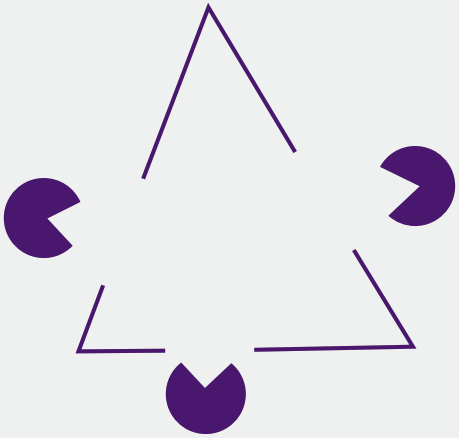
VISION



David Marr

FOREWORD BY
Shimon Ullman

AFTERWORD BY
Tomaso Poggio

INTEL-1028
9,641,841

VISION

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr

The MIT Press
Cambridge, Massachusetts
London, England

© 2010 Lucia M. Vaina

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was originally published in 1982 by W. H. Freeman and Company.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Garamond by Newgen. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Marr, David, 1945–1980
Vision : a computational investigation into the human representation and processing of visual information / David Marr.
p. cm.
Originally published: San Francisco : W. H. Freeman, c1982.
Inludes bibliographical references and index.
ISBN 978-0-262-51462-0 (pbk. : alk. paper) 1. Vision—Data processing. 2. Vision—Mathematical models. 3. Human information processing. I. Title.
QP475.M27 2010
612.8′4—dc22
2009048460

10 9 8 7 6 5 4 3 2 1

Detailed Contents

Foreword by Shimon Ullman xvii

Preface xxiii

PART I
INTRODUCTION AND PHILOSOPHICAL PRELIMINARIES

GENERAL INTRODUCTION 3

Chapter 1

The Philosophy and the Approach 8

Background 8

Understanding Complex Information-Processing Systems 19
- Representation and description 20
- Process 22
- The three levels 24
- Importance of computational theory 27
- The approach of J. J. Gibson 29

A Representational Framework for Vision 31
- The purpose of vision 32
- Advanced vision 34
- To the desirable via the possible 36

PART II
VISION

Chapter 2

Representing the Image 41

Physical Background of Early Vision 41
Representing the image 44
Underlying physical assumptions 44
Existence of surfaces 44
Hierarchical organization 44
Similarity 47
Spatial continuity 49
Continuity of discontinuities 49
Continuity of flow 50
General nature of the representation 51

Zero-Crossings and the Raw Primal Sketch 54
Zero-Crossings 54
Biological implications 61
The psychophysics of early vision 61
The physiological realization of the $\nabla^2 G$ filters 64
The physiological detection of zero crossings 64
The first complete symbolic representation of the image 67
The raw primal sketch 68
Philosophical aside 75

Spatial Arrangement of an Image 79

Light Sources and Transparency 86
Other light source effects 88
Transparency 89
Conclusions 90

Grouping Processes and the Full Primal Sketch 91
Main points in the argument 96
The computational approach and the psychophysics of texture discrimination 96

Chapter 3

From Images to Surfaces 99

Modular Organization of the Human Visual Processor 99
Processes, Constraints, and the Available Representations of an Image 103
Stereopsis 111
Measuring stereo disparity 111
Computational theory 111

Algorithms for stereo matching 116
A cooperative algorithm 116
Cooperative algorithms and the stereo matching problem 122
Biological evidence 125
A second algorithm 127
Uniqueness, cooperativity, and the pulling effect 140
Panum's fusional area 144
Impressions of depth from larger disparities 144
Have we solved the right problem? 148
Vergence movements and the 2½-D sketch 149
Neural implementation of stereo fusion 152
Computing distance and surface orientation from disparity 155
Computational theory 155
Distance from the viewer to the surface 155
Surface orientation from disparity change 156
Algorithm and implementation 159

Directional Selectivity 159
Introduction to visual motion 159
Computational theory 165
An algorithm 167
Neural implementation 169
Using directional selectivity to separate independently moving surfaces 175
Computational theory 175
Algorithm and implementation 177
Looming 182

Apparent Motion 182
Why apparent motion? 183
The two halves of the problem 184
The correspondence problem 188
Empirical findings 188
What is the input representation? 188
Two dimensionality of the correspondence process 193
Ullman's theory of the correspondence process 196
A critique of Ullman's theory 199
A new look at the correspondence problem 202
One problem or two? 202
Separate systems for structure and object constancy 204
Structure from Motion 205
The problem 205
A previous approach 207
The rigidity constraint 209
The rigidity assumption 210
A note about the perspective projection 211

Optical flow 212
The input representation 212
Mathematical results 213

Shape Contours 215
Some examples 216
Occluding contours 218
Constraining assumptions 219
Implications of the assumptions 222
Surface orientation discontinuities 225
Surface contours 226
The puzzle and difficulty of surface contours 228
Determining the shape of the contour generator 229
The effects of more than one contour 230

Surface Texture 233
The isolation of texture elements 234
Surface parameters 234
Possible measurements 234
Estimating scaled distance directly 238
Summary 239

Shading and Photometric Stereo 239
Gradient space 240
Surface illumination, surface reflectance, and image intensity 243
The reflectance map 245
Recovery of shape from shading 248
Photometric stereo 249

Brightness, Lightness, and Color 250
The Helson–Judd approach 252
Retinex theory of lightness and color 253
Algorithms 255
Extension to color vision 256
Comments on the retinex theory 257
Some physical reasons for the importance of simultaneous contrast 259
Hypothesis of the superficial origin of nonlinear changes in intensity 26
Implications for measurements on a trichromatic image 262
Summary of the approach 264
Summary 264

Chapter 4

The Immediate Representation of Visible Surfaces 268

Introduction 268

Image Segmentation 270

Reformulating the Problem 272
The Information to be Represented 275
General Form of the 2½-D Sketch 277
Possible Forms for the Representation 279
Possible Coordinate Systems 283
Interpolation, Continuation, and Discontinuities 285
Computational Aspects of the Interpolation Problem 288
  Discontinuities 289
  Interpolation methods 290
Other Internal Computations 291

Chapter 5

Representing Shapes for Recognition 295

Introduction 295
Issues Raised by the Representation of Shape 296
  Criteria for judging the effectiveness of a shape representation 296
    Accessibility 297
    Scope and uniqueness 297
    Stability and sensitivity 298
  Choices in the design of a shape representation 298
    Coordinate systems 298
    Primitives 300
    Organization 302
The 3-D Model Representation 302
  Natural coordinate systems 303
  Axis-based descriptions 304
  Modular organization of the 3-D model representation 305
  Coordinate system of the 3-D model 307
Natural Extensions 309
Deriving and Using the 3-D Model Representation 313
  Deriving a 3-D model description 313
  Relating viewer-centered to object-centered coordinates 317
  Indexing and the catalogue of 3-D models 318
  Interaction between derivation and recognition 321
    Finding the correspondence between image and catalogued model 322
    Constraint analysis 322
Psychological Considerations 325

Chapter 6

Synopsis 329

PART III
EPILOGUE

Chapter 7

In Defense of the Approach 335

Introduction 335

A Conversation 336

Afterword by Tomaso Poggio 362

Glossary 368

Bibliography 375

Index 393

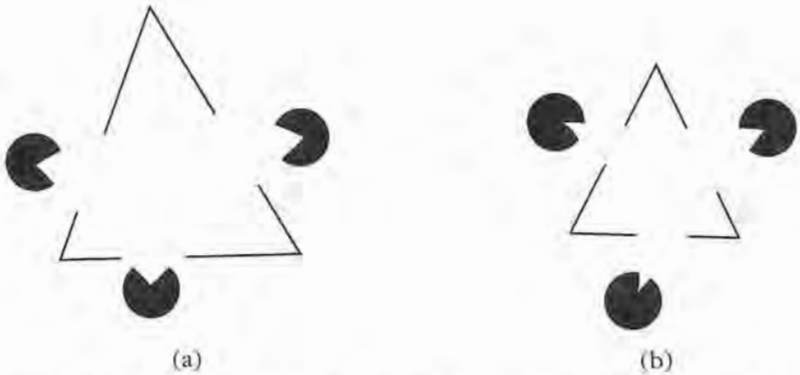


Figure 2–6. Subjective contours. The visual system apparently regards changes in depth as so important that they must be made explicit everywhere, including places where there is no direct visual evidence for them.

movements in the images of that object. If the object is rigid, the motions of the images of nearby portions of the object's surface are similar. Hence, the motions of portions of the object that are close to one another in the image are usually similar. In particular, the velocity field of motion in the image varies continuously almost everywhere, and if it is ever discontinuous at more than an isolated point, then a failure of rigidity (like an object boundary) is present in the outside world. In particular, *if direction of motion is ever discontinuous at more than one point—along a line, for example,—then an object boundary is present.*

General Nature of the Representation

The important message of these physical constraints is that although the basic elements in our image are the intensity changes, the physical world imposes on these raw intensity changes a wide variety of spatial organizations, roughly independently at different scales. This organization is reflected in the structure of images, and since it yields important clues about the structure of the visible surfaces, it needs to be captured by the early representations of the image. Specifically, I propose doing this by a set of "place tokens" that roughly correspond to oriented *edge* or *boundary* segments or to points of *discontinuity* in their orientations, to *bars* (roughly parallel edge pairs) or to their *terminations*; or to *blobs*—roughly, doubly terminated bars. These primitives can be defined in very concrete ways—from pure discontinuities in intensity—or in rather abstract ways. A blob

can be defined from a cloud of dots, for example, or a boundary from certain (but not all) kinds of texture change or from the lining up of a set of tokens that are themselves defined in quite complex ways, as in the example of Figure 2–4.

A rough illustration of the general idea appears in Figure 2–7; this representational scheme is called the *primal sketch* (Marr, 1976). The critical ideas behind it are the following:

1. The primal sketch consists of primitives of the same general kind at different scales—a blob has a rough position, length, width, and orientation at whatever scale it is defined—but the primitives can be defined from an image in a variety of ways, from the very concrete (a black ink mark) to the very abstract (a cloud of dots).

2. These primitives are built up in stages in a constructive way, first by analyzing and representing the intensity changes and forming tokens directly from them, then by adding representations of the local geometrical structure of their arrangement, and then by operating on these things with active selection and grouping processes to form larger-scale tokens that reflect larger-scale structures in the image, and so forth.

3. On the whole, the primitives that are obtained, the parameters associated with them, and the accuracy with which they are measured are designed to capture and to match the structure in an image so as to facilitate the recovery of information about the underlying geometry of the visible surfaces. This gives rise to a complex trade-off between the accuracy of the discriminations that can be made and the value of making them. For example, projected orientations in the image do change if the surface orientation changes, but on the whole by only a rather small amount and probably usually less than the typical variation in orientation to be found in the objective distribution of markings on a surface. This means that except in special situations, it is not worth having a very powerful apparatus for making subtle orientation discriminations. On the other hand, because only a very small relative movement is compelling evidence that two surfaces are separate, it is worth being very sensitive to relative movement.

The three main stages in the processes that derive the primal sketch are (1) the detection of zero-crossings (Marr and Poggio, 1979; Marr, Poggio, and Ullman, 1979; Marr and Hildreth, 1980); (2) the formation of the raw primal sketch (Marr, 1976; Marr and Hildreth, 1980; Hildreth 1980); and (3) the creation of the full primal sketch (Marr, 1976).

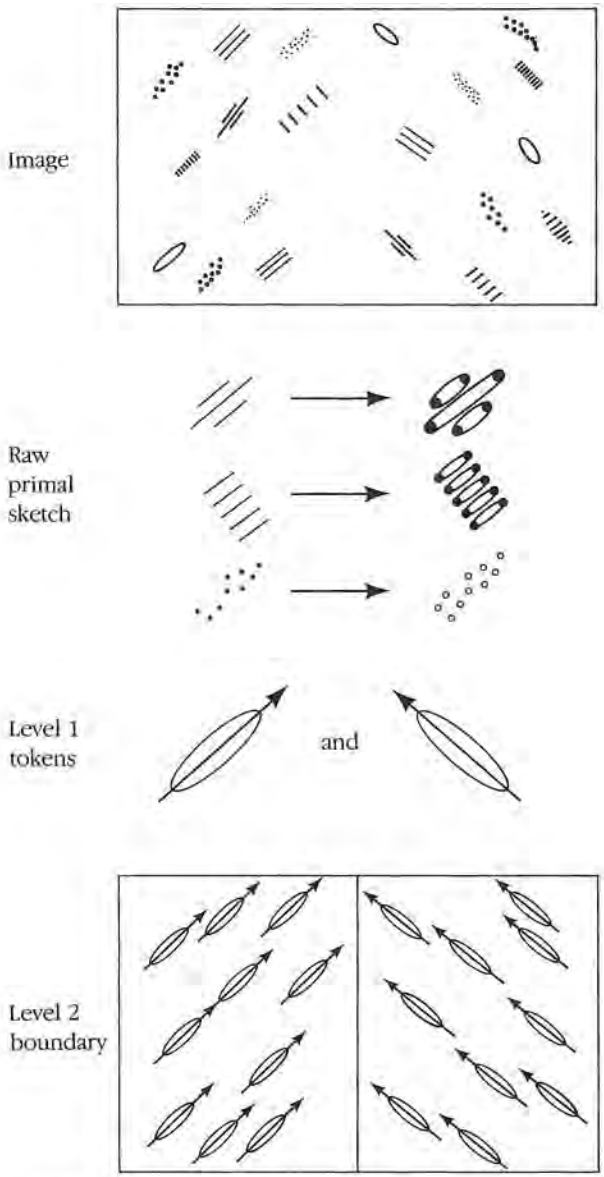


Figure 2–7. A diagrammatic representation of the descriptions of an image at different scales which together constitute the primal sketch. At the lowest level, the raw primal sketch faithfully follows the intensity changes and also represents terminations, denoted here by filled circles. At the next level, oriented tokens are formed for the groups in the image. At the next level, the difference in orientations of the groups in the two halves of the image causes a boundary to be constructed between them. The complexity of the primal sketch depends upon the degree to which the image is organized at the different scales.

2.2 ZERO-CROSSINGS AND THE RAW PRIMAL SKETCH

Zero-Crossings

The first of the three stages described above concerns the detection of intensity changes. The two ideas underlying their detection are (1) that intensity changes occur at different scales in an image, and so their optimal detection requires the use of operators of different sizes; and (2) that a sudden intensity change will give rise to a peak or trough in the first derivative or, equivalently, to a *zero-crossing* in the second derivative, as illustrated in Figure 2–8. (A zero-crossing is a place where the value of a function passes from positive to negative).

These ideas suggest that in order to detect intensity changes efficiently, one should search for a filter that has two salient characteristics. First and foremost, it should be a differential operator, taking either a first or second spatial derivative of the image. Second, it should be capable of being tuned to act at any desired scale, so that large filters can be used to detect blurry shadow edges, and small ones to detect sharply focused fine detail in the image.

Marr and Hildreth (1980) argued that the most satisfactory operator fulfilling these conditions is the filter $\nabla^2 G$, where ∇^2 is the Laplacian operator $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$ and G stands for the two-dimensional Gaussian distribution

$$G(x,y) = e^{-\frac{x^2+y^2}{2\pi\sigma^2}}$$

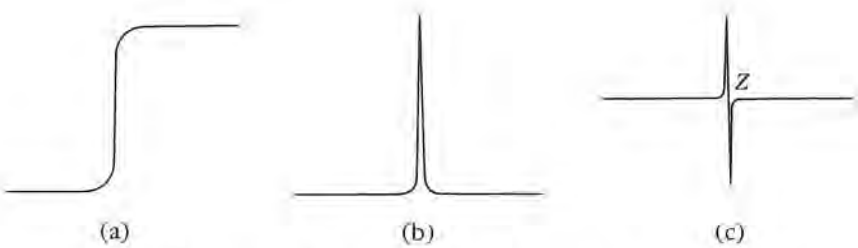


Figure 2–8. The notion of a zero-crossing. The intensity change (a) gives rise to a peak (b) in its first derivative and to a (steep) zero-crossing Z (c) in its second derivative.

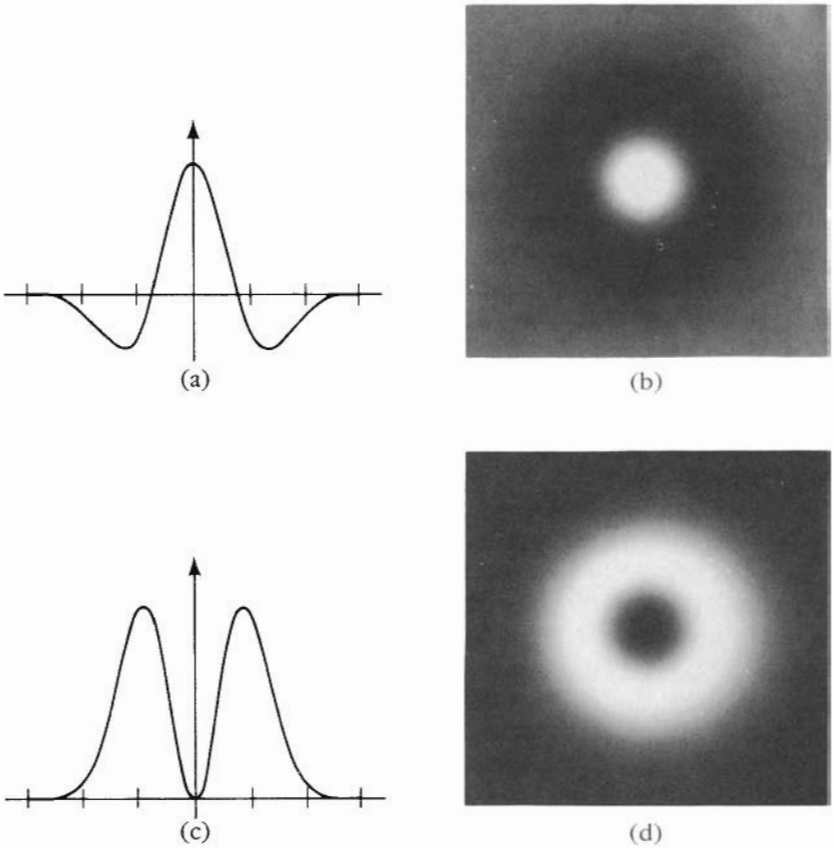
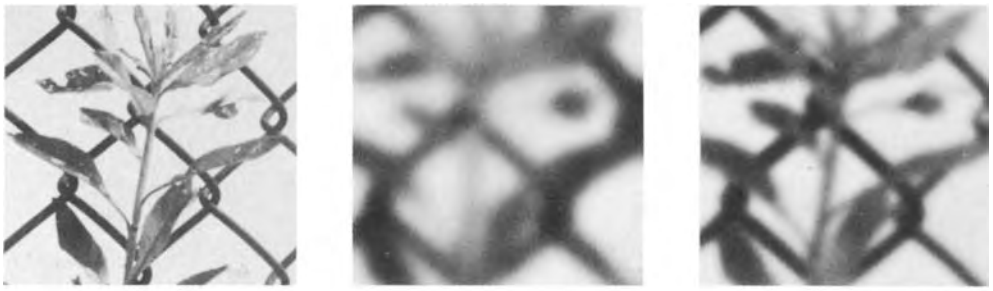


Figure 2–9. $\nabla^2 G$ is shown as a one-dimensional function (a) and in two-dimensions (b) using intensity to indicate the value of the function at each point. (c) and (d) show the Fourier transforms for the one- and two-dimensional cases respectively. (Reprinted by permission from D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R Soc. Lond. B 204,* pp. 301–328.)

which has standard deviation σ. $\nabla^2 G$ is a circularly symmetric Mexican-hat-shaped operator whose distribution in two dimensions may be expressed in terms of the radial distance r from the origin by the formula

$$\nabla^2 G(r) = \frac{-1}{\pi\sigma^4}\left(1 - \frac{r^2}{2\sigma^2}\right) e^{\frac{-r^2}{2\sigma^2}}$$

Figure 2–9 illustrates the one- and two-dimensional forms of this operator, as well as their Fourier transforms.



(a) (b) (c)

Figure 2–10. Blurring images is the first step in detecting intensity changes in them. (a) In the original image, 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 an image that has been blurred with a Gaussian filter, because there is, in effect, an upper limit to the rate at which changes can take place. The first part of the edge detection process can be thought of as decomposing the original image into a set of copies, each filtered with a different-sized Gaussian, and then detecting the intensity changes separately in each. (b) The image filtered with a Gaussian having $\sigma = 8$ pixels; in (c), $\sigma = 4$. The image is 320 by 320 elements. (Reprinted by permission from D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B 204,* pp. 301–328.)

There are two basic ideas behind the choice of the filter $\nabla^2 G$. The first is that the Gaussian part of it, G, blurs the image, effectively wiping out all structure at scales much smaller than the space constant σ of the Gaussian. To illustrate this, Figure 2–10 shows an image that has been convolved with two different-sized Gaussians whose space constants σ were 8 pixels (Figure 2–10b) and 4 pixels (Figure 2–10c). The reason why one chooses the Gaussian for this purpose, rather than blurring with a cylindrical pillbox function (for instance), is that the Gaussian distribution has the desirable characteristic of being smooth and localized in both the spatial and frequency domains and, in a strict sense, being the unique distribution that is simultaneously optimally localized in both domains. And the reason, in turn, why this should be a desirable property of our blurring function is that if the blurring is as smooth as possible, both spatially and in the frequency domain, it is least likely to introduce any changes that were not present in the original image.

The second idea concerns the derivative part of the filter, ∇^2. The great advantage of using it is economy of computation. First-order directional derivatives, like $\partial/\partial x$ or $\partial/\partial y$, could be used, in which case one would subsequently have to search for their peaks or troughs at each orientation (as illustrated in Figure 2–8b); or, second-order directional derivatives, like $\partial^2/\partial x^2$ or $\partial^2/\partial y^2$, could be used, in which case intensity changes would

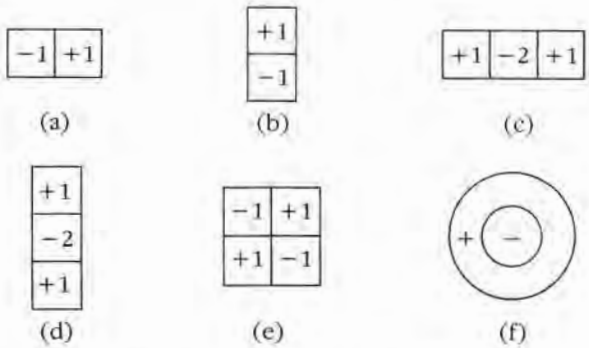
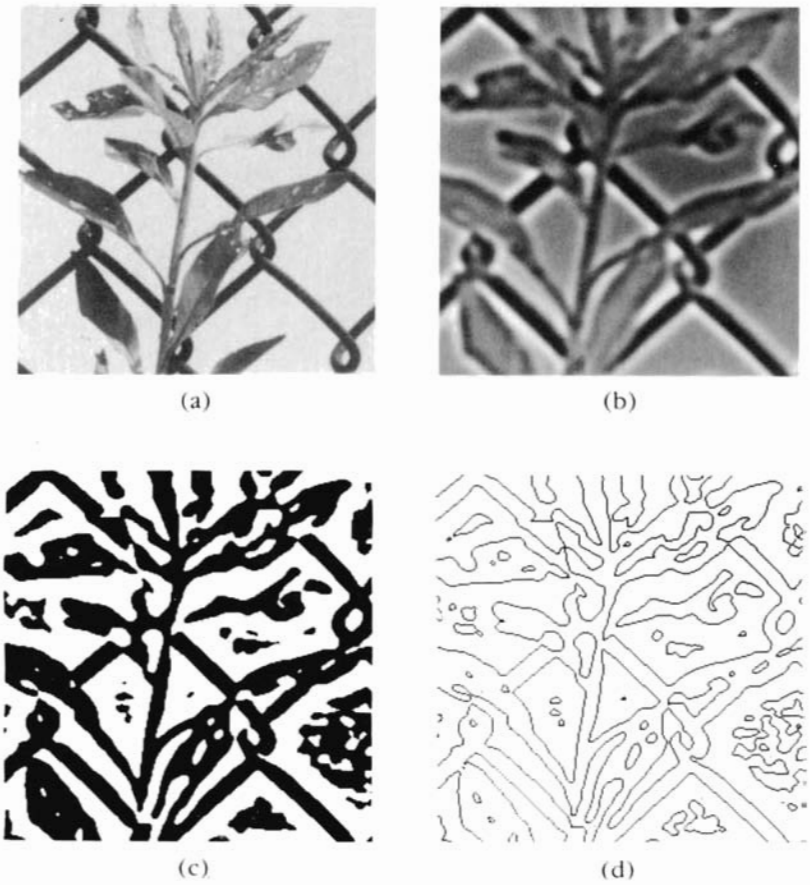


Figure 2–11. The spatial configuration of low-order differential operators. Operators like $\partial/\partial x$ can be roughly realized by filters with the receptive fields illustrated in the figure. (a) $\partial/\partial x$ can be thought of as measuring the difference between the values at two neighboring points along the x-axis. Similarly, (b) shows $\partial/\partial y$. The operator $\partial^2/\partial x^2$ can be thought of as the difference between two neighboring values of $\partial/\partial x$, and so it takes the form shown in (c). The other two second-order operators, $\partial^2/\partial y^2$ and $\partial^2/\partial x \partial y$, appear in (d) and (e), respectively. Finally, the lowest-order isotropic operator, the Laplacian $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$, which we denote by ∇^2, has the circularly symmetric form shown in (f).

correspond to their zero-crossings (see Figure 2–8c). However, the disadvantage of all these operators is that they are directional; they all involve an orientation (see Figure 2–11, which illustrates the spatial organizations, or "receptive fields," in neurophysiological terms of the various first- and second-order differential operators). In order to use the first derivatives, for example, both $\partial I/\partial x$ and $\partial I/\partial y$ have to be measured, and the peaks and troughs in the overall amplitude have to be found. This means that the signed quantity $[(\partial I/\partial x)^2 + (\partial I/\partial y)^2]^{-1/2}$ must also be computed.

Using second-order directional derivative operators involves problems that are even worse than the ones involved in using first-order derivatives. The only way of avoiding these extra computational burdens is to try to choose an orientation-independent operator. The lowest-order isotropic differential operator is the Laplacian ∇^2, and fortunately it so happens that this operator can be used to detect intensity changes provided the blurred image satisfies some quite weak requirements (Marr and Hildreth, 1980).* Images on the whole do satisfy these requirements locally,

*The mathematical notation for blurring an image intensity function $I(x, y)$ with a Gaussian function G is $G * I$ which is read G convolved with I. The Laplacian of this is denoted by $\nabla^2 (G * I)$ and a mathematical identity allows us to move the ∇^2 operator inside the convolution giving $\nabla^2 (G * I) = (\nabla^2 G) * I$.



Figures 2–12, 13, 14. These three figures show examples of zero-crossing detection using $\nabla^2 G$. In each figure, (a) shows the image (320 × 320 pixels); (b) shows the image's convolution with $\nabla^2 G$, with $w_{2-D} = 8$ (zero is represented by gray); (c) shows the positive values in white and the negative in black; (d) shows only the zero-crossings.

so in practice one can use the Laplacian. Hence, in practice, the most satisfactory way of finding the intensity changes at a given scale in an image is first to filter it with the operator $\nabla^2 G$, where the space constant of G is chosen to reflect the scale at which the changes are to be detected, and then to locate the zero-crossings in the filtered image.

Figures 2–12 to 2–14 show what an image looks like when processed in this way. The numerical values in the $\nabla^2 G$-filtered image are both positive

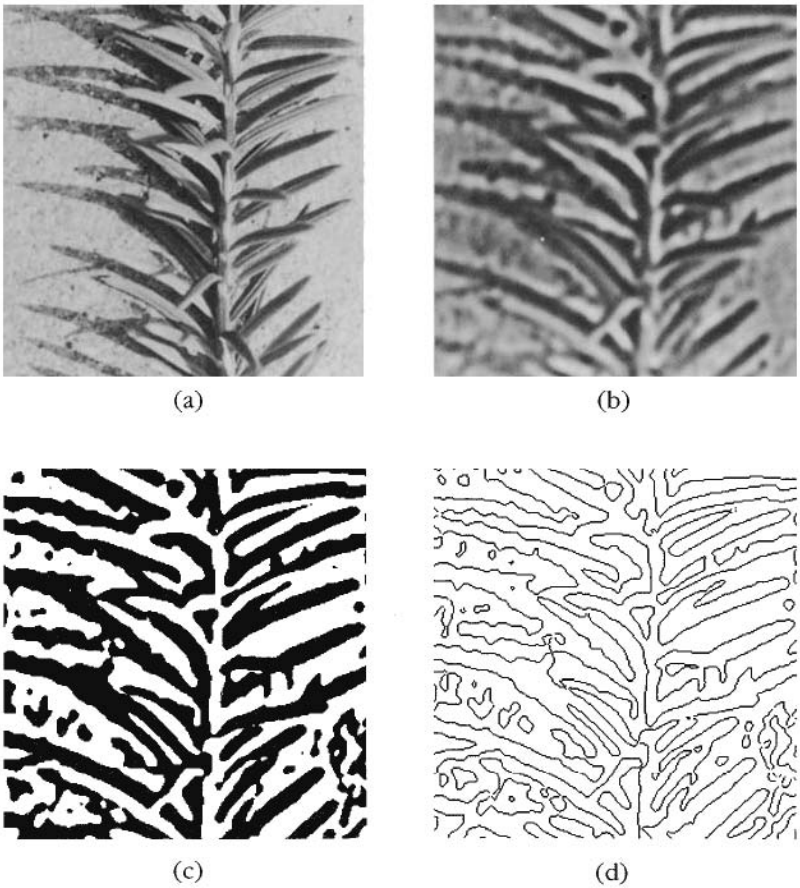


Figure 2–13.

and negative, the overall average being zero. Positive values are represented here by whites, negative by blacks, and the value zero by an intermediate gray. As we have seen, the critical fact about the operator $\nabla^2 G$ is that its zero-crossings mark the intensity changes, as seen at the Gaussian's particular scale. The figures show this well. In Figure 2–12(c), for instance, the filtered image has been "binarized"—that is, positive values were all set to $+1$ and negative values to -1, and in Figure 2–12(d) the zero-crossings alone are shown. The advantage of the binarized representation is that it also shows the sign of the zero-crossing—which side in the image is the darker.

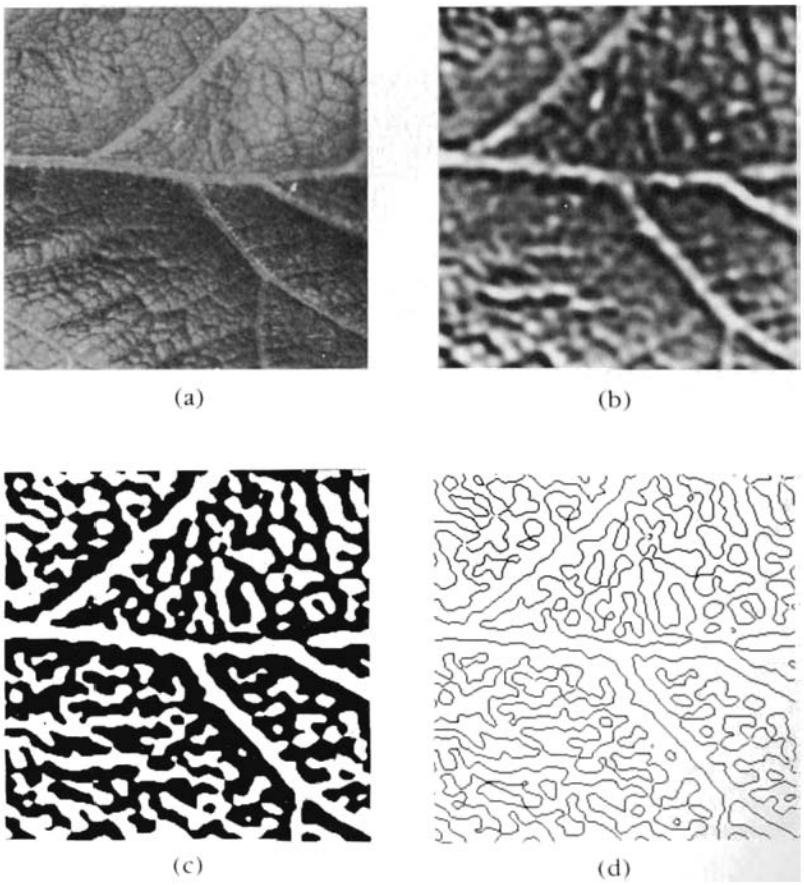
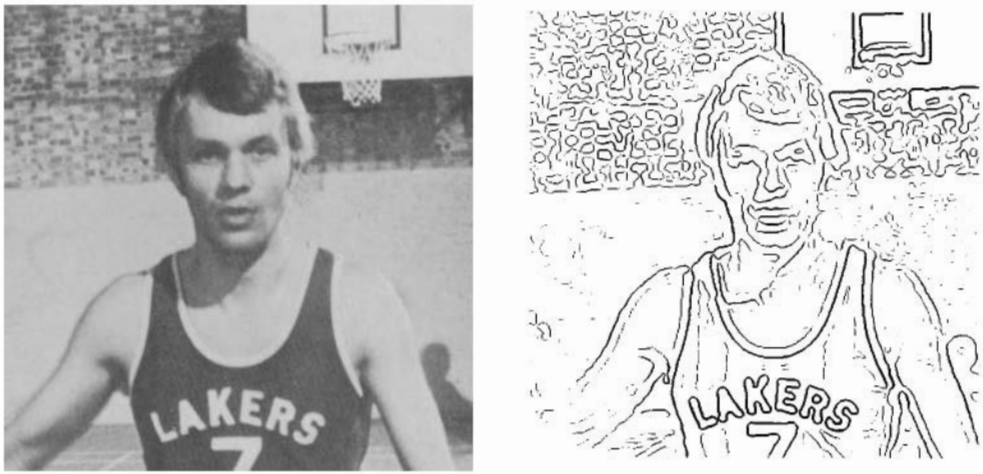


Figure 2–14.

In addition, the slope of the zero-crossing depends on the contrast of the intensity change, though not in a very straightforward way. This is illustrated by Figure 2–15, which shows an original image together with zero-crossings that have been marked with curves of varying intensity. The more contrasty the curve, the greater the slope of the zero-crossing at that point, measured perpendicularly to its local orientation.

Zero-crossings like those of Figures 2–12 to 2–15 can be represented symbolically in various ways. I choose to represent them by a set of oriented primitives called *zero-crossing segments,* each describing a piece of the contour whose intensity slope (rate at which the convolution changes across the segment) and local orientation are roughly uniform. Because of their eventual physical significance, it is also important to make explicit



(a) (b)

Figure 2–15. Another example of zero-crossings; here, the intensity of the lines has been made to vary with the slope of the zero-crossing, so that it is easier to see which lines correspond to the greater contrast. (Courtesy BBC Horizon.)

those places at which the orientation of a zero-crossing changes "discontinuously." The quotation marks are necessary because one can in fact prove that the zero-crossings of $\nabla^2 G * I$ can never change orientation discontinuously, but one can nevertheless construct a practical definition of discontinuity. In addition, small, closed contours are represented as blobs, each also with an associated orientation, average intensity slope, and size defined by its extent along a major and minor axis. Finally, in keeping with the overall plan, several sizes of operator will be needed to cover the range of scales over which intensity changes occur.

Biological Implications

This computational scheme for the very first stages in visual processing leads to an interpretation of many results from the psychophysical and neurophysiological investigations into early vision and to a proposal for the overall strategy behind the design of the first part of the visual pathway.

The psychophysics of early vision

In 1968, Campbell and Robson carried out some adaptation experiments. They found that the sensitivity of subjects to high-contrast gratings was