining the signals from P and Q through a logical AND operation, one can construct an operator for detecting when a zero-crossing segment (at some unknown orientation) passes

between P and Q (figure 9*a*). By adding nonlinear AND operations in the longitudinal direction, one can, in a similar way, construct an operator that detects oriented zero-crossing segments. It is easy to see that the pure logical operator of figure 9*b* will respond only to zero-crossing segments whose orientations lie within its sensitivity range (shown roughly dotted). We therefore propose:

(2) *Detection and representation of zero-crossing segments.* Part of the function of one subclass of simple cells is to detect zero-crossing segments. Their receptive fields include the construction shown in figure 9*b*, with the proviso that the non-linearities may be weaker than the pure logical ANDs shown there. It is, however, a critical feature of this model that the (P AND Q) interaction (figure 9*a*) across the zero-crossing segment should contain a strong nonlinear component and that the longitudinal interaction (e.g. between the ends in figure 9*b*) contains at least a weak nonlinear component. Marr & Ullman's (1979) full model for simple cells contains this organization, but includes additional machinery for detecting the direction of movement of the zero-crossing segment, and it is this that provides a role for the two larger transient channels.

(3) *Signalling amplitude.* Ideally, the output of the cell should be gated by the logical AND function of (2), but its value should be the average local amplitude ν associated with the zero-crossings along the segment. As we saw earlier, this may be found by measuring the average local value of the slope of the zero-crossings, which (in suitable units) is equal to the sum of the inputs to the cell.

(4) *Sampling density.* Finally, for this scheme to be successful, the sampling density of the function $\nabla^2 G * I$ must be great enough to ensure that the zero-crossings may subsequently be localized accurately enough to account for the findings about hyperacuity (see, for example, Westheimer & McKee 1977), which means roughly to within 5′. This implies an extremely high precision of representation, but in layer IV of the monkey's striate cortex, there apparently exists a myriad of small, centre–surround, non-oriented cells (Hubel & Wiesel 1968). Barlow (1979) and Crick *et al.* (1980) have suggested that these cells may be involved in the reconstruction of the $\nabla^2 G$ function to an adequate precision for hyperacuity.

The empirical consequences of this overall scheme are set out by Marr & Ullman (1979).

Combination of zero-crossings

Empirical predictions for psychophysics

There are several aspects of our algorithm, for combining zero-crossings from different channels, that are accessible to psychophysical experiment. They are: (*a*) the phase relations; (*b*) combination of zero-crossings from different channels, and (*c*) the special cases that arise when zero-crossings lie close to one another.

(1) *Phase relations.* Our theory predicts that descriptors need exist only for sets

of zero-crossings, from different channels, that coincide spatially (i.e. have a phase relation of 0 or π). Interestingly, Atkinson & Campbell (1974) superimposed 1 and 3 cycles/deg sinusoidal gratings of the same orientation, and found that the number of perceptual fluctuations per minute (which they called rate of monocular rivalry) was low near the in-phase, 0, and out-of-phase, π, positions, but reached a high plateau for intermediate phase positions. They concluded (p.161) that the visual system contains a device that 'seems to be designed to respond only to 0 and π phase relation. When ...[it] ... is active, it gives rise to a stable percept that is the sum of the two spatial frequency selective channels' (cf. also Maffei & Fiorentini 1972*a*). Our theory would predict these results, if the additional assumption were made that units exist that represent explicitly the edge segment descriptor formed by combining appropriately arranged zero-crossing segments.

(2) *The parsing process*. The main point here is that the description of an edge (its width, amplitude and orientation) can be obtained from the (smallest) channel whose zero-crossing there has maximum slope. As Marr (1976*b*, pp. 496–497) observed, this is consistent with Harmon & Julesz's (1973) finding that noise bands spectrally adjacent to the spectrum of a picture are most effective at suppressing recognition, since these have their greatest effect on mask response amplitudes near the important mask sizes. It also explains why removal of the middle spatial frequencies from such an image leaves a recognizable image of Lincoln behind a visible graticule (see Harmon & Julesz 1973). The reason is that the zero-crossings from different mask sizes fail to coincide, and the gap in the spectrum means that the small bar descriptors fail to account for this discrepancy. Hence, the assumption of spatial coincidence cannot be used, and the outputs from the different mask sizes are assumed to be due to different physical phenomena. Accordingly, they give rise to independent descriptions.

There is another possible but weaker consequence. If one makes the extra assumption, that the selection criterion is implemented by inhibitory connections between zero-crossing segment detectors that are spatially coincident and lying adjacent in the frequency domain, then one would expect to find an inhibitory interaction between channels at the cortical, orientation-dependent level. There is, in fact, evidence that this occurs (see, for example, Tolhurst 1972; de Valois 1977*a*).

(3) *Bar-detectors*. Case (2) of our parsing algorithm requires the specific detection of close, parallel, zero-crossing segments. This requires the existence of units sensitive, at each orientation, to one of the four cases (black bar, white bar, two dark edges, two light edges) and sensitive to their width (i.e. the distance separating the edges) rather than to spatial frequency characteristics of the whole pattern. Adaptation studies that lead to these conclusions for white bars and for black bars have recently been published (Burton *et al.* 1977; de Valois 1977*b*). If our algorithm is implemented by the human visual system, the analogous result should hold for the remaining two cases (see figure 7*a*).

(4) *Blob-detectors and terminations*. Case (3) of our parsing algorithm requires

the explicit representation of (oriented) blobs and terminations. Units that represent them should be susceptible to psychophysical adaptation, and, in fact, Nakayama & Roberts (1972) and Burton & Ruddock (1978) have found evidence for units that are sensitive to bars whose length does not exceed three times the width.

Consequences for neurophysiology

There are several ways of implementing the parsing process that we have described, but it is probably not worth setting them out in detail until we have good evidence from psychophysics about the parsing algorithm that is actually used and we know whether simple cells, in fact, implement the detection of zero-crossing segments. Without these pieces of information firm predictions cannot be made, but we offer the following suggestions as a possible framework for the neural implementation. (1) The four types of 'bar' detectors could be implemented at the very first, simple cell level (along the lines of figure 9, but being fed by three rows of centre–surround cells instead of two). (2) For relatively isolated edges, there should exist oriented edge-segment–detecting neurons that combine zero-crossing segment detectors (simple cells) from different channels when, and only when, the segments are spatially coincident. (3) Detectors for terminations and blobs (doubly-terminated oriented bars) seem to have been found already (Hubel & Wiesel 1962, 1968). Interestingly, Schiller *et al.* (1976*a*) found that even some simple cells are stopped. Our scheme is consistent with this since it requires such detectors at a very early stage.

Discussion

The concept of an 'edge' has a partly visual and partly physical meaning. One of our main purposes in this article is to make explicit this dual dependence: our definition of an edge rests lightly on the early assumptions of theorem 1 about directional derivatives and heavily on the constraint of spatial localization.

Our theory is based on two main ideas. First, one simplifies the detection of intensity changes by dealing with the image separately at different resolutions. The detection process can then be based on finding zero-crossings in a second derivative operator, which, in practice, can be the (non-oriented) Laplacian. The representation at this point consists of zero-crossing segments and their slopes. This representation is probably complete and is, therefore, in principle, invertible. This had previously been given only an empirical demonstration by Marr and by R. Woodham (see Marr 1978, fig. 7).

The subsequent step, of combining information from different channels into a single description, rests on the second main idea of the theory, which we formulated as the spatial coincidence assumption. Physical edges will produce roughly coincident zero-crossings in channels of nearby sizes. The spatial coincidence assumption asserts that the converse of this is true, that is the coincidence of zero-

crossings is sufficient evidence for the existence of a real physical edge. If the zero-crossings in one channel are not consistent with those in the others, they are probably caused by different physical phenomena, so descriptions need to be formed from both sources and kept somewhat separate.

Finally, the basic idea, that some simple cells detect and represent zero-crossing segments and that this is carried out simultaneously at different scales, has some implications for Marr & Poggio's (1979) stereo theory. According to various neurophysiological studies (Barlow *et al.* 1967; Poggio & Fischer 1978; von der Heydt *et al.* 1978), there exist disparity sensitive simple cells. The existence of such cells is consistent with our suggestion that they detect zero-crossing segments, but not with the idea that they perform a linear convolution equivalent to a directional derivative, since it is the primitive symbolic descriptions provided by zero-crossing segments that need to be matched between images, not the raw convolution values.

We thank K. Nishihara, T. Poggio and S. Ullman for their illuminating and helpful comments. This work was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defence and monitored by the Office of Naval Research, under contract number N00014-75-C-0643. D. M. was also supported by N.S.F. contract number 77-07569-MCS.

References

Alpern, M., Rushton, W. A. H. & Torii, S. 1970 The size of rod signals. *J. Physiol., Lond.* **206**, 193–208.

Atkinson, J. & Campbell, F. W. 1974 The effect of phase on the perception of compound gratings. *Vision Res.* **14**, 159–162.

Barlow, H. B. 1969 Pattern recognition and the responses of sensory neurons. *Ann. N.Y. Acad. Sci.* **156**, 872–881.

Barlow, H. B. 1979 Reconstructing the visual image in space and time. *Nature, Lond.* **279**, 189–190.

Barlow, H. B., Blakemore, C. & Pettigrew, J. D. 1967 The neural mechanism of binocular depth discrimination. *J. Physiol., Lond.* **193**, 327–342.

Bracewell, R. 1965 *The Fourier transform and its applications*. New York: MacGraw-Hill.

Burton, G. J., Nagshineh, S. & Ruddock, K. H. 1977 Processing by the human visual system of the light and dark contrast components of the retinal image. *Biol. Cybernetics* **28**, 1–9.

Burton, G. J. & Ruddock, K. H. 1978 Visual adaptation to patterns containing two-dimensional spatial structure. *Vision Res.* **18**, 93–99.

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. 1971 Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. *J. Physiol., Lond.* **217**, 473–496.

Cowan, J. D. 1977 Some remarks on channel bandwidths for visual contrast detection. *Neurosci. 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* (ed. F. O. Schmidt & F. G. Worden).

De Valois, K. K. 1977*a* Spatial frequency adaptation can enhance contrast sensitivity. *Vision Res.* **17**, 1057–1065.

De Valois, K. K. 1977*b* Independence of black and white: phase-specific adaptation. *Vision Res.* **17**, 209–215.

Enroth-Cugell, C. & Robson, J. G. 1966 The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol., Lond.* **187**, 517–552.

Graham, N. 1977 Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Res.* **17**, 637–652.

Harmon, L. D. & Julesz, B. 1973 Masking in visual recognition: effects of two-dimensional filtered noise. *Science N.Y.* **180**, 1194–1197.

von der Heydt, R., Adorjani, Cs., Hanny, P. & Baumgartner, G. 1978 Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Exp. Brain Res.* **31**, 523–545.

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. 1974 Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J. comp. Neurol.* **158**, 295–306.

Kulikowski, J. J. & King-Smith, P. E. 1973 Spatial arrangement of line, edge, and grating detectors revealed by subthreshold summation. *Vision Res.* **13**. 1455–1478.

Leipnik, R. 1960 The extended entropy uncertainty principle. *Inf. Control* **3**, 18–25.

Logan, B. F. Jr 1977 Information in the zero-crossings of bandpass signals. *Bell Syst. tech. J.* **56**. 487–510.

Macleod, I. D. G. & Rosenfeld, A. 1974 The visibility of gratings: spatial frequency channels or bar-detecting units? *Vision Res.* **14**, 909–915.

Maffei, L. & Fiorentini, A. 1972*a* Process of synthesis in visual perception. *Nature, Lond.* **240**, 479–481.

Maffei, L. & Fiorentini, A. 1972*b* Retinogeniculate convergence and analysis of contrast. *J. Neurophysiol.* **35**, 65–72.

Maffei, L. & Fiorentini, A. 1977 Spatial frequency rows in the striate visual cortex. *Vision Res.* **17**, 257–264.

Marr, D. 1970 A theory for cerebral neocortex. *Proc. R. Soc. Lond.* B **176**, 161–234.

Marr, D. 1976*a* Analyzing natural images: a computational theory of texture vision. *Cold Spring Harbor Symp. quant. Biol.* **40**, 647–662.

Marr, D. 1976*b* Early processing of visual information. *Phil. Trans. R. Soc. Lond.* B **275**, 483–524.

Marr, D. 1978 Representing visual information. A.A.A.S. 143rd Annual Meeting, Symposium on: Some mathematical questions in biology, February 1977. Published in *Lectures on mathematics in the life sciences* **10**, 101–180. Also available as *M.I.T. A.I. Lab. Memo* 415.

Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B **204**, 301–328.

Marr, D., Poggio, T. & Ullman, S. 1979 Bandpass channels, zero-crossings, and early visual information processing. *J. opt. Soc. Am.* **69**, 914–916.

Marr, D. & Ullman, S. 1979 Directional selectivity and its use in early visual processing. (In preparation.)

Mayhew, J. E. W. & Frisby, J. P. 1978 Suprathreshold contrast perception and complex random textures. *Vision Res.* **18**, 895–897.

Nakayama, K. & Roberts, D. J. 1972 Line-length detectors in the human visual system: evidence from selective adaptation. *Vision Res.* **12**, 1709–1713.

Peichl, L. & Wässle, H. 1979 Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *J. Physiol., Lond.* **291**, 117–141.

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.

Pollen, D. A., Lee, J. R. & Taylor, J. H. 1971 How does the striate cortex begin the reconstruction of the visual world? *Science N.Y.* **173**, 74–77.
Ratliff, F. 1965 *Mach bands: quantitative studies on neural networks in the retina*. San Francisco: Holden-Day.
Rodieck, R. W. & Stone, J. 1965 Analysis of receptive fields of cat retinal ganglion cells. *J. Neurophysiol.* **28**, 833–849.
Rosenfeld, A. & Kak, A. C. 1976 *Digital picture processing*. New York: Academic Press.
Sachs, M. B., Nachmias, J. & Robson, J. G. 1971 Spatial-frequency channels in human vision. *J. opt. Soc. Am.* **61**, 1176–1186.
Shapley, R. M. & Tolhurst, D. J. 1973 Edge detectors in human vision. *J. Physiol., Lond.* **229**, 165–183.
Schiller, P. H., Finlay, B. L. & Volman, S. F. 1976*a* 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. 1976*b* Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *J. Neurophysiol.* **39**, 1320–1333.
Ullman, S. 1976 On visual detection of light sources. *Biol. Cybernetics* **21**, 205–212.
Westheimer, G. & McKee, S. P. 1977 Spatial configurations for visual hyperacuity. *Vision Res.* **17**, 941–947.
Wilson, H. R. & Bergen, J. R. 1979 *A four mechanism model for spatial vision*. *Vision Res.* **19**, 19–32.
Wilson, H. R. & Giese, S. C. 1977 Threshold visibility of frequency gradient patterns. *Vision Res.* **17**. 1177–1190.

Appendix A

Theorem 1

Let l be an open line segment of the y-axis, containing the origin O. Suppose that $f(x, y)$ is twice continuously differentiable and that $N(l)$ is an open two-dimensional neighbourhood of l. Assume that $\partial^2 f/\partial x^2 = 0$ on l. Then, if $\partial f/\partial y$ is constant in $N(l)$, the slope of the second directional derivative taken perpendicular to l (i.e., the slope of $\partial^2 f/\partial x^2$) is greater than the slope of the zero-crossing along any other line through O.

Proof

Consider the line segment $\Omega = (r\cos\theta,\ r\sin\theta)$ for fixed θ and values of r sufficiently small that Ω lies entirely within $N(l)$ (see figure 10). Now writing f_{xx} for $\partial^2 f/\partial x^2$ etc., we have

$$\begin{aligned}(\partial^2 f/\partial\Omega^2)_{r,\theta} &= (f_{xx}\cos^2\theta + f_{xy}\,2\sin\theta\cos\theta + f_{yy}\sin^2\theta)_{r,\theta}\\ &= (f_{xx}\cos^2\theta)_{r,\theta},\end{aligned}$$

since the condition of the theorem that f_y be constant implies that f_{xy} and f_{yy} are both zero. As required, therefore, the above quantity is zero at $r = 0$ and has maximum slope when $\theta = 0$.

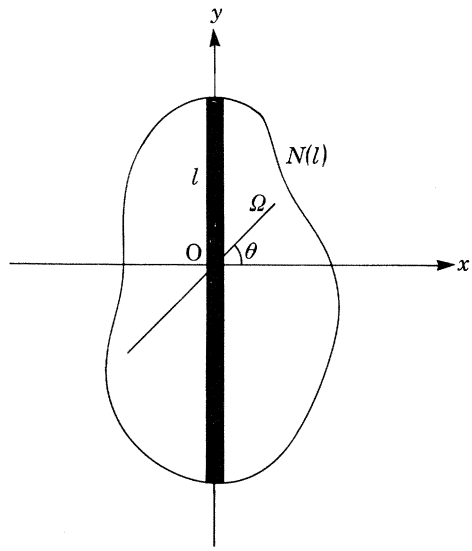


FIGURE 10. Diagram for theorems 1 and 2: l is a segment of the y-axis, containing the origin; $N(l)$ is a neighbourhood of it. Provided that $\partial f/\partial y$ is constant in $N(l)$, theorem 1 states that the orientation of the line of zero-crossings is perpendicular to the orientation at which the zero-crossings have maximum slope.

THEOREM 2

Let $f(x, y)$ be a real-valued, twice continuously differentiable function on the plane. Let l be an open line segment along the axis $x = 0$. Then the two conditions

$$\text{(i)}\quad \nabla^2 f = 0 \text{ on } l$$

$$\text{and (ii)}\quad \partial^2 f/\partial x^2 = 0 \text{ on } l$$

are equivalent if and only if $f(0, y)$ is constant or linear on l.

Proof

If $f(0, y)$ is linear on l, $\partial^2 f/\partial y^2 = 0$ on l. Hence, $\nabla^2 f = 0$ there implies that $\partial^2 f/\partial x^2 = 0$ on l too.

Conversely, if $\partial^2 f/\partial x^2 = \nabla^2 f = 0$ on l, then $\partial^2 f/\partial y^2 = 0$ on l, and so $f(0, y)$ varies at most linearly on l.

APPENDIX B

DOGS *and* $\nabla^2 G$

$\nabla^2 G$ is the limit of a DOG

Wilson's *DOG* function may be written

$$DOG(\sigma_e, \sigma_i) = [1/(2\pi)^{\frac{1}{2}}\sigma_e]\exp(-x^2/2\sigma_e^2) - [1/(2\pi)^{\frac{1}{2}}\sigma_i]\exp(-x^2/2\sigma_i^2), \qquad (3)$$

where σ_e and σ_i are the excitatory and inhibitory space constants. Writing $\sigma_e = \sigma$, and $\sigma_i = \sigma + \delta\sigma$, the right hand side varies with

$$(1/\sigma)\exp(-x^2/2\sigma^2) - [1/(\sigma+\delta\sigma)]\exp[-x^2/2(\sigma+\delta\sigma)^2] = \delta\sigma\,(\partial/\partial\sigma)\,(1/\sigma\exp[-x^2/2\sigma^2]). \quad (4)$$

This derivative is equal to $-(1/\sigma^2 - x^2/\sigma^4)\exp(-x^2/2\sigma^2)$, which equals G'' up to a constant (text equation 5).)

Approximation of $\nabla^2 G$ *by a* DOG

The function

$$DOG(\sigma_e, \sigma_i) = [1/(2\pi)^{\frac{1}{2}}\sigma_e]\exp(-x^2/2\sigma_e^2) - [1/(2\pi)^{\frac{1}{2}}\sigma_i]\exp(-x^2/2\sigma_i^2) \quad (5)$$

has Fourier transform

$$\widetilde{DOG}(\omega) = \exp(-\sigma_e^2\,\omega^2/2) - \exp(-\sigma_i^2\,\omega^2/2) \quad (6)$$

Notice that $\widetilde{DOG}(\omega)$ behaves like ω^2 for values of ω that are small compared with σ_e and σ_i, so that these filters, in common with $\nabla^2 G$, approximate a second derivative operator.

The problem with using a DOG to approximate $\nabla^2 G$ is to find a space constant that keeps the bandwidth of the filter small and yet allows the filter adequate sensitivity: for, clearly, as the space constants approach one another, the contributions of the excitatory and inhibitory components become identical and the sensitivity of the filter is reduced.

The bandwidths at half sensitivity and at half power and the peak sensitivity all depend together on the value of σ_i/σ_e in a way that is shown in figure 11. From this we see that: (i) the bandwidth at half sensitivity increases very slowly up to about $\sigma_i/\sigma_e = 1.6$, increases faster from there to $\sigma_i/\sigma_e = 3.0$, and is thereafter approximately constant; (ii) the peak sensitivity of the filter is desultory for small σ_i/σ_e, reaching about 33 % at $\sigma_i/\sigma_e = 1.6$. Since our aim is to create a narrow bandpass differential operator, we should choose σ_i/σ_e to minimize the bandwidth. Since the bandwidth is approximately constant for $\sigma_i/\sigma_e < 1.6$, and since sensitivity is low there, the minimal value one would in practice choose for σ_i/σ_e is around 1.6, giving a half-sensitivity bandwidth of 1.8 octaves and a half power bandwidth of 1.3 octaves.

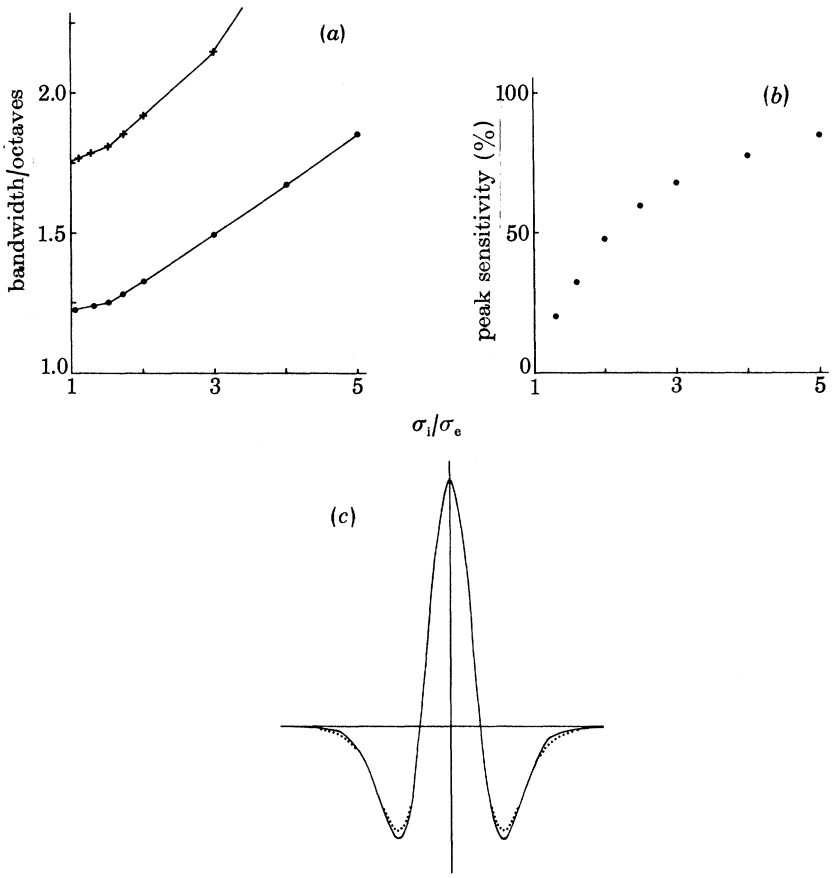


FIGURE 11. The values of certain parameters associated with difference-of-Gaussian (DOG) masks, with excitatory and inhibitory space constants σ_e and σ_i. (*a*) For various values of σ_e/σ_i, we show the half-sensitivity bandwidth (+) and the half-power bandwidth (●) of the filter. In (*b*) is shown its peak sensitivity in the Fourier plane. (The peak sensitivity of the excitatory component alone equals 100% on this scale.) (*c*) The arguments in the appendix show that the best engineering approximation to $\nabla^2 G$ using a DOG occurs with σ_i/σ_e around 1.6. In figure (*c*), this particular DOG is shown dotted against the operator $\nabla^2 G$ with the appropriate σ. The two profiles are very similar.

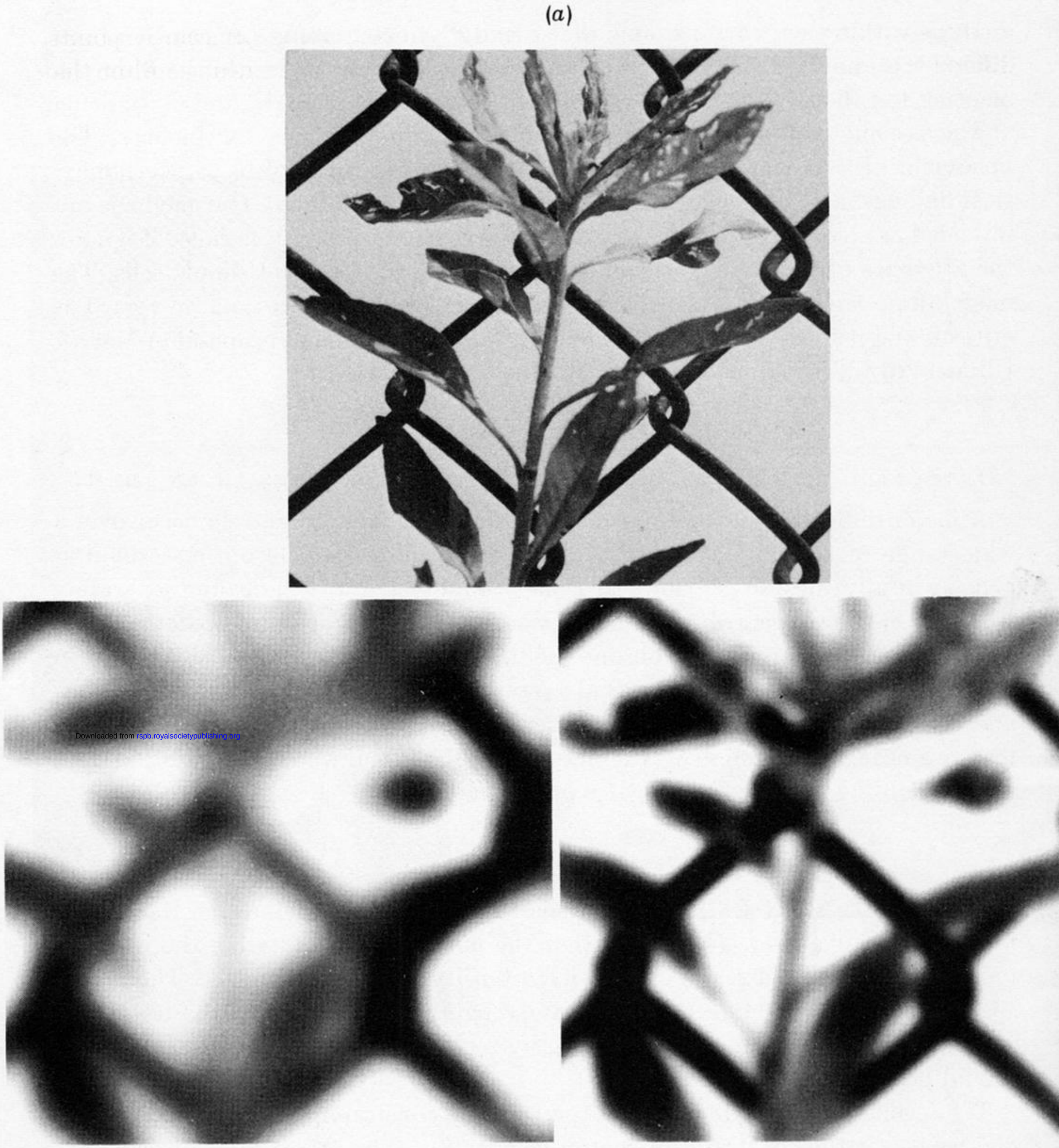
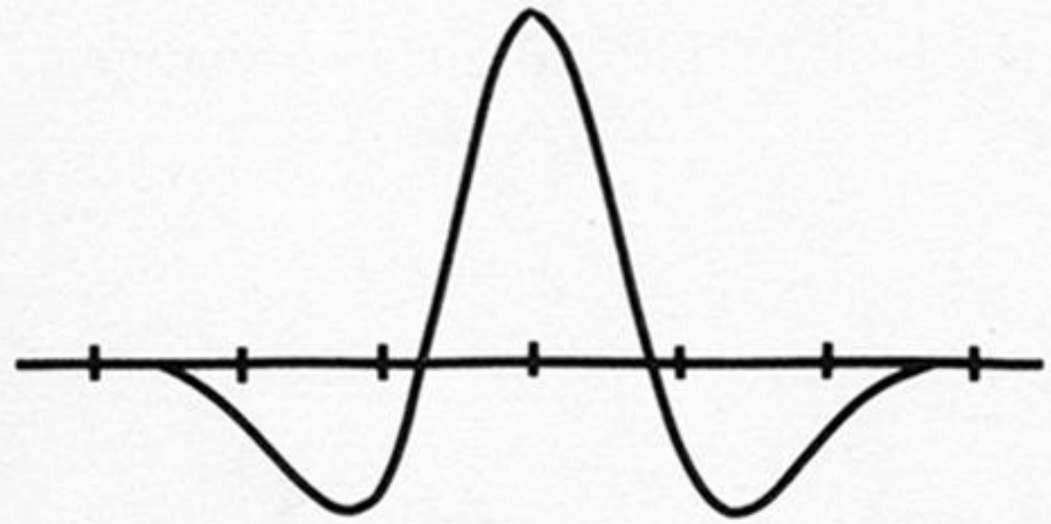
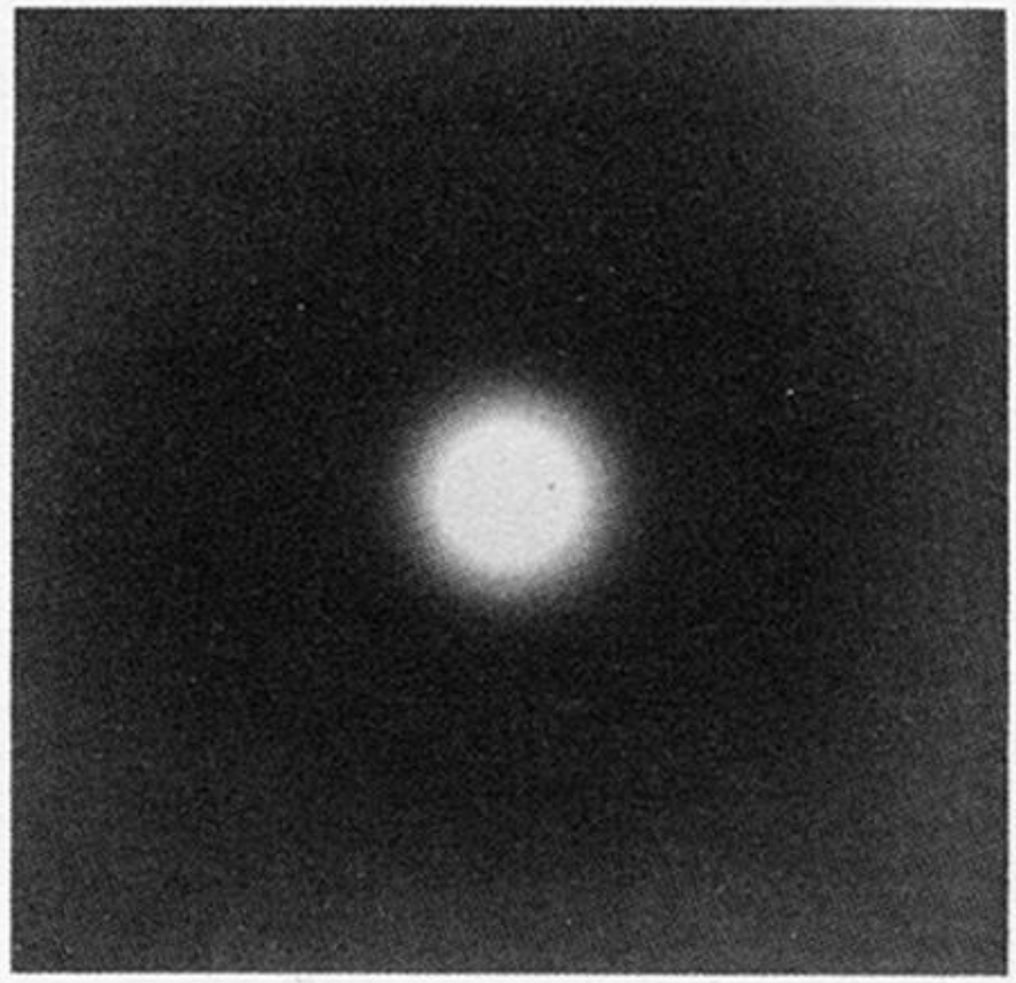


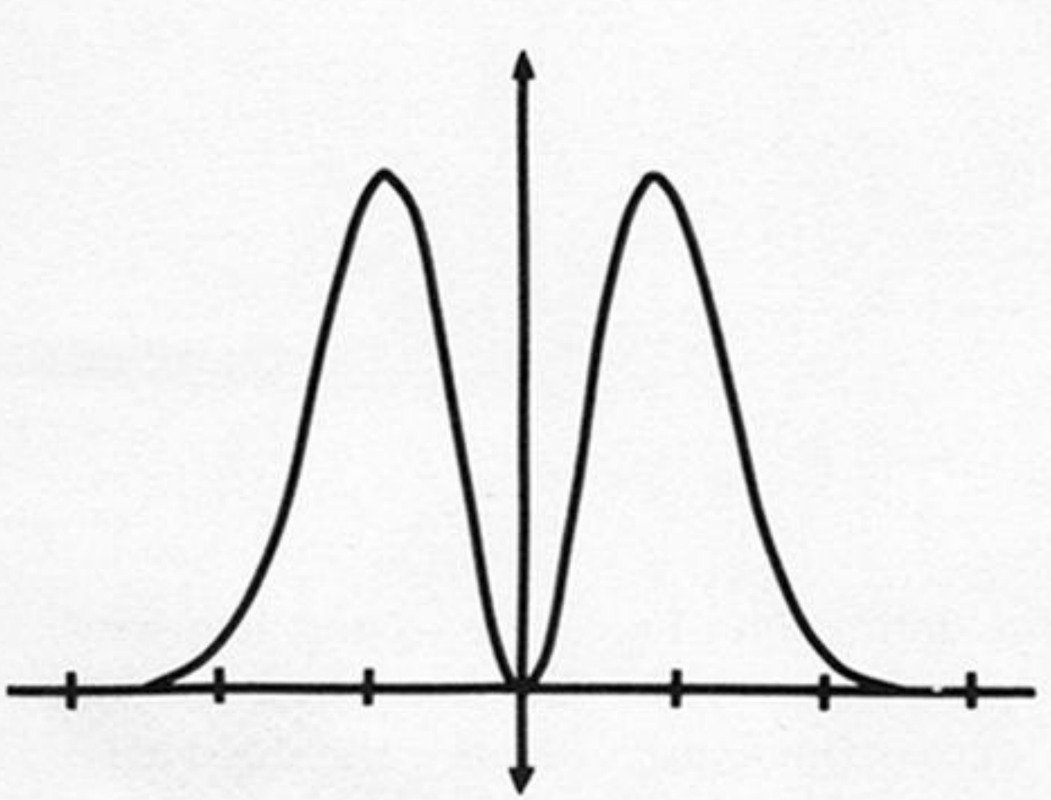
FIGURE 1. A local-average filtered image. In the original image (*a*), intensity changes can take place over a wide range of scales and no single operator will be very efficient at detecting all of them. The problem is much simplified in a Gaussian band-limited filtered image because there is effectively an upper limit to the rate at which changes can take place. The first part of our scheme can be thought of as decomposing the original image into a set of copies, each filtered like this, and detecting the intensity changes separately in each. In (*b*) the image is filtered with a Gaussian having $\sigma = 8$ picture elements, and, in (*c*), $\sigma = 4$. The image is 320×320 picture elements.

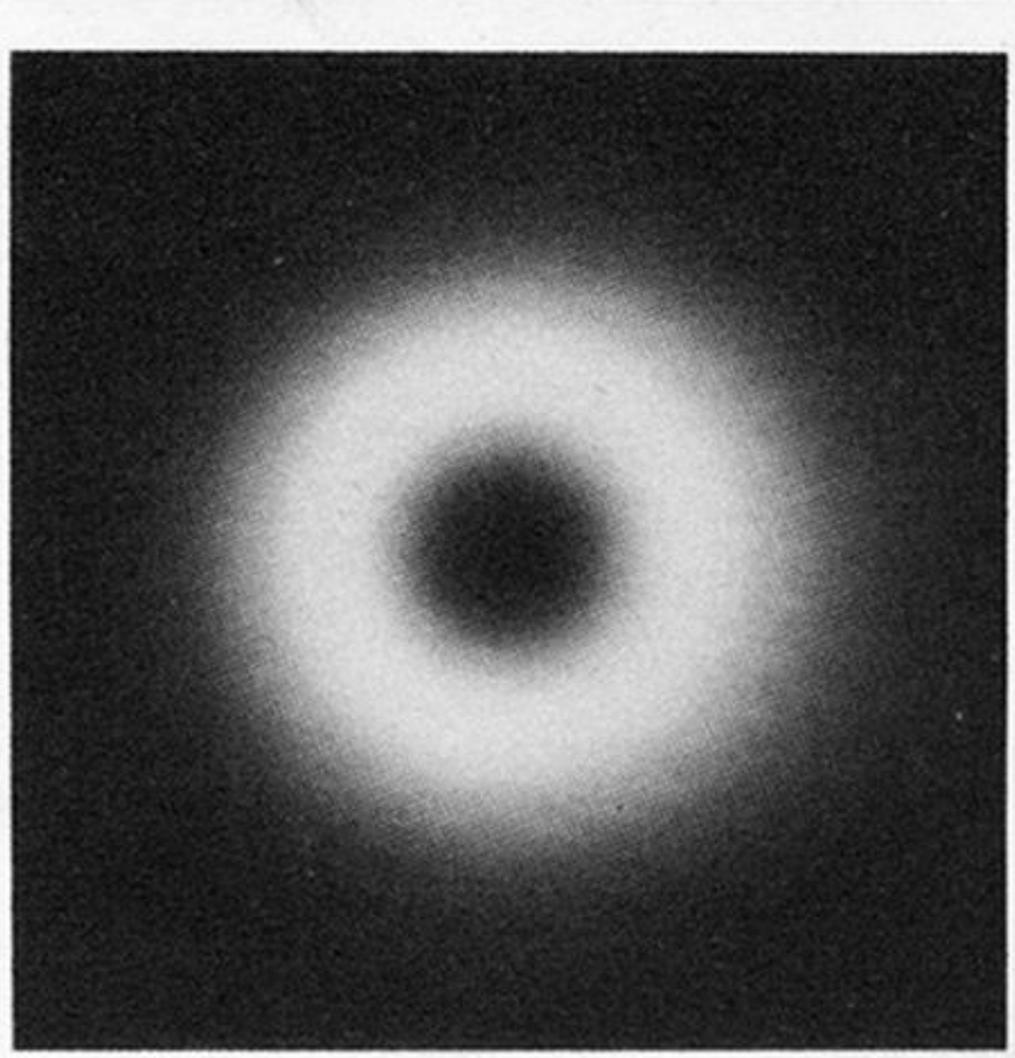


(a)



(c)





(d)

FIGURE 2. The operators G'' (equation 5) and $\nabla^2 G$: (*a*) shows G'', the second derivative of the one-dimensional Gaussian distribution; (*c*) shows $\nabla^2 G$, its rotationally symmetric two-dimensional counterpart; (*b*) and (*d*) exhibit their Fourier transforms.

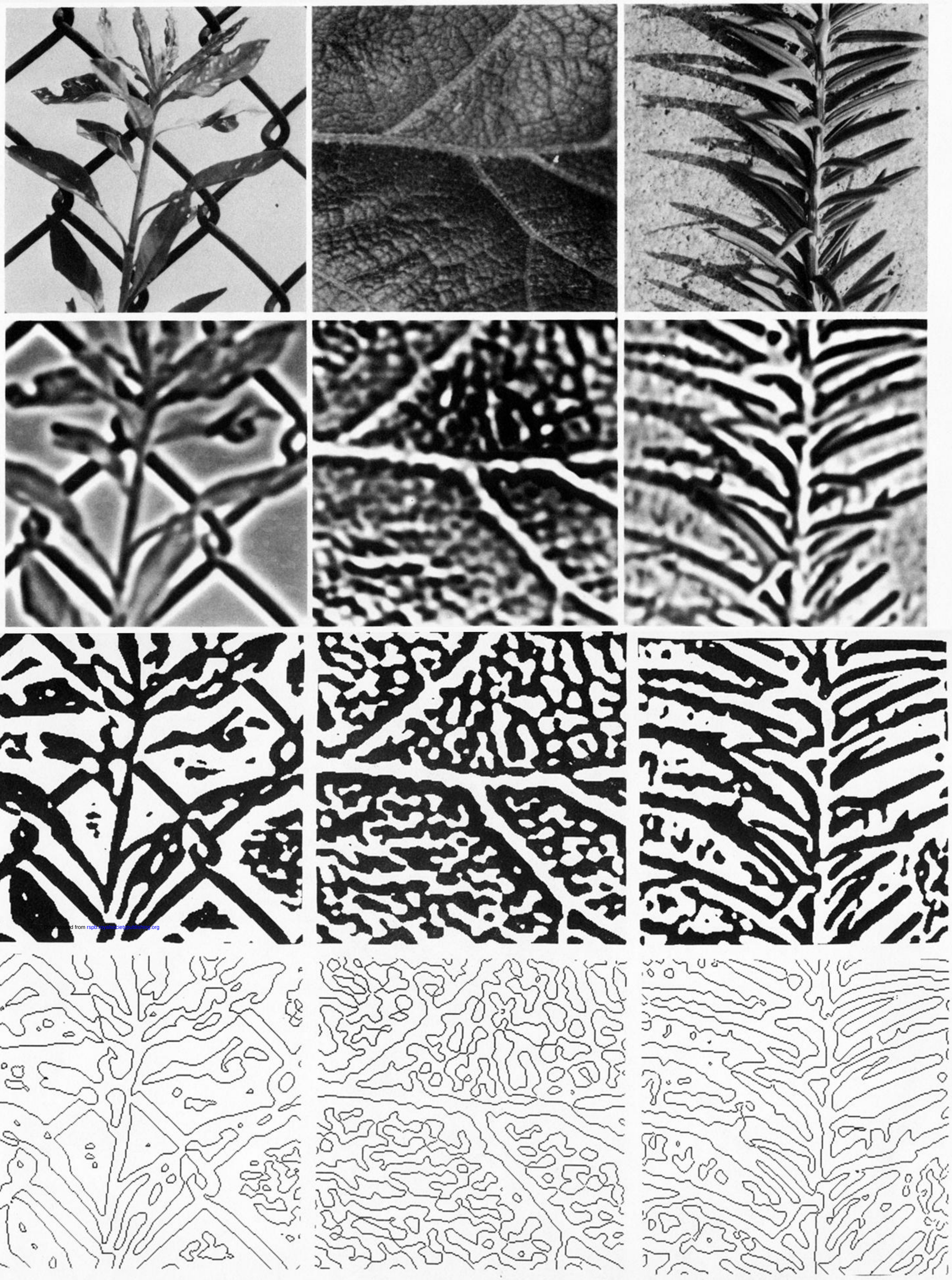
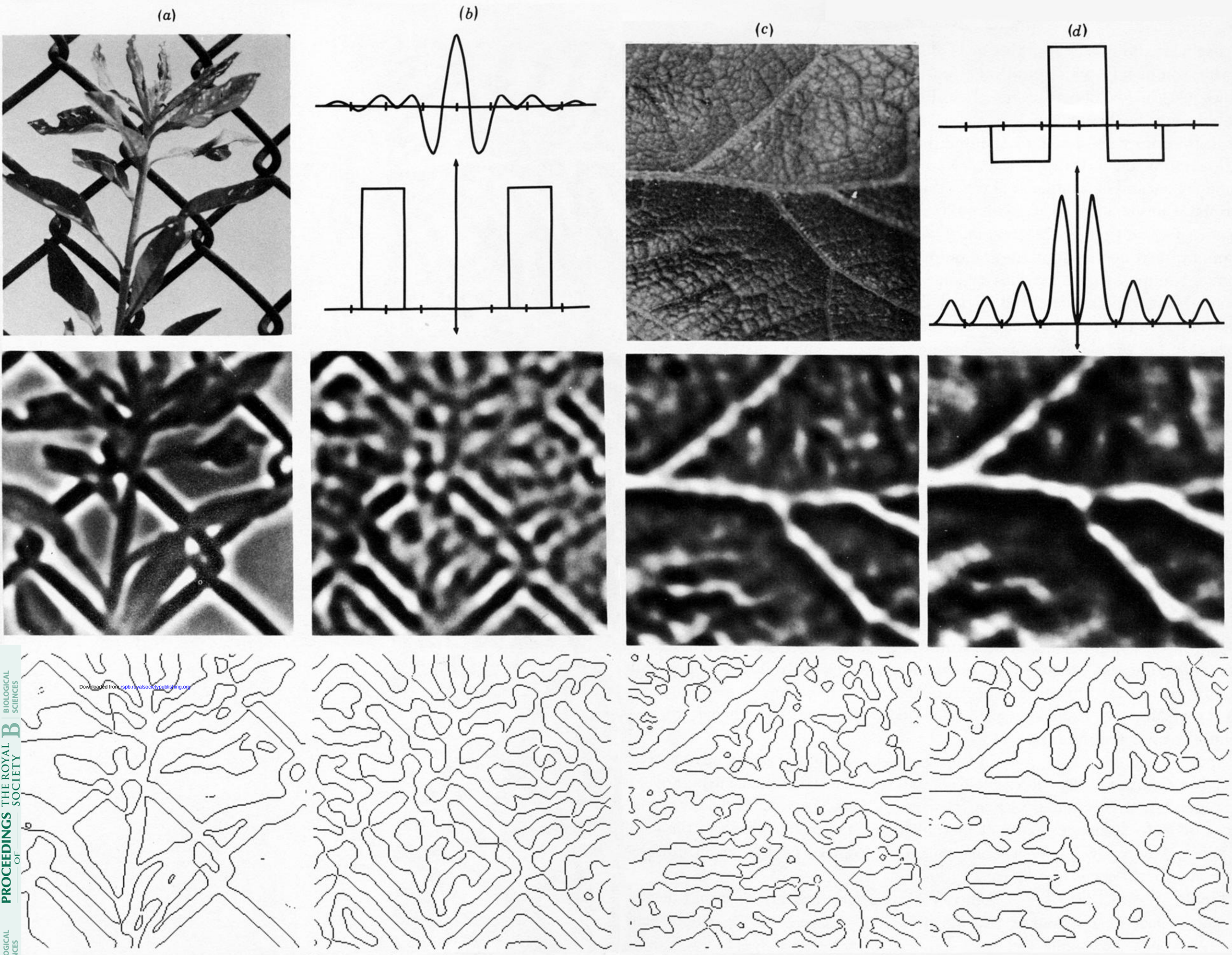


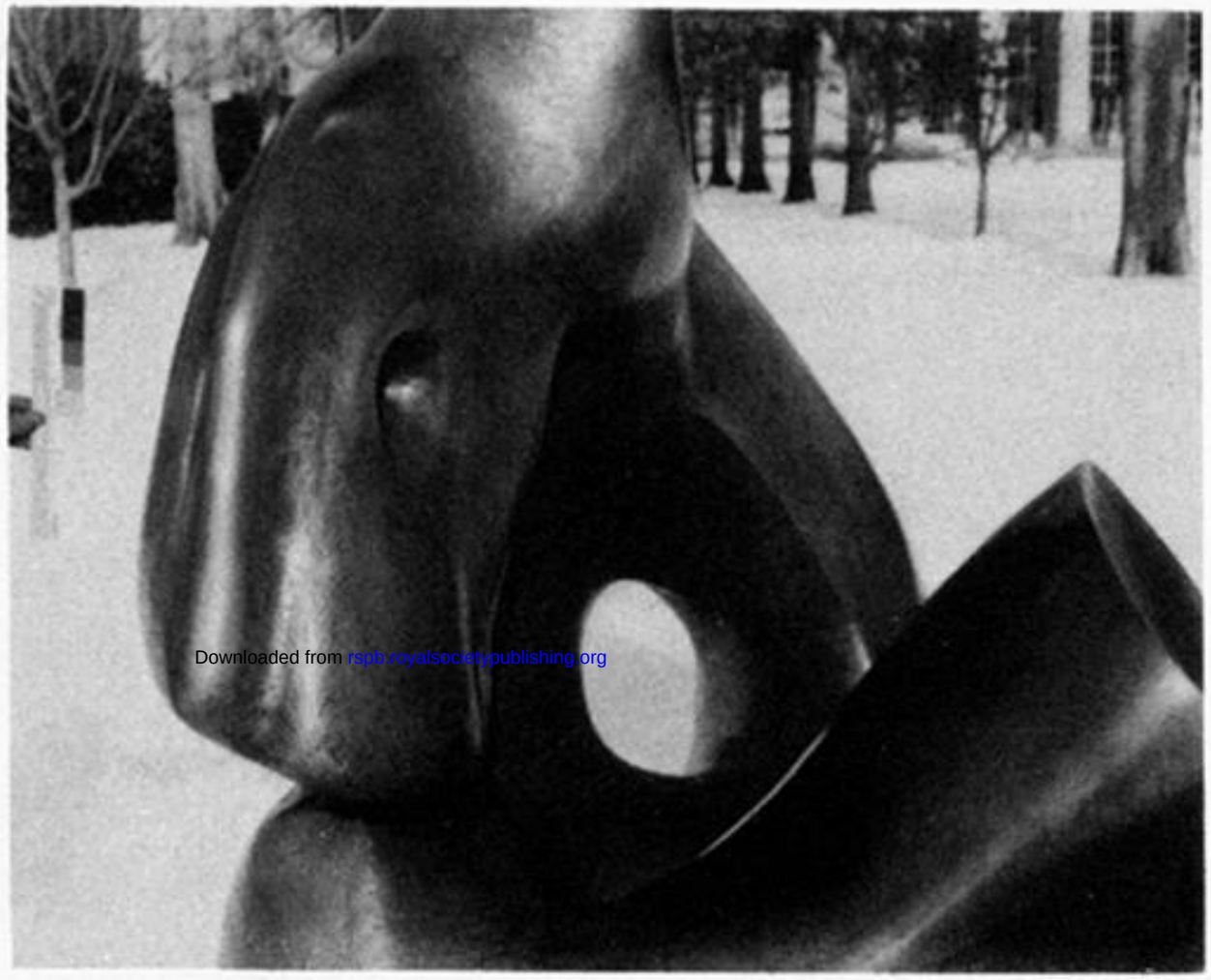
FIGURE 4. Examples of zero-crossing detection by means of $\nabla^2 G$. Row (*a*) shows three images and row (*b*) shows their convolutions with the $\nabla^2 G$ filter of figure 2 ($w = 2\sigma = 8$), zero being represented by an intermediate grey. In row (*c*), positive values are shown white, and negative, black; and in row (*d*) only the zero-crossings appear.

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



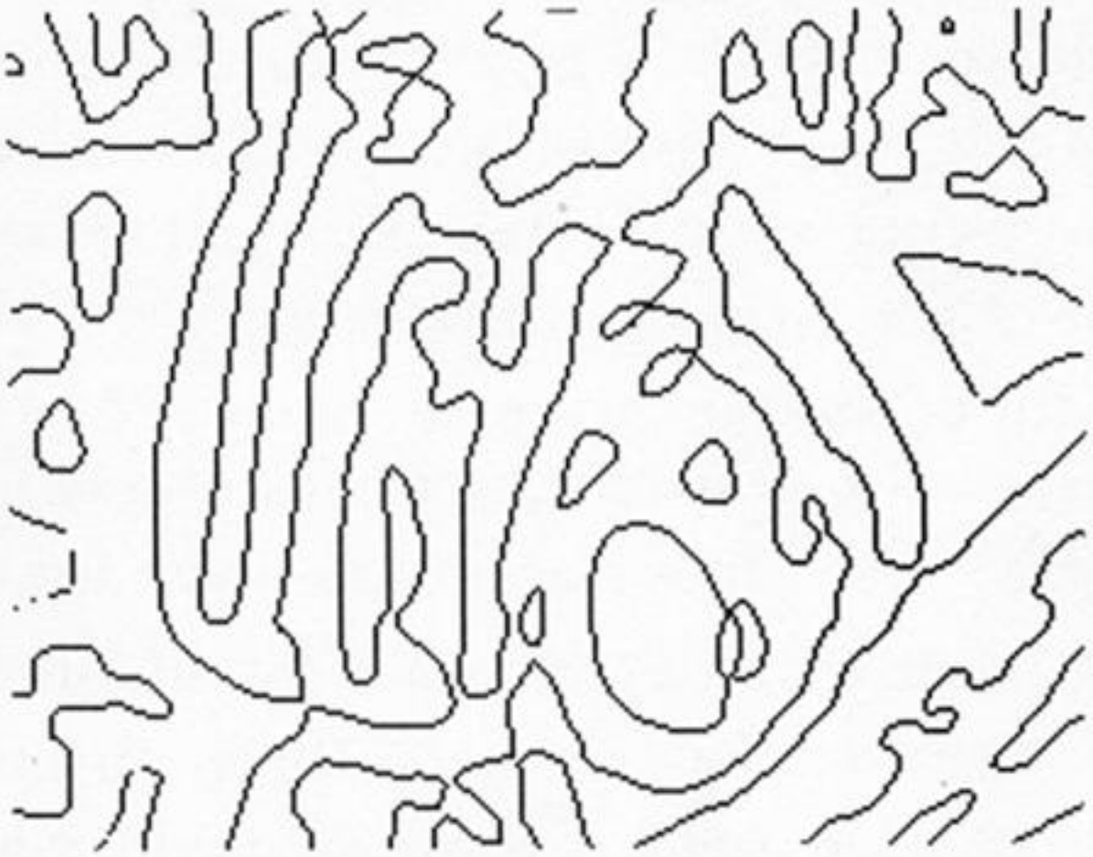
PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

FIGURE 5. Comparison of the performance of $\nabla^2 G$ with that of similar filters. Column (*a*) shows an image, its convolution with $\nabla^2 G$ and the resulting signed zero-crossings. Column (*b*) contains the same sequence, but for the pure one-octave bandpass filter shown, with its Fourier transform, at the top of the column. The zero-crossing array contains echoes of the strong edges in the image. Columns (*c*) and (*d*) exhibit the same analysis of another image, except that here, $\nabla^2 G$ is compared with a square-wave approximation to the second derivative. The widths of the central excitatory regions of the filters are the same for each comparison pair, being 12 for (*a*) and (*b*), and 18 for (*c*) and (*d*). The square-wave filter sees relatively few zero-crossings.



(a)

(b)



(d)

FIGURE 6. The image (*a*) has been convolved with $\nabla^2 G$ having $w = 2\sigma = 6$, 12 and 24 pixels. These filters span approximately the range of filters that operate in the human fovea. In (*b*), (*c*) and (*d*) are shown the zero-crossings thus obtained. Notice the fine detail picked up by the smallest. This set of figures neatly poses our next problem: how does one combine all this information into a single description?