Handbook of Image & Video Processing

EDITOR AL BOVIK

HANDBOOK OF

IMAGE AND VIDEO PROCESSING

Academic Press Series in Communications, Networking, and Multimedia

EDITOR-IN-CHIEF

Jerry D. Gibson
Southern Methodist University

This series has been established to bring together a variety of publications that represent the latest in cutting-edge research, theory, and applications of modern communication systems. All traditional and modern aspects of communications as well as all methods of computer communications are to be included. The series will include professional handbooks, books on communication methods and standards, and research books for engineers and managers in the world-wide communications industry.

HANDBOOK OF IMAGE AND VIDEO PROCESSING

EDITOR

AL BOVIK

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
THE UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS

ACADEMIC PRESS

A Harcourt Science and Technology Company

SAN DIEGO / SAN FRANCISCO / NEW YORK / BOSTON / LONDON / SYDNEY / TOKYO

This book is printed on acid-free paper. ♾

ACADEMIC PRESS
A Harcourt Science and Technology Company
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA
http://www.academicpress.com

Academic Press
Harcourt Place, 32 Jamestown Road, London, NW1 7BY, UK
http://www.hbuk.co.uk/ap/

Library of Congress Catalog Number: 99-69120

ISBN: 0-12-119790-5

Printed in Canada

00 01 02 03 04 05 FR 9 8 7 6 5 4 3 2 1

Preface

This *Handbook* represents contributions from most of the world's leading educators and active research experts working in the area of *Digital Image and Video Processing.* Such a volume comes at a very appropriate time, since finding and applying improved methods for the acquisition, compression, analysis, and manipulation of visual information in digital format has become a focal point of the ongoing revolution in information, communication and computing. Moreover, with the advent of the world-wide web and digital wireless technology, digital image and video processing will continue to capture a significant share of "high technology" research and development in the future. This *Handbook* is intended to serve as the basic reference point on image and video processing, both for those just entering the field as well as seasoned engineers, computer scientists, and applied scientists that are developing tomorrow's image and video products and services.

The goal of producing a truly comprehensive, in-depth volume on *Digital Image and Video Processing* is a daunting one, since the field is now quite large and multidisciplinary. Textbooks, which are usually intended for a specific classroom audience, either cover only a relatively small portion of the material, or fail to do more than scratch the surface of many topics. Moreover, any textbook must represent the specific point of view of its author, which, in this era of specialization, can be incomplete. The advantage of the current *Handbook* format is that every topic is presented in detail by a distinguished expert who is involved in teaching or researching it on a daily basis.

This volume has the ambitious intention of providing a resource that covers introductory, intermediate and advanced topics with equal clarity. Because of this, the *Handbook* can serve equally well as reference resource and as classroom textbook. As a reference, the *Handbook* offers essentially all of the material that is likely to be needed by most practitioners. Those needing further details will likely need to refer to the academic literature, such as the *IEEE Transactions on Image Processing.* As a textbook, the *Handbook* offers easy-to-read material at different levels of presentation, including several introductory and tutorial chapters and the most basic image processing techniques. The *Handbook* therefore can be used as a basic text in introductory, junior- and senior-level undergraduate, and graduate-level courses in digital image and/or video processing. Moreover, the *Handbook* is ideally suited for short courses taught in industry forums at any or all of these levels. Feel free to contact the Editor of this volume for one such set of computer-based lectures (representing 40 hours of material).

The *Handbook* is divided into ten major sections covering more than 50 Chapters. Following an Introduction, Section 2 of the *Handbook* introduces the reader to the most basic methods of gray-level and binary image processing, and to the essential tools of image Fourier analysis and linear convolution systems. Section 3 covers basic methods for image and video recovery, including enhancement, restoration, and reconstruction. Basic Chapters on Enhancement and Restoration serve the novice. Section 4 deals with the basic modeling and analysis of digital images and video, and includes Chapters on wavelets, color, human visual modeling, segmentation, and edge detection. A valuable Chapter on currently available software resources is given at the end. Sections 5 and 6 deal with the major topics of image and video compression, respectively, including the JPEG and MPEG standards. Sections 7 and 8 discuss the practical aspects of image and video acquisition, sampling, printing, and assessment. Section 9 is devoted to the multimedia topics of image and video databases, storage, retrieval, and networking. And finally, the *Handbook* concludes with eight exciting Chapters dealing with applications. These have been selected for their timely interest, as well as their illustrative power of how image processing and analysis can be effectively applied to problems of significant practical interest.

As Editor and Co-Author of this *Handbook*, I am very happy that it has been selected to lead off a major new series of handbooks on Communications, Networking, and Multimedia to be published by Academic Press. I believe that this is a real testament to the current and growing importance of digital image and video processing. For this opportunity I would like to thank Jerry Gibson, the series Editor, and Joel Claypool, the Executive Editor, for their faith and encouragement along the way.

Last, and far from least, I'd like to thank the many co-authors who have contributed such a fine collection of articles to this *Handbook*. They have been a model of professionalism, timeliness, and responsiveness. Because of this, it was my pleasure to carefully read and comment on every single word of every Chapter, and it has been very enjoyable to see the project unfold. I feel that this *Handbook of Image and Video Processing* will serve as an essential and indispensable resource for many years to come.

Al Bovik
Austin, Texas
1999

Editor

Al Bovik is the General Dynamics Endowed Fellow and Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin, where he is the Associate Director of the Center for Vision and Image Sciences. He has published nearly 300 technical articles in the general area of image and video processing areas and holds two U.S. patents.

Dr. Bovik is a recipient of the IEEE Signal Processing Society Meritorious Service Award (1998), and is a two-time Honorable Mention winner of the international Pattern Recognition Society Award. He is a Fellow of the IEEE, is the Editor-in-Chief of the *IEEE Transactions on Image Processing*, serves on many other boards and panels, and was the Founding General Chairman of the IEEE International Conference on Image Processing, which was first held in Austin, Texas in 1994.

Contributors

Scott T. Acton
Oklahoma State University
Stillwater, Oklahoma

Jake K. Aggarwal
The University of Texas at Austin
Austin, Texas

Jan P. Allebach
Purdue University
West Lafayette, Indiana

Rashid Ansari
University of Illinois at Chicago
Chicago, Illinois

Supavadee Aramvith
University of Washington
Seattle, Washington

Gonzalo Arce
University of Delaware
Newark, Delaware

Barry Barnett
The University of Texas at Austin
Austin, Texas 78759

Keith A. Bartels
Southwest Research Institute
San Antonio, Texas

Jan Biemond
Delft University of Technology
Delft, The Netherlands

Charles G. Boncelet, Jr.
University of Delaware
Newark, Delaware

Charles A. Bouman
Purdue University
West Lafayette, Indiana

Alan C. Bovik
The University of Texas at Austin
Austin, Texas

Kevin W. Bowyer
University of South Florida
Tampa, Florida

Walter Carrara
Nonlinear Dynamics, Inc.
Ann Arbor, Michigan

Rama Chellappa
University of Maryland
College Park, Maryland

Tsuhan Chen
Carnegie Mellon University
Pittsburgh, Pennsylvania

Rolf Clackdoyle
Medical Imaging Research Laboratory
University of Utah

Lawrence K. Cormack
The University of Texas at Austin
Austin, Texas

Edward J. Delp
Purdue University
West Lafayette, Indiana

Mita D. Desai
The University of Texas at San Antonio
San Antonio, Texas

Kenneth R. Diller
The University of Texas at Austin
Austin, Texas

Eric Dubois
University of Ottawa
Ottawa, Ontario, Canada

Adriana Dumitras
University of British Columbia
Vancouver, British Columbia, Canada

Touradj Ebrahimi
EPFL
Lausanne, Switzerland

Berna Erol
University of British Columbia
Vancouver, British Columbia, Canada

Brian L. Evans
The University of Texas at Austin
Austin, Texas

P. Fieguth
University of Waterloo
Ontario, Canada

Nikolas P. Galatsanos
Illinois Institute of Technology
Chicago, Illinois

Joydeep Ghosh
The University of Texas at Austin
Austin, Texas

Ron Goodman
ERIM International, Inc.
Ann Arbor, Michigan

Ulf Grendander
Brown University
Providence, Rhode Island

G. M. Haley
Ameritech
Hoffman Estates, Illinois

Soo-Chul Han
Lucent Technologies
Murray Hill, New Jersey

Joe Havlicek
University of Oklahoma
Norman, Oklahoma

Michael D. Heath
University of South Florida
Tampa, Florida

William E. Higgins
Pennsylvania State University
University Park, Pennsylvania

Shih-Ta Hsiang
Rensselaer Polytechnic Institute
Troy, New York

Thomas S. Huang
University of Illinois at Urbana-Champaign
Urbana, Illinois

Anil Jain
Michigan State University
East Lansing, Michigan

Lina J. Karam
Arizona State University
Tempe, Arizona

William C. Karl
Boston University
Boston, Massachusetts

Aggelos K. Katsaggelos
Northwestern University
Evanston, Illinois

Mohammad A. Khan
Georgia Institute of Technology
Atlanta, Georgia

Janusz Konrad
INRS Télécommunications
Verdun, Quebec, Canada

Faouzi Kossentini
University of British Columbia
Vancouver, British Columbia, Canada

Murat Kunt
Signal Processing Laboratory, EPFL
Lausanne, Switzerland

Reginald L. Lagendijk
Delft University of Technology
Delft, The Netherlands

Sridhar Lakshmanan
University of Michigan – Dearborn
Dearborn, Michigan

Richard M. Leahy
University of Southern California
Los Angeles, California

Wei-Ying Ma
Hewlett-Packard Laboratories
Palo Alto, California

Chhandomay Mandal
The University of Texas at Austin
Austin, Texas

B. S. Manjunath
University of California
Santa Barbara, California

Petros Maragos
National Technical University of Athens
Athens, Greece

Nasir Memon
Polytechnic University
Brooklyn, New York

Fatima A. Merchant
Perceptive Scientific Instruments, Inc.
League City, Texas

Michael I. Miller
Johns Hopkins University
Baltimore, Maryland

Phillip A. Mlsna
Northern Arizona University
Flagstaff, Arizona

Baback Moghaddam
Mitsubishi Electric Research Laboratory (MERL)
Cambridge, Massachusetts

Pierre Moulin
University of Illinois
Urbana, Illinois

John Mullan
University of Delaware
Newark, Delaware

T. Naveen
Tektronix
Beaverton, Oregon

Sharath Pankanti
IBM T. J. Watson Research Center
Yorktown Heights, New York

Thrasyvoulos N. Pappas
Northwestern University
Evanston, Illinois

Jose Luis Paredes
University of Delaware
Newark, Delaware

Alex Pentland
Massachusetts Institute of Technology
Cambridge, Massachusetts

Lucio F. C. Pessoa
Motorola, Inc.
Austin, Texas

Ioannis Pitas
University of Thessaloniki
Thessaloniki, Greece

Kannan Ramchandran
University of California, Berkeley
Berkeley, California

Joseph M. Reinhardt
University of Iowa
Iowa City, Iowa

Jeffrey J. Rodriguez
The University of Arizona
Tucson, Arizona

Peter M. B. van Roosmalen
Delft University of Technology
Delft, The Netherlands

Yong Rui
Microsoft Research
Redmond, Washington

Martha Saenz
Purdue University
West Lafayette, Indiana

Robert J. Safranek
Lucent Technologies
Murray Hill, New Jersey

Paul Salama
Purdue University
West Lafayette, Indiana

Dan Schonfeld
University of Illinois at Chicago
Chicago, Illinois

Timothy J. Schulz
Michigan Technological University
Houghton, Michigan

K. Clint Slatton
The University of Texas at Austin
Austin, Texas

Mark J. T. Smith
Georgia Institute of Technology
Atlanta, Georgia

Michael A. Smith
Carnegie Mellon University
Pittsburgh, Pennsylvania

Shridhar Srinivasan
Sensar Corporation
Princeton, New Jersey

Anuj Srivastava
Florida State University
Tallahassee, Florida

Ming-Ting Sun
University of Washington
Seattle, Washington

A. Murat Tekalp
University of Rochester
Rochester, New York

Daniel Tretter
Hewlett Packard Laboratories
Palo Alto, California

H. Joel Trussell
North Carolina State University
Raleigh, North Carolina

Chun-Jen Tsai
Northwestern University
Evanston, Illinois

Baba C. Vemuri
University of Florida
Gainesville, Florida

George Voyatzis
University of Thessaloniki
Thessaloniki, Greece

D. Wang
Samsung Electronics
San Jose, California

Dong Wei
Drexel University
Philadelphia, Pennsylvania

Miles N. Wernick
Illinois Institute of Technology
Chicago, Illinois

Ping Wah Wong
Gainwise Limited
Cupertino, California

John W. Woods
Rensselaer Polytechnic Institute
Troy, New York

Zixiang Xiong
Texas A&M University
College Station, Texas

Jun Zhang
University of Wisconsin at Milwaukee
Milwaukee, Wisconsin

Huaibin Zhao
The University of Texas at Austin
Austin, Texas

Contents

SECTION VII Image and Video Acquisition

SECTION VIII Image and Video Rendering and Assessment

SECTION IX Image and Video Storage, Retrieval and Communication

SECTION X Applications of Image Processing

4.1

Computational Models of Early Human Vision

Lawrence K. Cormack
The University of Texas at Austin

"The nature of things, hidden in darkness, is revealed only by analogizing. This is achieved in such a way that by means of simpler machines, more easily accessible to the senses, we lay bare the more intricate." *Marcello Malpighi, 1675*

1 Introduction

1.1 Aim and Scope

The author of a short chapter on computational models of human vision is faced with an *embarras de richesse*. One wishes to make a choice between breadth and depth, but even this is virtually impossible within a reasonable space constraint. It is hoped that this chapter will serve as a brief overview for engineers interested in processing done by the early levels of the human visual system. We will focus on the representation of luminance information at three stages: the optics and initial sampling, the representation at the output of the eyeball itself, and the representation at primary visual cortex. With apologies, I have allowed us a very brief foray into the historical roots of the quantitative analysis of vision, which I hope may be of interest to some readers.

1.2 A Brief History

The first known quantitative treatment of image formation in the eyeball by Alhazan predated the Renaissance by four centuries. In 1604, Kepler codified the fundamental laws of physiological optics, including the then-controversial inverted retinal image, which was then verified by direct observation of the image *in situ* by Pater Scheiner in 1619 and later (and more famously) by Rene Descarte. Over the next two centuries there was little advancement in the study of vision and visual perception *per se*, with the exception of Newton's formulation of laws of color mixture, However, Newton's seemingly innocuous suggestion that "the Rays to speak properly are not coloured" [1] anticipated the core feature of modern quantitative models of visual perception: the computation of higher *perceptual* constructs (e.g., color) based upon the activity of peripheral receptors differentially sensitive to a *physical* dimension (e.g., wavelength).[1]

In 1801, Thomas Young proposed that the eye contained but three classes of photoreceptor, each of which responded with a sensitivity that varied over a broad spectral range [2]. This theory, including its extensions by Helmholtz, was arguably the first modern computational theory of visual perception. The Young/Helmholtz theory *explicitly* proposed that the properties of objects in the world are not sampled directly, but that certain properties of light are encoded by the nervous system, and that

[1] Newton was pointing out that colors must arise in the brain, because a given color can arise from many wavelength distributions, and some colors can *only* arise from multiple wavelengths. The purples, for example, and even unique red (red that observers judge as tinged with neither orange nor violet), are colors that cannot be experienced by viewing a monochromatic light.

the resulting neural activity was transformed and combined by the nervous system to result in perception. Moreover, the neural activity was assumed to be quantifiable in nature, and thus the output of the visual system could be precisely predicted by a mathematical model. In the case of color, it could be firmly stated that sensation "may always be represented as simply a function of three variables" [3]. While not a complete theory of color perception, this has been borne out for a wide range of experimental conditions.

Coincident with the migration of trichromatic theory from England to Central Europe, some astronomical data made the same journey, and this resulted in the first *applied* model of visual processing. The data were observations of stellar transit times from the Greenwich Observatory taken in 1796. There was a half-second discrepancy between the observations by Maskelyne (the director) and Kinnebrook (his assistant), and for this Kinnebrook lost his job. The observations caught the notice of Bessel in Prussia at a time when the theory of variability was being given a great deal of attention because of the work of Laplace, Gauss, and others. Unable to believe that such a large, systematic error could be due to sloppy astronomy, Bessel developed a linear model of observers' reaction times to visual stimuli (i.e., stellar transits) relative to one another. These models, which Bessel called "personal equations," could then be used to correct the data for the individual making the observations.

It was no accident that the nineteenth century saw the genesis of models of visual behavior, for it was at that time that several necessary factors came together. First, it was realized that an understanding of the eyeball itself begged rather than yielded an explanation of vision.

Second, the brain had be to viewed as explainable, that is, viewed in a mechanistic fashion. While this was not entirely new to the nineteenth century, the measurement of the conduction velocity of a neural impulse by Helmholtz in 1850 probably did more than any other single experiment to demonstrate that the senses did not give rise to immediate, qualitative (and therefore incalculable) impressions, but rather transformed and conveyed information by means that were ultimately quantifiable.

Third, the stimulus had to be understood to some degree. To make tangible progress in modeling the *early* levels of the visual system, it was necessary to think not in terms of objects and meaningful structures in the environment, but in terms of light, of wavelength, of intensity, and its spatial and temporal derivatives. The enormous progress in optics in the nineteenth century created a climate in which vision could be thought of quantitatively; light was not understood, but its veils of magic were quickly falling away.

Finally, theories of vision would have to constrained and testable in a quantitative manner. Experiments would have to be done in which observers made well-defined responses to well-controlled stimuli in order to establish quantitative input–output relationships for the visual system, which could then in turn be modeled. This approach, called *psychophysics*, was born with the publication of *Elemente der Psychophysik* by Gustav Fechner in 1860.

With the historical backdrop painted, we can now proceed to a selective survey of quantitative treatments of early human visual processing.

1.3 A Short Overview

Figure 1 shows a schematic overview of the major structures of the early visual system and some of the functions they perform. We start with the visual world, which varies with space, time, and wavelength, and which has an amplitude spectrum roughly proportional to $1/f$, where f is the spatial frequency of luminance variation. The first major operations by the visual system are passive: low-pass filtering by the optics and sampling by the receptor mosaic, and both of these operations, and the relationship between them, vary with eccentricity.

The retina of the eyeball filters the image further. The photoreceptors themselves filter along the dimensions of time and wavelength, and the details of the filtering varies with receptor type. The output cells of the retina, the retinal ganglion cells, synapse onto the lateral geniculate nucleus of the thalamus (known as the LGN). We will consider the LGN primarily as a relay station to cortex, and the properties of retinal ganglion cells and LGN cells will be treated as largely interchangeable.

LGN cells come in two major types in primates, magnocellular ("M") and parvocellular ("P"); the terminology was adopted for morphological reasons, but important functional properties distinguish the cell types. To grossly simplify, M cells are tuned to low spatial frequencies and high temporal frequencies, and they are insensitive to wavelength variation. In contrast, P cells are tuned to high spatial frequencies and low temporal frequencies, and they encode wavelength information. These two cell types work independently and in parallel, emphasizing different aspects of the same visual stimuli. In the two-dimensional (2-D) Fourier plane, both are essentially circularly-symmetric bandpass filters.

In the primary visual cortex, several properties emerge. Cells become tuned to orientation; they now inhabit something like a Gaussian blob on the spatial Fourier plane. Cells also become tuned to direction of motion (displacement across time) and binocular disparity (displacement across eyeballs). A new dichotomy also emerges, that between so-called simple and complex cells. Simple cells behave much as wavelet-like linear filters, although they demonstrate some response nonlinearities critical to their function. The complex cells are more difficult to model, as their sensitivity shows no obvious spatial structure.

We will now explore the properties of each of these functional divisions, and their consequences, in turn.

2 The Front End

A scientist in biological vision is likely to refer to anything between the front of the cornea and the area on which he or she is

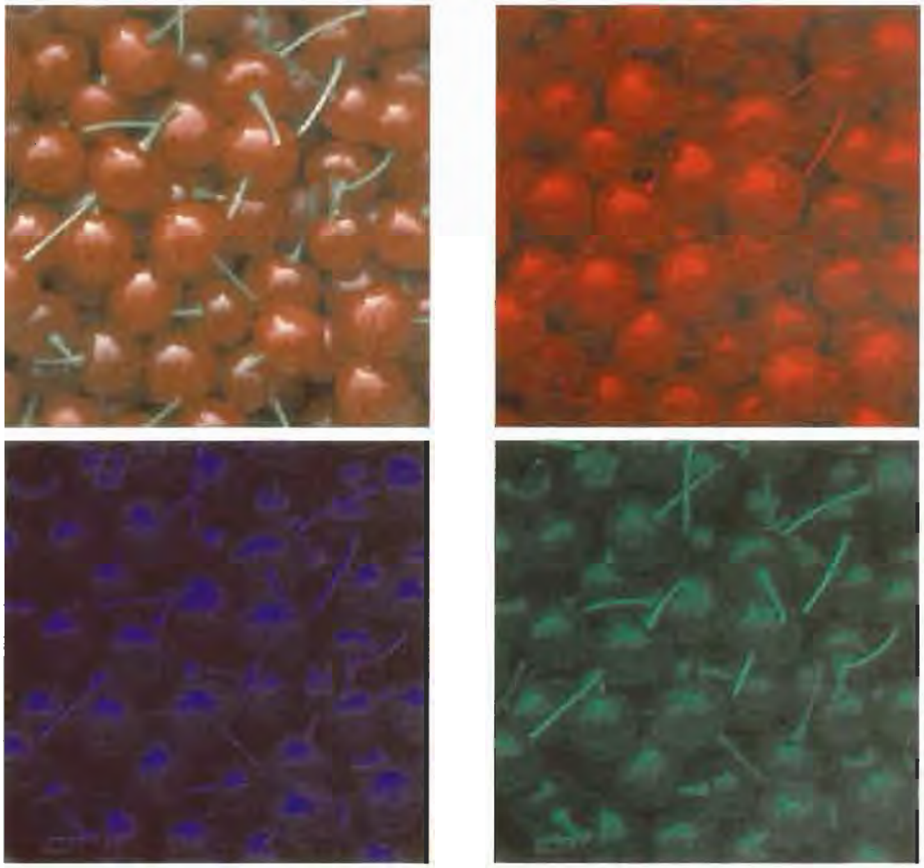


FIGURE 1.13 Color image of "cherries" (top left), and (clockwise) its red, green, and blue components.

FIGURE 3.2.7 Center WM filter applied to each component independently.

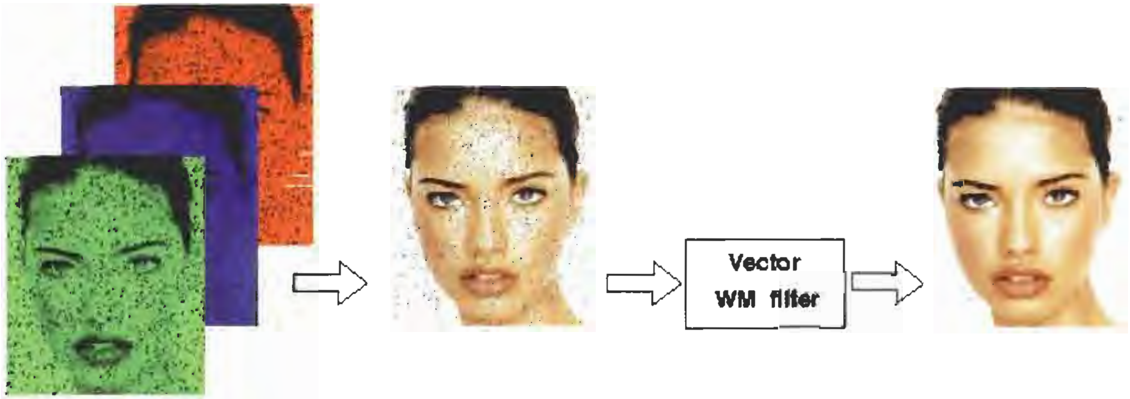


FIGURE 3.2.8 Center vector WM filter applied in the three-dimensional space.

FIGURE 3.2.10 Impulse noise cleaning with a 5×5 CWM smoother: (a) original "portrait" image, (b) image with salt- and-pepper noise, (c) CWM smoother with $W_c = 16$, (d) CWM smoother with $W_c = 5$.

FIGURE 3.2.11 (Enlarged) Noise-free image (left), 5×5 median smoother output (center), and 5×5 mean smoother (right).

FIGURE 3.2.12 (Enlarged) CWM smoother output (left), recursive CWM smoother output (center), and permutation CWM smoother output (right). Window size is 5×5.

FIGURE 3.2.13 (a) Original image, (b) filtered image using a marginal WM filter, (c) filtered image using a vector WM filter.

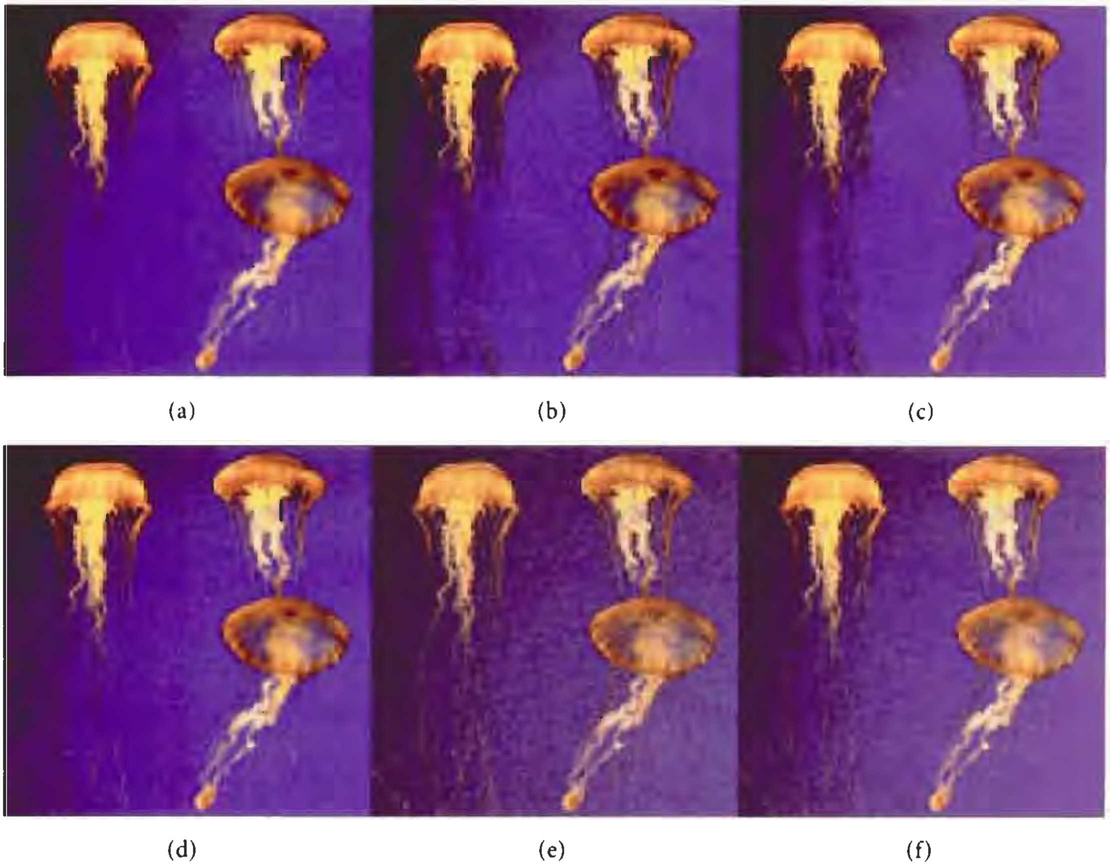


FIGURE 3.2.19 (a) Original image sharpened with (b) the FIR sharpener, and (c) with the WM sharpener. (d) Image with added Gaussian noise sharpened with (e) the FIR sharpener, and (f) the WM sharpener.

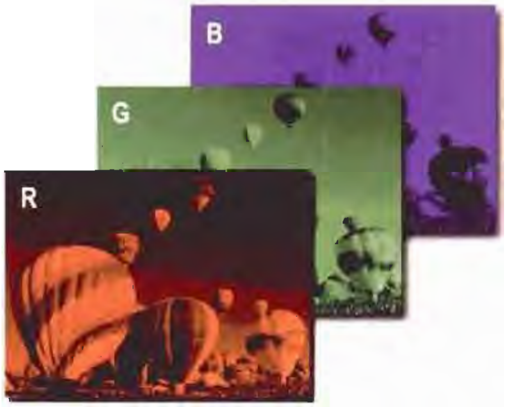


FIGURE 3.7.1 Example of a multichannel image. A color image consists of three color components (channels) that are highly correlated with one another. Similarly, a video image sequence consists of a collection of closely related images.

FIGURE 3.7.2 Example of a multichannel LMMSE restoration: original (upper left), degraded (upper right), restored single-channel statistics obtained from original (middle left), restored single-channel statistics obtained from degraded original (middle right), restored multichannel statistics obtained from original (lower left), restored multichannel statistics obtained from degraded (lower right).

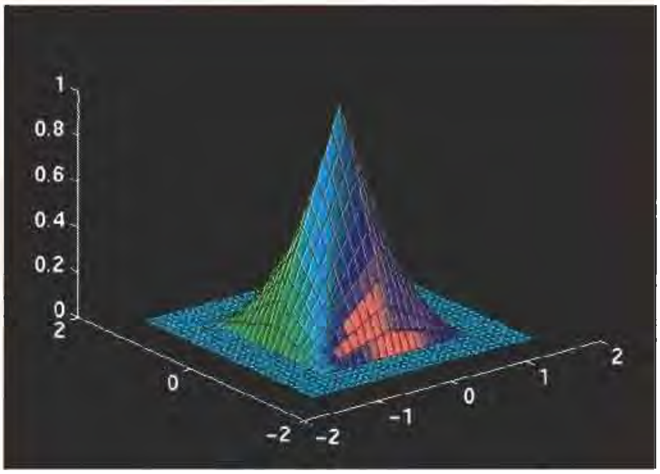


FIGURE 3.12.4 Bilinear basis function.

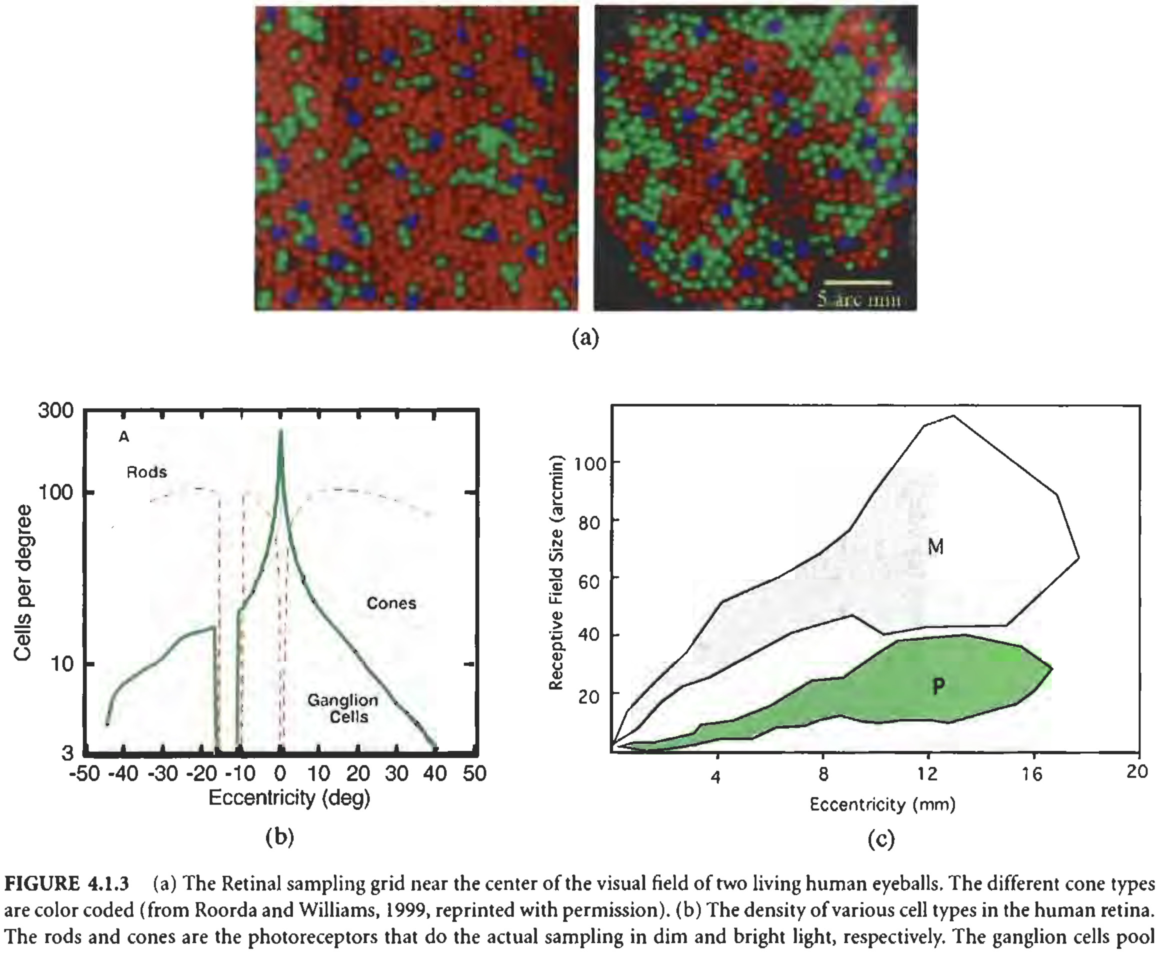


FIGURE 4.1.3 (a) The Retinal sampling grid near the center of the visual field of two living human eyeballs. The different cone types are color coded (from Roorda and Williams, 1999, reprinted with permission). (b) The density of various cell types in the human retina. The rods and cones are the photoreceptors that do the actual sampling in dim and bright light, respectively. The ganglion cells pool the photoreceptor responses and transmit information out of the eyeball (from Geisler and Banks, 1995) reprinted with permission. (c) The dendritic field size (assumed to be roughly equal to the receptive field size) of the two main types of ganglion cell in the human retina (redrawn from Dacy, 1993). The gray shaded region shows the parasol (or M) cells, and the green region shows the midget (or P) cells. The two cell types seem to independently and completely tile the visual world. The functional properties of the two cell types are summarized in Table 1.

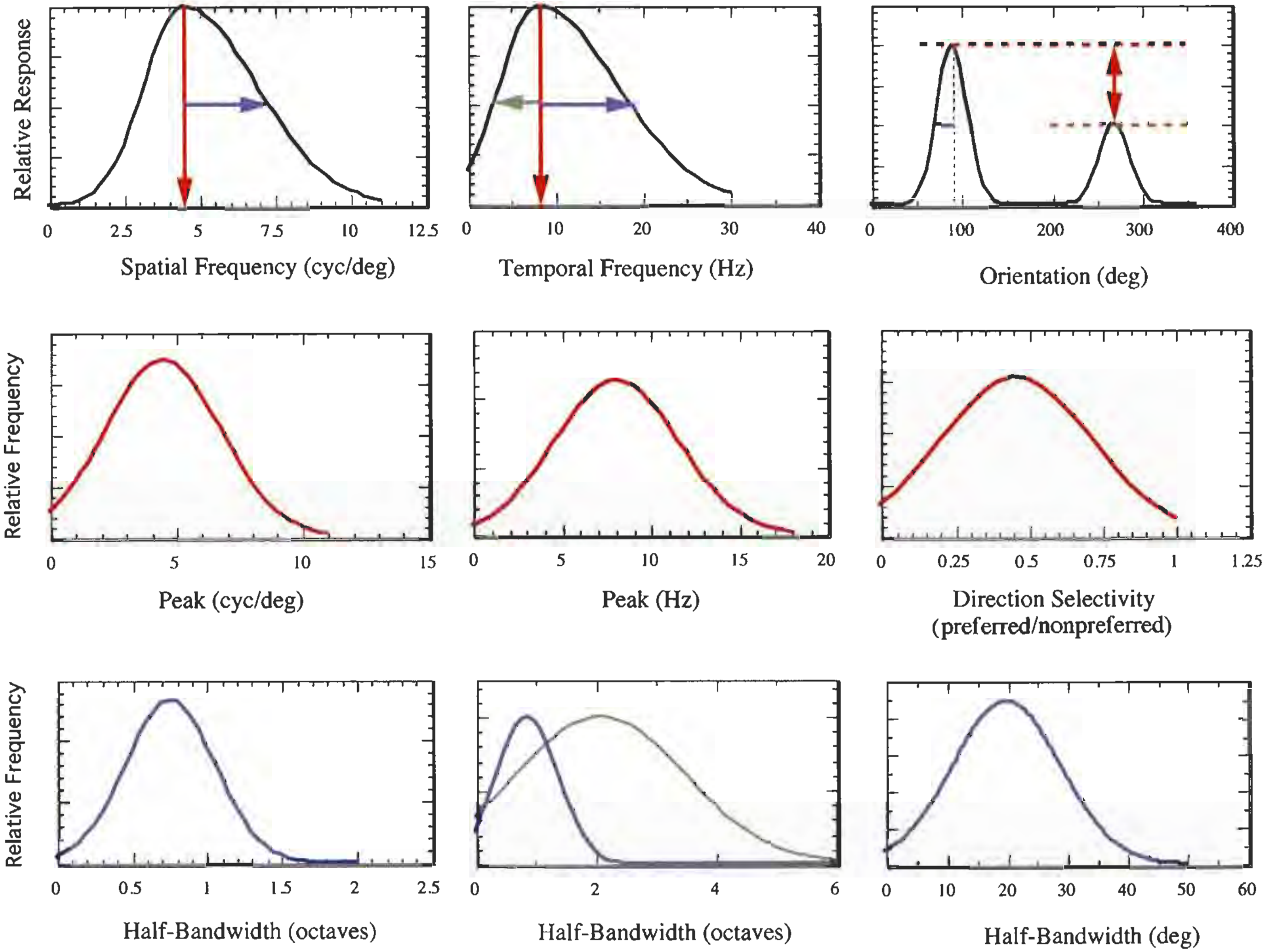


FIGURE 1.13 Left column: the upper panel shows a spatial frequency tuning profile typical of cell such as shown in Fig. 5; the middle and lower panels show distribution estimates of the two parameters of peak sensitivity (middle) and half-bandwidth in octaves (lower) for cells in macaque visual cortex. Middle column: same as the left column, but showing the temporal frequency response. As the response is asymmetric in octave bandwidth, the lower figure shows separate distributions for the upper and lower half-bandwidths (blue and green, respectively). Right column: the upper panel shows the response of a typical cortical cell to the orientation of a drifting sinusoidal grating. The ratio of responses between the optimal direction and its reciprocal is taken as an index of directional selectivity; the estimated distribution of this ratio is plotted in the middle panel (the index cannot exceed unity by definition). The estimate of half-bandwidth for Macaque cortical cells is shown in the lower panel.

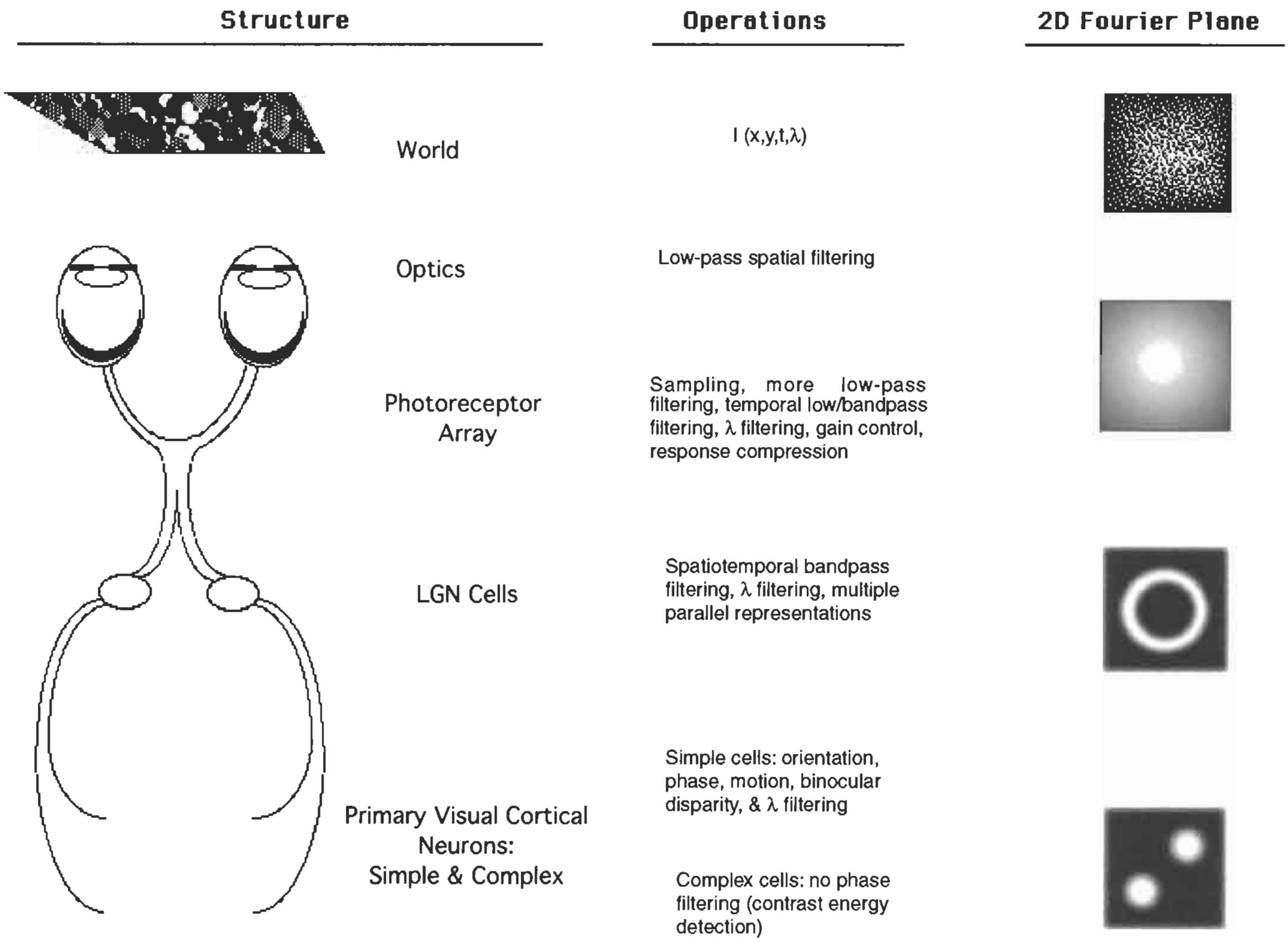


FIGURE 1 Schematic overview of the processing done by the early visual system. On the left, are some of the major structures to be discussed; in the middle, are some of the major operations done at the associated structure; in the right, are the 2-D Fourier representations of the world, retinal image, and sensitivities typical of a ganglion and cortical cell.

working as "the front end." Herein, we use the term to refer to the optics and sampling of the visual system and thus take advantage of the natural division between optical and neural events.

2.1 Optics

The optics of the eyeball are characterized by its 2-D spatial impulse response function, the point-spread function [4]:

$$h(r) = 0.952e^{-2.59|r|^{1.36}} + 0.048e^{-2.43|r|^{1.74}}, \tag{1}$$

in which r is the radial distance in minutes of arc from the center of the image.

This function, plotted in Fig. 2 (or its Fourier transform, the modulation-transfer function), completely characterizes the optics of the eye within the central visual field. The optics deteriorate substantially in the far periphery, so a spatially variant point-spread function is actually required to fully characterize image formation in the human eyeball. For most purposes, however, the point-spread function may be simply convolved with an input image,

$$i(x, y) = I(x, y) * h(x, y), \tag{2}$$

to compute the central retinal image for an arbitrary stimulus, and thus derive the starting point of vision.

2.2 Sampling

While sampling by the retina is a complex spatiotemporal neural event, it is often useful to consider the spatial sampling to be a passive event governed only by the geometry of the receptor grid and the stationary probability of a single receptor absorbing a photon. In the human retina, there are two parallel sampling grids to consider, one comprising the rod photoreceptors and operating in dim light, and the other comprising the

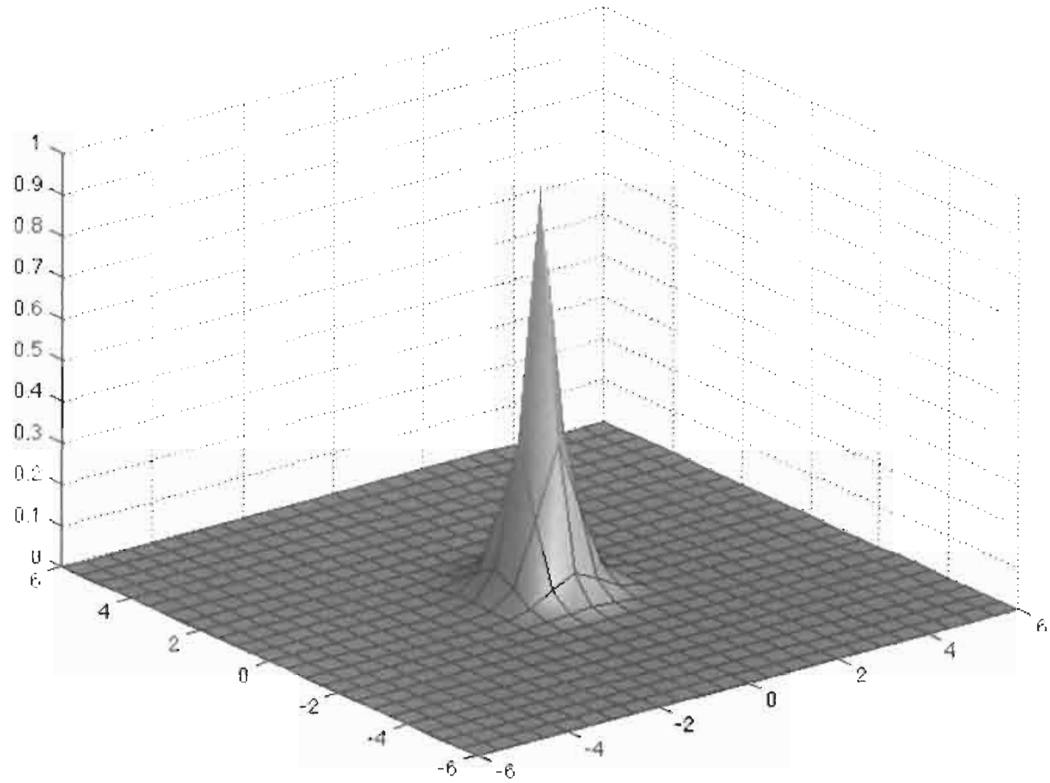


FIGURE 2 Point-spread function of the human eyeball. The x and y axes are in minutes of arc, and the z axis is in arbitrary units. The spacing of the grid lines is equal to the spacing of the photoreceptors in the central visual field of the human eyeball, which is approximately 30 arc sec.

cone photoreceptors (on which we concentrate) and operating in bright light. Shown in Fig. 3(a) are images of the cone sampling grid 1° from the center of the fovea taken in two living, human eyes, using aberration-correcting adaptive optics (similar to those used for correcting atmospheric distortions for terrestrial telescopes) [5]. The short-, medium-, and long-wavelength sensitive cones have been pseudo-colored blue, green, and red, respectively. At the central fovea, the average interreceptor distance is ~2.5 μm, which is ~30 arc sec in the human eyeball. Locally, the lattice is roughly hexagonal, but it is irregular over large areas and seems to become less regular as eccentricity increases. Theoretical performance has been compared in various visual tasks using both actual foveal receptor lattices taken from anatomical studies of the macaque[2] retina and idealized hexagonal lattices of the same receptor diameter, and little difference was found [6].

While the use of a regular hexagonal lattice is convenient for calculations in the space domain, it is often more efficient to work in the frequency domain. In the central retina, one can take the effective sampling frequency to be $\sqrt{3/2}$ times the average interreceptor distance (due to the hexagonal lattice), and then treat the system as sampling with an equivalent 2-D comb (sampling) function. In the peripheral retina, where the optics of the eye pass frequencies above the theoretical sampling limits of the retina, it is possible that the irregular nature of the array helps prevent some of the effects of aliasing. However, visual discriminations in the periphery can be made above the Nyquist frequency by the detection of aliasing [7], so a 2-D comb function of appropriate sampling density can probably suffice for representing the peripheral retina under some conditions.

The photoreceptor density as a function of eccentricity for the rod and cone receptor types in the human eye is shown in Fig. 3(b). The cone lattice is *foveated*, peaking in density at a central location and dropping off rapidly away from this point. Also shown is the variation in the density of retinal ganglion cells that transmit the information out of the eyeball. The ganglion cells effectively sample the photoreceptor array in *receptive fields*, whose size also varies with eccentricity. This variation for the two main types of ganglion cells (which will be discussed below) is shown in Fig. 3(c). The ganglion cell density falls more rapidly than cone density, indicating that ganglion cell receptive fields in the periphery summate over a larger number of receptors, thus sacrificing spatial resolution. This is reflected in measurements of visual acuity as a function of eccentricity, which fall in accord with the ganglion cell data.

The other main factor to consider is the probability of given receptor absorbing a photon, which is governed by the area of

[2]The macaque is an old-world monkey, *macaca fascicularis*, commonly used in vision research because of the great similarity between the macaque and human visual systems.

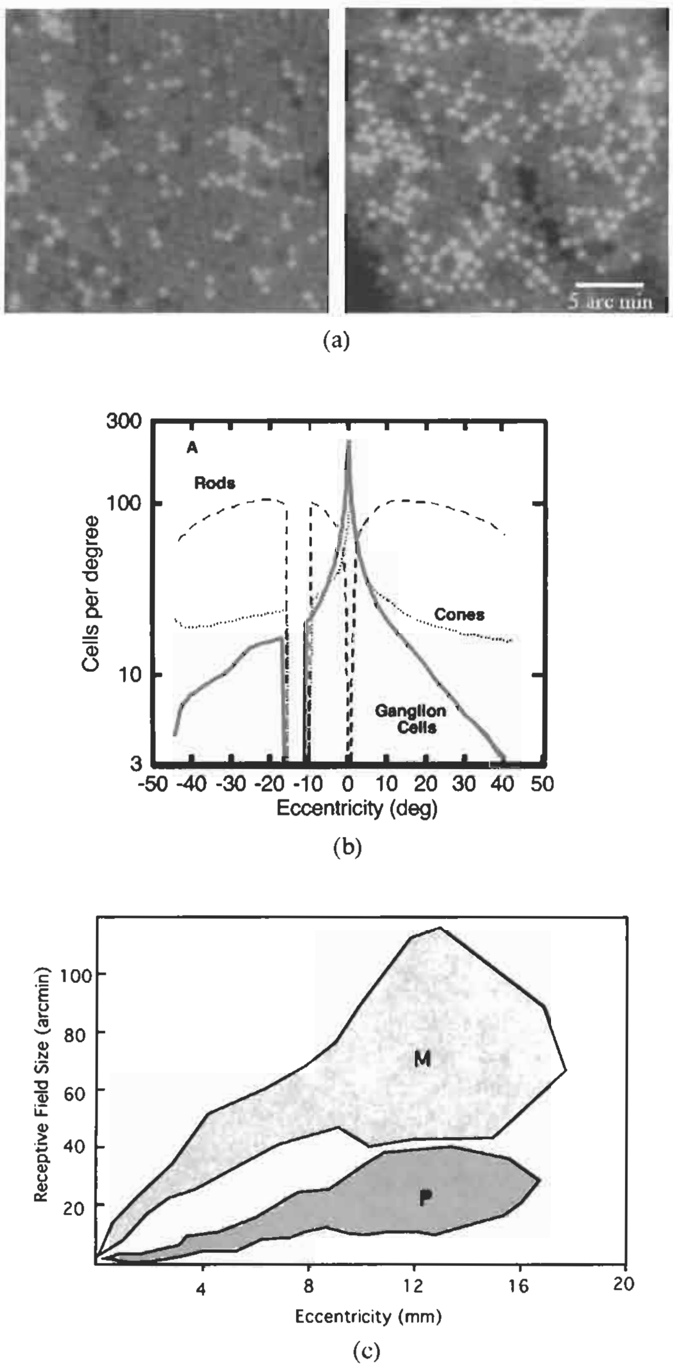


FIGURE 3 (a) The Retinal sampling grid near the center of the visual field of two living human eyeballs. The different cone types are color coded (from Roorda and Williams, 1999, reprinted with permission). (b) The density of various cell types in the human retina. The rods and cones are the photoreceptors that do the actual sampling in dim and bright light, respectively. The ganglion cells pool the photoreceptor responses and transmit information out of the eyeball (from Geisler and Banks, 1995) reprinted with permission. (c) The dendritic field size (assumed to be roughly equal to the receptive field size) of the two main types of ganglion cell in the human retina (redrawn from Dacy, 1993). The gray shaded region shows the parasol (or M) cells, and the green region shows the midget (or P) cells. The two cell types seem to independently and completely tile the visual world. The functional properties of the two cell types are summarized in Table 1. (See color section, p. C–7.)

the effective aperture of the photoreceptor and the probability that a photon entering the aperture will be absorbed. This latter probability is obtained from Beer's Law, which gives the ratio of radiant flux reaching the back of the receptor outer segment to that entering the front [8]:

$$v(\lambda) = 10^{-lc\varepsilon(\lambda)} \tag{3}$$

in which l is the length of the receptor outer segment, c is the concentration of unbleached photopigment, and $\varepsilon(\lambda)$ is the absorption spectrum of the photopigment.

For many modeling tasks, it is most convenient to express the stimulus in terms of $n(\lambda)$, the number of quanta per second as a function of wavelength. This is given by [9]

$$n(\lambda) = 2.24 \times 10^3 A \frac{L(\lambda)}{V(\lambda)} t(\lambda)\lambda, \tag{4}$$

in which A is area of the entrance pupil, $L(\lambda)$ is the spectral luminance distribution of the stimulus, $V(\lambda)$ is the standard spectral sensitivity of human observers, and $t(\lambda)$ is the transmittance of the ocular media. Values of these functions are tabulated in [8].

Thus, for any receptor, the number of absorptions per second, N, is given approximately by

$$N = \int a(1 - v(\lambda))n(\lambda)\, d\lambda \tag{5}$$

in which a is the receptor aperture.

These equations are of fundamental import because they describe the data that the visual system collects about the world. Any comprehensive model of the visual system must ultimately use these data as input. In addition, since these equations specify the information available to the visual system, they allow us to specify how well a particular visual task could be done in principle. This specification is done with a special type of model called an *ideal observer.*

2.3 Ideal Observers

An ideal observer is a mathematical model that performs a given task as well as possible given the information in the stimulus. It is included in this section because it was traditionally used to assess the visual system in terms of quantum efficiency, f, which is the ratio of the number of quanta theoretically required to do a task to the number actually required [e.g., 10]. It is therefore more natural to introduce the topic in terms of optics. However, ideal observers have been used to assess the information loss at various neurophysiological sites in the visual system [6, 11]; the only requirement is that the information present at a given site can be quantitatively expressed.

An ideal observer performs a given task optimally (in the Bayesian sense), and it thus provides an *absolute* theoretical limit on performance in any given task (it thus gives to psychophysics and neuroscience what absolute zero gives to thermodynamics: a fundamental baseline). For example, the smallest offset between a pair of abutting lines (such as on a vernier scale on a pair of calipers) that a human observer can reliably discriminate (75% correct, say) from a stimulus with no offset is almost unbelievably low – a few *seconds* of arc. Recalling from above that foveal cone diameters and receptor spacing are of the order of a half a *minute* of arc, such performance seems rather amazing. But amazing relative to what? The ideal observer gives us the answer by defining what the best possible performance is. In our example, a human observer would be less than 1% efficient as measured at the level of the photoreceptors, meaning that the human observer would require of the order of 10^3 more quanta to achieve the same level of discrimination performance. In this light, human performance ceases to appear quite so amazing, and attention can be directed toward determining how and where the information loss is occurring.

An ideal observer consists of two main parts, a model of the visual system and a Bayesian classifier. The latter is usually expressed as a likelihood ratio:

$$l(s) = \frac{P(s \mid a)}{P(s \mid b)}, \tag{6}$$

in which the numerator and denominator are the conditional probabilities of making observations given that the stimulus was actually a or b, respectively. If the likelihood ratio, or more commonly its logarithm, exceeds a certain amount, stimulus a is judged to have occurred. For a simple discrimination, s would be a vector containing observed quantum catches in a set of photoreceptors, and the probability of this observation given hypotheses a and b would be calculated with the Poisson distribution of light and the factors described above in Sections 2.1 and 2.2.

The beauty of the ideal observer is that it can be used to parse the visual system into layers, and to examine the information loss at each layer. Thus, it becomes a tool by which we can learn which patterns of behavior result from the physics of the stimulus and the structure of the early visual system, and which patterns of behavior result from nonoptimal strategies or algorithms employed by the human visual system. For example, there exists an asymmetry in visual search in which a patch of low-frequency texture in a background of high-frequency texture is much easier to find than when the figure and ground are reversed. It is intuitive to think that if only low-level factors were limiting performance, detecting A on a background of B should be equivalent to detecting B on a background of A (by almost any measure, the contrast of A on B would be equal to that of B on A). However, an ideal-observer analysis proves this intuition false, and an ideal-observer based model of visual search produces the aforementioned search asymmetry [12].

3 Early Filtering and Parallel Pathways

In this section, we discuss the nature of the information that serves as the input to visual cortex. This information is contained in the responses of the retinal ganglion cells (the output of the eyeball) and the LGN.[3] Arguably, this is the last stage that can be comfortably modeled as a strictly data-driven system in which neural responses are independent of activity from other cells in the same or subsequent layers.

3.1 Spatiotemporal Filtering

One difficulty with modeling neural responses in the visual system, particularly for someone new to reading the physiology literature, is that people have an affinity for dichotomies. This is especially evident from a survey of the work on retinogeniculate processing. Neurons have been dichotomized a number of dimensions. In most studies, only one or perhaps two of these dimensions are addressed, which leaves the relationships between the various dimensions somewhat unclear.

With that caveat in mind, the receptive field shown in Fig. 4 is fairly typical of that encountered in retinal ganglion cells or cells of the lateral geniculate nucleus. Figure 4(a) shows the hypothetical cell's sensitivity as a function of spatial position. The receptive field profile shown is a difference of Gaussians, which agrees well with physiological recordings of the majority of ganglion cell receptive field profiles [13, 14], and it is given by

$$DOG(x, y) = a_1 e^{[(x^2-y^2)/(s_1^2)]} - a_2 e^{[(x^2-y^2)/(s_2^2)]}, \tag{7}$$

in which a_1 and a_2 normalize the areas, and s_1 and s_2 are space constants in a ratio of about 1:1.6. Their exact values will vary as a function of eccentricity as per Fig. 3(c).

This representation is fairly typical of that seen in the early work on ganglion cells [e.g., 15], in which the peak response of a neuron to a small stimulus at a given location in the receptive field was recorded, but the location in *time* of this peak response was somewhat indefinite. Thus, a receptive field profile as shown represents a slice in time of the neuron's response some tens of milliseconds after stimulation and, further, the slice of time represented in one spatial location isn't necessarily the same as that represented in another (although for the majority of ganglion cells, the discrepancy would not be too large).

Since the receptive field is spatially symmetric, we can get a more complete picture by looking at a plot of one spatial dimension against time. Such an x–t plot is shown in Fig. 4(b), in which the x dimension is in arcminutes and the t dimension is in milliseconds. The response is space–time separable; the value at any given point is simply the value of the spatial impulse response at that spatial location scaled by the value of the temporal impulse response at that point in time. Thus, the response is given by

$$r(x, t) = DOG(x)[h(t)] \tag{8}$$

[3]Thus we regrettably omit a discussion of the response properties of the photoreceptors *per se* and of the circuitry of the retina. These are fascinating topics — the retina is a marvelous computational structure — and interested readers are referred to [40].

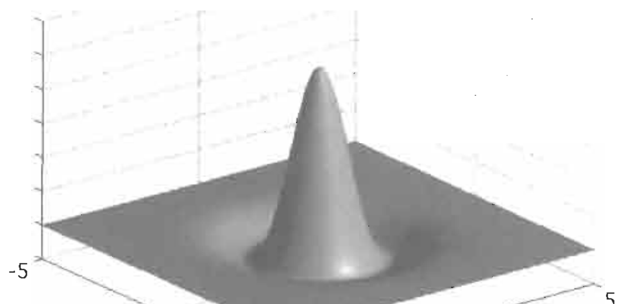


FIGURE 4 (a) Receptive field profile of a retinal ganglion cell modeled as a difference of Gaussians. The x and y axes are in minutes of arc, so this cell would be typical of an M cell near the center of the retina, or a P cell at an eccentricity of 10° to 15° (see Fig. 2). (b) Space–time plot of the same receptive field, illustrating its biphasic temporal impulse response. (The x-axis is in minutes of arc, and the y-axis is in milliseconds).

in which $h(t)$ is a biphasic temporal impulse response function. This response function, $h(t)$, was constructed by subtracting two cascaded low-pass filters of different order [cf. 16]. These low-pass filters are constructed by successive autocorrelation of an impulse response function of the form

$$h(t) = H(t)e^{-t/\tau}, \tag{9}$$

in which $H(t)$ is the Heaviside unit step:

$$H(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}. \tag{10}$$

A succession of n autocorrelations gives

$$h_n(t) = \frac{H(t)(t/\tau)^n e^{-t/\tau}}{\tau n!}, \tag{11}$$

which is a monophasic (low-pass) filter of order n. A difference of two filters of different orders produces the biphasic bandpass response function, and the characteristics of the filter can be adjusted by using component filters of various orders.

The most important implication of this receptive field structure, obvious from the figure, is that the cell is bandpass in both spatial and temporal frequency. As such, the cell discards information about absolute luminance and emphasizes change across space (likely to denote the edge of an object) or change across time (likely to denote the motion of an object). Also obvious from the receptive field structure is that the cell is not selective for orientation (the direction of the spatial change) or the direction of motion.

The cell depicted in the figure is representative in terms of its qualitative characteristics, but the projection from retina to cortex comprises of the order of 10^6 such cells that vary in their specific spatiotemporal tuning properties. Rather than being continuously distributed, however, the cells seem to form functional subgroups that operate on the input image in parallel.

3.2 Early Parallel Representations

The early visual system carries multiple representations of the visual scene. The earliest example of this is at the level of the photoreceptors, where the image can be sampled by rods, cones, or both (at intermediate light levels). An odd aspect of the rod pathway is that it ceases to exist as a separate entity at the output of the retina; there is no such thing as a "rod retinal ganglion cell." This is an interesting example of a need for a separate sensor system for certain conditions combined with a need for neural economy. The pattern analyzing mechanisms in primary visual cortex and beyond are used for both rod and cone signals with (probably) no information about which system is providing the input.

Physiologically, the most obvious example of separate, parallel projections from the retina to the cortex is the presence of the so-called ON and OFF pathways. All photoreceptors have the same sign of response. In the central primate retina, however, each photoreceptor makes direct contact with at least two *bipolar* cells — cells intermediate between the receptors and the ganglion cells — one of which preserves the sign of the photoreceptor response, and the other of which inverts it. Each of these bipolar cells in turn serves as the excitatory center of a ganglion cell receptive field, thus forming two parallel pathways: an ON pathway, which responds to increases in light in the receptive field center, and an OFF pathway, which responds to decreases in light in the receptive field center. Each system forms an independent tiling of the retina, resulting in two complete, parallel neural images being transmitted to the brain.

Another fundamental dichotomy is between midget (or "P" for reasons to become clear in a moment) and parasol (or "M") ganglion cells. Like the ON–OFF subsystems, the midget and parasol ganglion receptive fields perform a separate and parallel tiling of the retina. On average, the receptive fields of parasol

TABLE 1 Important properties of the two major cell types providing input to the visual cortex

Property	P cells	M cells	Comments
Percent of cells	80	10	The remainder project to subcortical streams.
Receptive field size	relatively small, single cone center in fovea, increases with eccentricity (see Fig. 3)	Relatively large, ~3× larger than P cells at any given eccentricity	RF modeled well by a difference of Gaussians.
Contrast sensitvity	poor (factor of 8–10 lower than for M cells), driven by high contrast	good, saturation at high contrasts	
Contrast gain	low	high (~6× higher)	Possible gain control in M cells.
Spatial frequency response	peak and high-frequency cutoff at relatively low spatial frequency	Peak and high-frequency cutoff at relatively high spatial frequency	Unclear dichotomy: physiological differences tend to be less pronounced than predicted by anatomy.
temporal frequency response	low-pass, fall off at 20–30 Hz.	bandpass, peaking at or above 20 Hz	
Spatial linearity	almost all have linear summation	most have linear summation, some show marked nonlinearities	Estimated proportion of nonlinear neurons depends on how the distinction is made.
Wavelength opponency	yes	no	
Conduction velocity	slow (6 m/s)	fast (15 m/s)	

ganglion cells are about a factor of 3 larger than those of midget ganglion cells at any given eccentricity, as shown in Fig. 3(c), so the two systems can be thought of as operating in parallel at different spatial scales. This separation is strictly maintained in the projection to the LGN, which is layered like a wedding cake. The midget cells project exclusively to what are termed the *parvocellular* layers of the LGN (the dorsalmost four layers), and the parasol cells project exclusively to the *magnocellular* layers (the ventralmost two layers). Because of this separation and the important physiological distinctions that exist, visual scientist now generally speak in terms of the parvocellular (or "P") pathway, and the magnocellular (or "M") pathway.

There is a reliable difference in the temporal frequency response between the cells of the M and P pathways [17]. In general, the parvocellular cells peak at a lower temporal frequency than magnocellular cells (<10 Hz vs. 10–20 Hz), have a lower high-frequency cutoff (~20 Hz vs. ~60 Hz), and shallower low-frequency rolloff (with many P cells showing a DC response). The temporal frequency response envelopes of both cell types can be functionally modeled as a difference of exponentials in the frequency domain.

Another prevalent distinction is based upon linear versus nonlinear summation within a cell's receptive field. Two major classes of retinal ganglion cell have been described in the cat, termed X and Y cells, based on the presence or absence of a null phase when stimulated with a sinusoidal grating [15]. The response of a cell such as shown in Fig. 4 will obviously depend strongly on the spatial phase of the stimulus. For such a cell, a spatial phase of a grating can be found such that the grating can be exchanged with a blank field of equal mean luminance with no effect on the output of the cell. These X cells compose the majority. For other cells, termed Y cells, no such null phase can be found, indicating that something other than linear summation across space occurs.

In the primate, nonlinear spatial summation is much less prevalent at the level of the LGN; although nonlinear cells do exist, and are more prevalent in M cells than in P cells [17]. It may be that nonlinear processing, which is very important, has largely shifted to the cortex in primates, just as have other important functions such as motion processing, which occurs much earlier in the visual systems of more phylogenically-challanged species.

At this point, there is a great body of evidence suggesting that the M–P distinction is a fundamental one in primates, and that most of the above dichotomies are either an epiphenomenon of it, or at least best understood in terms of it. We can summarize the important parameters of M and P cells as follows. Table 1 (cf. [18]) provides a fairly comprehensive, albeit qualitative, overview of what we could term the magnocellular and parvocellular "geniculate transforms" that serve as the input to the cortex. If, in fact, work on the visual cortex continues to show effects such as malleability of receptive fields, it may be that models of geniculate function will actually increase in importance, because it may be the last stage at which we can confidently rely on a relatively linear transform-type model. Attempts in this direction have been made [19, 20] but most modeling efforts seem to have been concentrated on either cortical cells or psychophysical behavior (i.e., modeling the output of the human as a whole, e.g., contrast threshold in response to some stimulus manipulation).

4 The Primary Visual Cortex and Fundamental Properties of Vision

4.1 Neurons of the Primary Visual Cortex

The most striking feature of neurons in the visual cortex is the presence of several emergent properties. We begin to see, for example, orientation tuning, binocularity, and selectivity for the direction of motion. The distinction between the magnocellular and parvocellular pathways remains — they synapse at different input layers in the visual cortex — but interactions between them begin to occur.

Perhaps the most obvious and fundamental physiological distinction in the cortex is between so-called simple and complex cells [21, 22]. This terminology was adopted (prior to wide application of linear systems analysis in vision) because the simple cells made sense. Much as with ganglion cells, mapping the receptive field was straightforward and, once the receptive field was mapped, the response of the cell to a variety of patterns could be intuitively predicted. Complex cells, in contrast, were more complex. The simple/complex distinction seems to have no obvious relationship with the magnocellular/parvocellular distinction, but it seems to be a manifestation of a computational scheme used within both processing streams.

The spatial receptive field of a generic simple cell is shown in Fig. 5(a). The cell is modeled as a Gabor function, in which sensitivity is given by

$$s(x, y) = ae^{-(x^2/\sigma_x^2+y^2/\sigma_y^2)} \sin(2\pi\omega x + \phi) \tag{12}$$

As the axes are in arcminutes, the cell is most sensitive to horizontal Fourier energy at ~3 cycles/deg. In this case, the cell is odd symmetric. While it would be elegant if cells were always even or odd symmetric, it seems that phase is continuously represented [23, 24], although this certainly does not preclude the use of pairs of cells in quadrature phase in subsequent processing.

As in Fig. 4, Fig. 5(b) shows the spatiotemporal receptive field of the model cell: the cell's sensitivity at $y = 0$ plotted as a function of x and t. Notice that, in this case, the cell is spatiotemporally inseparable; it is oriented in space–time and is directionally selective [25, 26]. Thus, the optimal stimulus would be drifting sinusoidal grating, in this case a 3 cycle/deg grating drifting at approximately 5 deg/s. Many, but not all, cortical cells are directionally selective (see below).

Cells in the primary visual cortex can be thought of as a bank or banks of spatiotemporal filters that tile the visual world on several dimensions and, in so doing, determine the envelope of information to which we have access. We can get a feel for this envelope by looking at the distribution of cell tuning along various dimensions. This is done in Fig. 6 using data from cells in the Macaque primary visual cortex reported in Geisler and Albrecht [27]. In the upper row, the response of a typical cell is shown as a function of the spatial frequency of a counterphasing grating (left column), the temporal frequency of same stimulus at optimal spatial frequency (middle column), or the orientation of a drifting grating of optimal spatiotemporal frequency (right column). The middle and lower rows show the normalized frequency distributions of the parameters of the tuning functions for the population of cells surveyed ($n = 71$).[4]

At this point, we can sketch a sort of standard model of the spatial response properties of simple and complex cortical cells [e.g., 27, 28]. The basic elements of such a model are illustrated in Fig. 7(a). The model comprises four basic components, the first of which is a contrast gain control, which causes a response saturation to occur (see below). Typically, it takes the form of

$$r(c) = \frac{c^n}{c^n + c_{50}^n}, \tag{13}$$

in which c is the image contrast, c_{50} is the contrast at which half the maximum response is obtained, and n is the response exponent, which averages ~2.5 for Macaque cortical cells.

Next is the sampling of the image by a Gabor or Gabor-like receptive field, which is a linear spatial summation:

$$f(x, y) = \sum c(x, y)h(x, y), \tag{14}$$

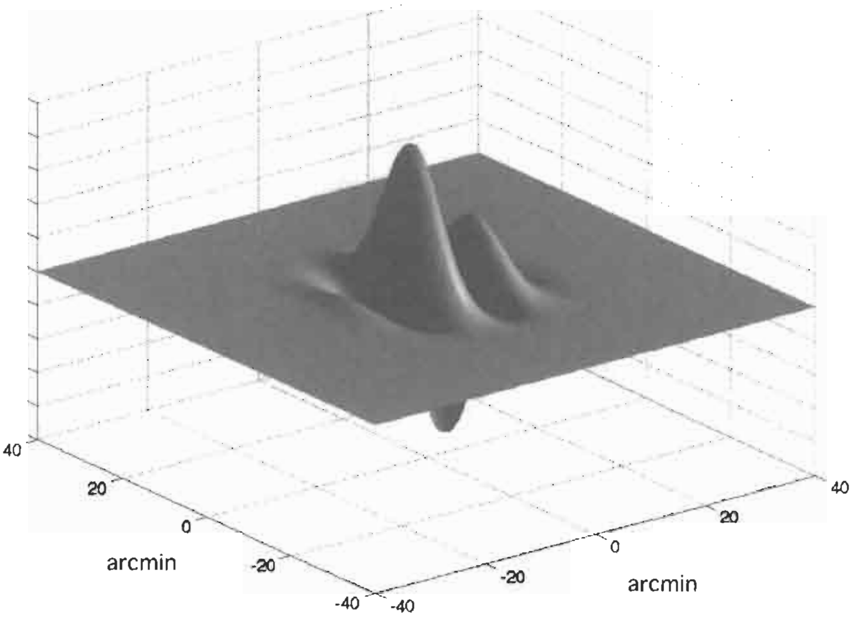
in which $h(x, y)$ is the spatial receptive field profile, and $c(x, y)$ is the effective contrast of the pixel at (x, y), i.e., the departure of the pixel value from the average pixel value in the image.

The third stage is a half-wave rectification (unlike ganglion cells, cortical cells have a low maintained discharge and thus can signal in only one direction) and an expansive nonlinearity, which serves to enhance the response disparity between optimal and nonoptimal stimuli. Finally, Poisson noise is incorporated, which provides a good empirical description of the response variability of cortical cells. The variance of the response of a cortical cell is proportional to the mean response with an average constant of proportionality of ~1.7.

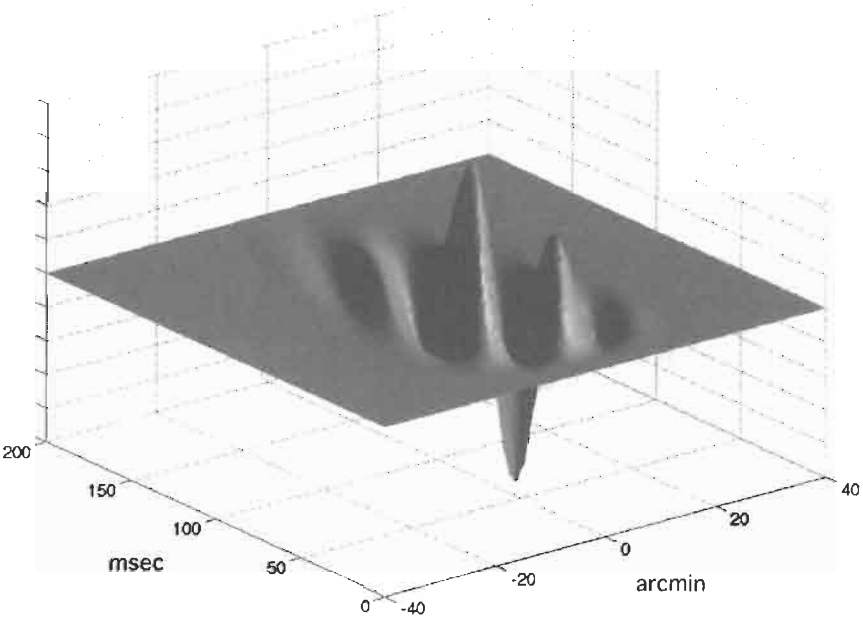
A model complex cell is adequately constructed by summing (or averaging) the output of two quadrature pairs of simple cells with opposite sign, as shown in Fig. 7(b) [e.g., 28]. Whether complex cells are actually constructed out of simple cells this way in primary visual cortex is not known; they could be constructed directly from LGN input. For modeling purposes, using simple cells to construct them is simply convenient. The important aspect is that their response is phase independent, and thus they behave as detectors of local contrast energy.

The contrast response of cortical cells deserves a little additional discussion. At first glance, the saturating contrast response function described above seems to be a rather mundane response limit, perhaps imposed by metabolic constraints. However, a subtle but key feature is that the response of a given cortical

[4]While these distributions are based on real data, they are schematized using a Gaussian assumption, which is probably not strictly valid. They do, however, convey a fairly accurate portrayal of the variability of the various parameters.



(a)



(b)

FIGURE 5 Receptive field profile of a cortical simple cell modeled as Gabor function: (a) spatial receptive field profile with the x and y axes in minutes of arc, and the z axis in arbitrary units of sensitivity; (b) space–time plot of the same receptive field with the x axis in minutes of arc and the y axis in milliseconds. The receptive field is space–time inseparable and the cell would be sensitive to rightward motion.

neuron saturates at the same *contrast*, regardless of overall response level (as opposed to saturating at some given *response* level). Why is this important? Neurons have a multidimensional sensitivity manifold, but a unidimensional output. Thus, if the output of a neuron increases from 10 to 20 spikes per second, say, then any number of things could have occurred to cause this. The contrast may have increased, the spatial frequency may have shifted to a more optimal one, etc., or any combination of such factors may have occurred. There is no way to identify which may have occurred from the output of the neuron.

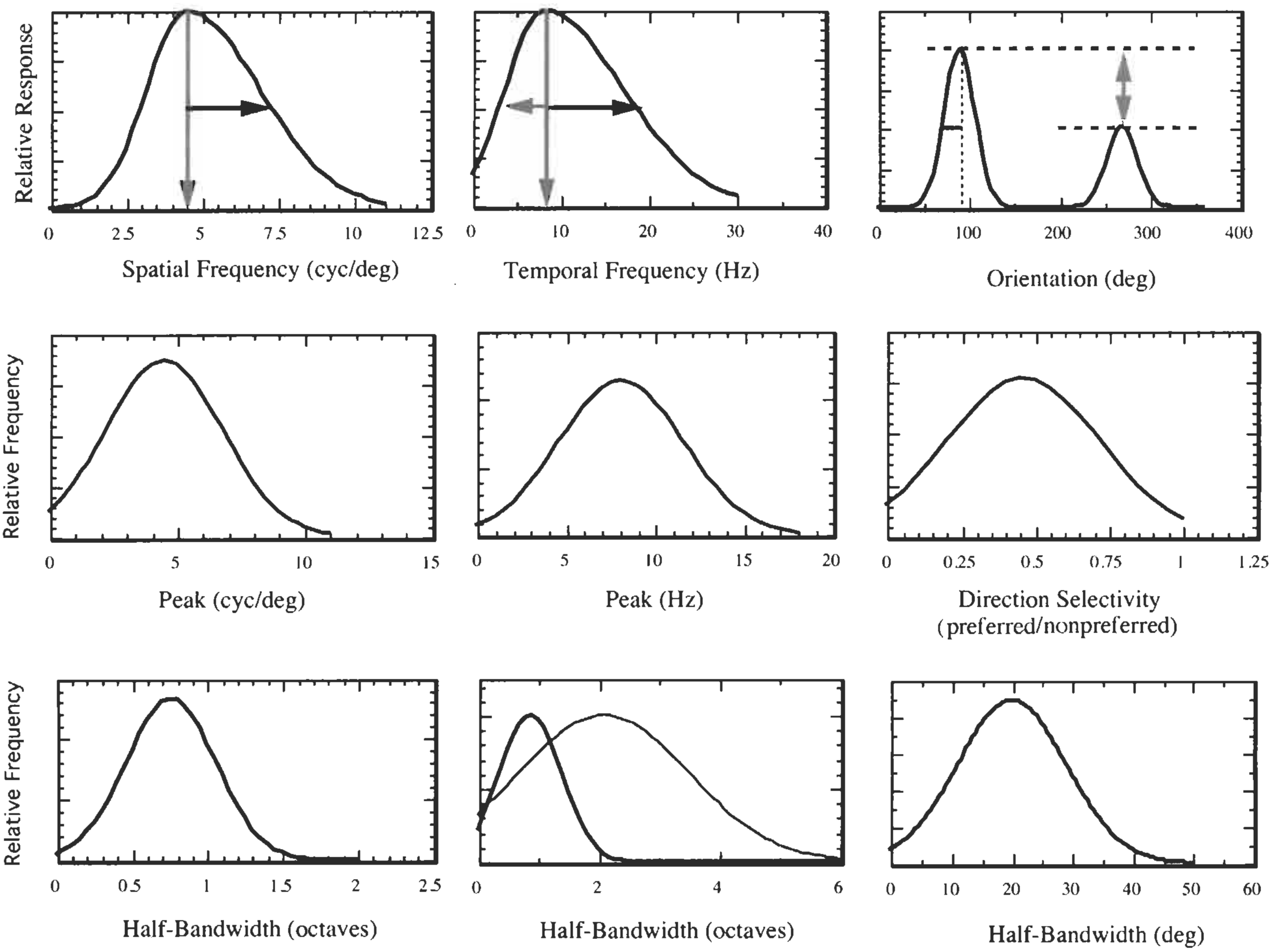


FIGURE 6 Left column: the upper panel shows a spatial frequency tuning profile typical of cell such as shown in Fig. 5; the middle and lower panels show distribution estimates of the two parameters of peak sensitivity (middle) and half-bandwidth in octaves (lower) for cells in macaque visual cortex. Middle column: same as the left column, but showing the temporal frequency response. As the response is asymmetric in octave bandwidth, the lower figure shows separate distributions for the upper and lower half-bandwidths (blue and green, respectively). Right column: the upper panel shows the response of a typical cortical cell to the orientation of a drifting sinusoidal grating. The ratio of responses between the optimal direction and its reciprocal is taken as an index of directional selectivity; the estimated distribution of this ratio is plotted in the middle panel (the index cannot exceed unity by definition). The estimate of half-bandwidth for Macaque cortical cells is shown in the lower panel. (See color section, p. C–8.)

But now consider the effect of the contrast saturation on the output of the neuron for both an optimal and a nonoptimal stimulus. Since the optimal stimulus is much more effective at driving the neuron, the saturation will occur at a higher *response* rate for the optimal stimulus. This effectively defeats the response ambiguity: *because of the contrast saturation, only an optimal stimulus is capable of driving the neuron to its maximum output.* Thus, if a neuron is firing at or near its maximum output, the stimulus is specified fairly precisely. Moreover, the expansive nonlinearity magnifies this by enhancing small differences in output. Thus, 95% confidence regions for cortical neurons on, for example, the contrast/spatial frequency plane are much narrower than the spatial frequency tuning curves themselves [29].

This suggests that it is important to rethink the manner in which subsequent levels of the visual system may use the information conveyed by neurons in primary visual cortex. Over the past $2^1/_2$ decades, linear system analysis has dominated the thinking in vision science. It has been assumed that the act of perception would involve a large-scale comparison of the outputs of many linear filters, outputs which would individually be very ambiguous. While such across-filter comparison is certainly necessary, it may be that the filters of primary visual cortex behave much more like "feature detectors" than we have been assuming.

I doubt that anyone reading a volume on image processing could look at receptive profiles in the cortex (such as shown in Fig. 5) and not be reminded of schemes such as a wavelet

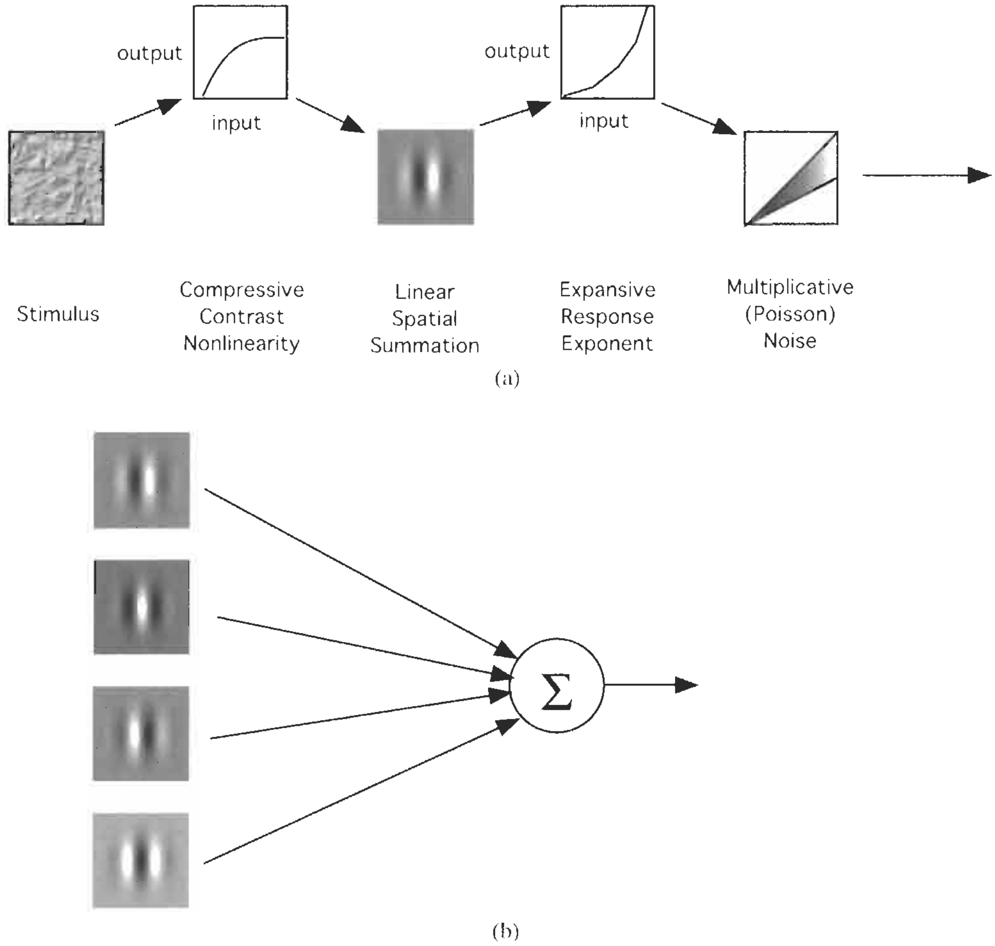


FIGURE 7 (a) Overview of a model neuron similar to that proposed by Heeger and colleagues (1991, 1996) and Geisler and Albrecht (1997). An early contrast saturation precedes linear spatial summation across the Gabor-like receptive field; the contrast saturation ensures that only optimal stimuli can maximally stimulate the cell (see text). An expansive nonlinearity such as half-squaring enhances small differences in output. Multiplicative noise is then added; the variance of cortical cell output is proportional to the mean response (with the constant of proportionality ~1.7), so the signal-to-noise ratio grows as the square root of output. (b) Illustration of the construction of a phase-independent (i.e., energy detecting) complex cell from simple cell outputs.

transform or Laplacian pyramid. Not surprisingly, then, most models of the neural image in primary visual cortex share the property of encoding the image in parallel at multiple spatial scales, and several such models have been developed. One model that is computationally very efficient and easy to implement is the cortex transform [30]. The cortex transform is not, nor was it meant to be, a full model of the cortical representation. For example, response nonlinearities, the importance of which were discussed above, are omitted. It does, however, produce a simulated neural image that shares many of the properties of the simple cell representation in primary visual cortex. Models such as this have enormous value in that they give vision scientists a sort of testbed that can be used to investigate other aspects of visual function, e.g., possible interactions between the different frequency and orientation bands, in subsequent visual processes such as the computation of depth from stereopsis.

4.2 Motion and Cortical Cells

As mentioned previously, ganglion cell receptive fields are space–time separable. The resulting symmetry around a constant-space axis [Fig. 4(b)] makes them incapable of coding the direction of motion. Many cortical cells, in contrast, are directionally selective.

In the analysis of motion, a representation in space–time is often most convenient. Figure 8 (top) shows three frames of a moving spot. The continuous space–time representation is shown beneath, and it is simply an oriented bar in space–time. The next row of the figure shows the space–time representation of both a rightward and leftward moving bar. The third row of the figure shows a space–time receptive field of a typical cortical cell as was also shown in Fig. 5 (for clarity, it is shown enlarged relative to the stimulus). The orientation of the

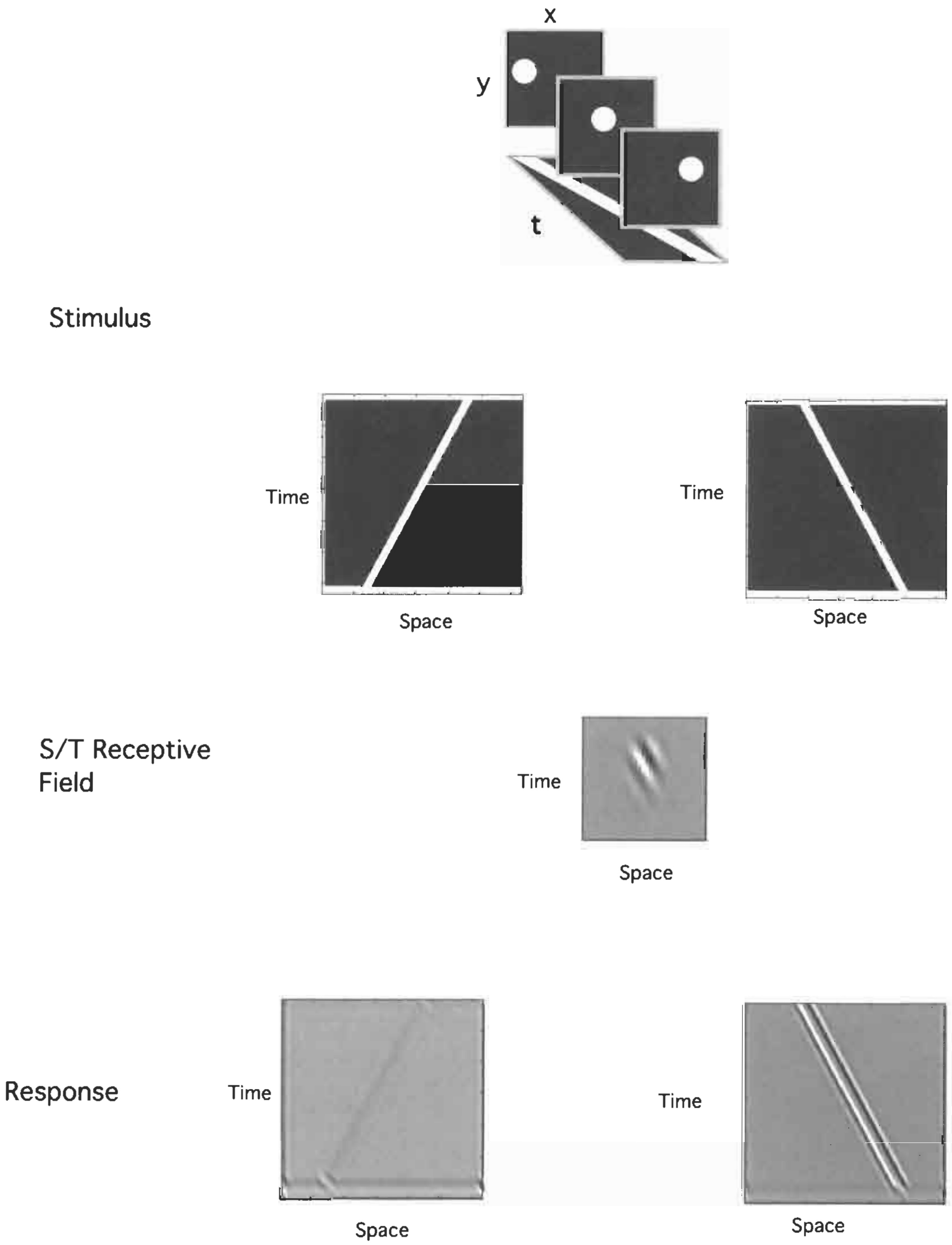


FIGURE 8 Three x–y slices are shown of a spot moving from left to right, and directly below is the continuous x–t representation: a diagonal bar. Below this are the space–time representations of a leftward and rightward moving bar, the receptive field of a directionally selective cortical cell (shown enlarged for clarity), and the response of the cell to the leftward and rightward stimuli.

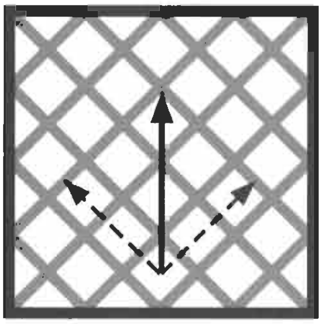
receptive field in space–time gives it a fairly well defined velocity tuning; it effectively performs an autocorrelation along a space–time diagonal. Such space–time inseparable receptive fields are easily constructed from ganglion cell inputs by summing pairs of space–separable receptive fields (such as those shown in Fig. 4), which are in quadrature in both the space and time domains [25, 26].

The bottom row of the figure shows the response of such cells to the stimuli shown in the second row obtained by convolution. In these panels, each column represents the output of a cell as a function of time (row), and each cell has a receptive field centered at the spatial location represented by its column. Clearly, each cell produces vigorous output modulation in response to motion in the preferred direction (with a relative time delay proportional to its spatial position, obviously), and almost no output in response to motion in the opposite direction.

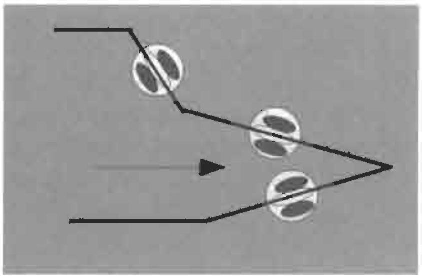
For most purposes, it would be desirable to sense "motion energy." That is, one desires units that would respond to motion in one direction regardless of the sign of contrast or the phase of the stimulus. Indeed, such motion energy units may be thought of as the spatiotemporal equivalent of the complex cells described above. Similar to the construction of complex cells, such energy detectors are easily formed by, for example, summing the squared output of quadrature pairs of simple velocity sensitive units. Such a model captures many of the basic attributes of human motion perception, as well as a some common motion illusions [25].

Motion sensing is vital. If nothing else, a primitive organism asks its visual system to sense moving things, even if it is only the change in a shadow which triggers a sea scallop to close. It is perhaps not surprising, then, that there seems to be a specialized cortical pathway, an extension of the magnocellular pathway an earlier levels, for analyzing motion in the visual field. A review of the physiology and anatomy of this pathway is clearly beyond the scope of this chapter. One aspect of the pathway worth mentioning here, however, is the behavior of neurons in an area of the cortex known as MT, which receives input from primary visual cortex (it also receives input from other areas, but for our purposes, we can consider only its V1 inputs).

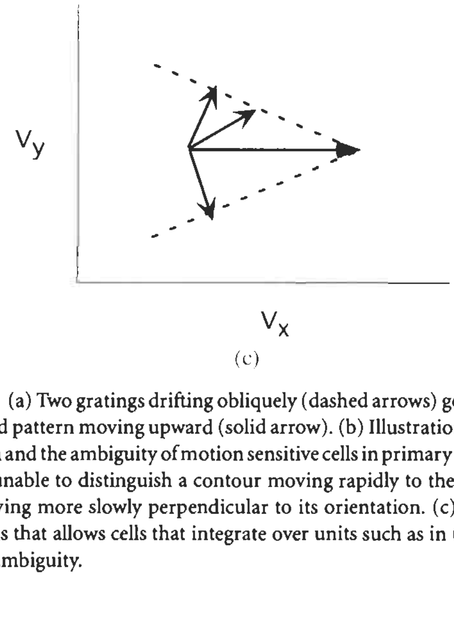
Consider a "plaid" stimulus, as illustrated in Fig. 9(a) composed of two drifting gratings differing in orientation by 90° — one drifting up and to the right and the other up and to the left. When viewing such a stimulus, a human observer sees an array of alternating dark and light areas — the intersections of the plaid — drifting upward. The response of cells such as pictured in Fig. 5, however, would be quite different. Such cells would respond in a straightforward way according to the Fourier energy in the pattern, and would thus signal a pair of motion vectors corresponding to the individual grating components of the stimulus. Obviously, then, the human visual system incorporates some mechanism that is capable of combining motion estimates from filters such as the cells in primary visual cortex to yield estimates of motion for more complex structures. These mechanisms, corresponding to cells in area MT, can be parsimoniously modeled by combining complex cell outputs in manner similar to that by which complex cells can be constructed from simple cell outputs [31, 32]. These cells effectively perform a local sum over the set of cells tuned to the appropriate orientation and spatiotemporal frequency combinations consistent with a real object moving in a given direction at a given rate. In effect, then, these cells are a neural implementation of the intersection-of-constraints solution to the aperture problem of edge (or grating) motion [33]. This problem is illustrated in Figs. 9(b) and 9(c). In Fig. 9(b), an object is shown moving to the right with some velocity. Various edges along the object will stimulate receptive fields with the appropriate orientation. Clearly, these individual cells have no way of encoding the true motion of the *object*. All they can sense is the motion of the *edge*, be it almost orthogonal to the motion of the object at a relatively low speed, or in the direction of the object at a relatively high speed. The set of motion vectors generated by the edges, however, must satisfy the intersection of motion constraint as illustrated in Fig. 9(c). The endpoints of the motion vectors generated by the moving edges lie on a pair of lines that intersect at the true motion of the object. Thus, a cell summing (or averaging) the outputs of receptive fields of the appropriate orientation and spatiotemporal frequency (i.e., speed) combinations will effectively be tuned to a particular



(a)



(b)



(c)

FIGURE 9 (a) Two gratings drifting obliquely (dashed arrows) generate a percept of a plaid pattern moving upward (solid arrow). (b) Illustration of the aperture problem and the ambiguity of motion sensitive cells in primary visual cortex. Each cell is unable to distinguish a contour moving rapidly to the right from a contour moving more slowly perpendicular to its orientation. (c) Intersection of constraints that allows cells that integrate over units such as in (b) to resolve the motion ambiguity.

velocity and largely independent of the structure moving at that velocity.

4.3 Stereopsis and Cortical Cells

Stereopsis refers to the computation of depth from the image displacements that result from the horizontal separation of the eyeballs. Computationally, stereopsis is closely related to motion, the former involving displacements across viewpoint rather than across time. For this reason, the development of models in the two domains has much in common. Early models tended to focus on local correlations between the images, and excitatory or inhibitory interactions in order to filter out false matches (spurious correlations).

As with motion, however, neurophysiological and psychophysical findings [e.g., 34] have served to concentrate efforts on models based on receptive field structures similar to those found in Fig. 5. Of course, this is not incompatible with disparity domain interactions, but ambiguity is more commonly eliminated via interactions between spatial scales.

The primary visual cortex is the first place along the visual system in which information from the two eyes converges on single cells; as such, it represents the beginning of the binocular visual processing stream. Traditionally, it has been assumed that in order to encode horizontal disparities, these binocular cells received monocular inputs from cells with different receptive field *locations* in the two eyes, thus being maximally stimulated by an object off the plane of fixation. It is now clear, however, that binocular simple cells in the primary visual cortex often have receptive fields like that shown in Fig. 5, but with different *phases* between the two eyes [35].[5] The relative phase relation between the receptive fields in the two eyes is distributed uniformly (not in quadrature pairs) for cells tuned to vertical orientations, whereas there is little phase difference for cells tuned to horizontal orientations, indicating that these phase differences are almost certainly involved in stereopsis. Just as in motion, however, these simple cells have many undesirable properties, such as phase sensitivity and phase ambiguity (a phase disparity $k\pi$ being indistinguishable from a phase disparities of $2nk\pi$, where n is an integer).

To obviate the former difficulty, an obvious solution would be to build a binocular version of the complex cell by summing across simple cells with the same disparity tuning but various monocular phase tunings [e.g., 36]. Such construction is analogous to the construction of phase-independent, motion-sensitive complex cells discussed earlier, except that the displacement of interest is across eyeballs instead of time. This has been shown to occur in cortical cells and, in fact, these cells show more precise disparity tuning than 2-D position tuning [37].

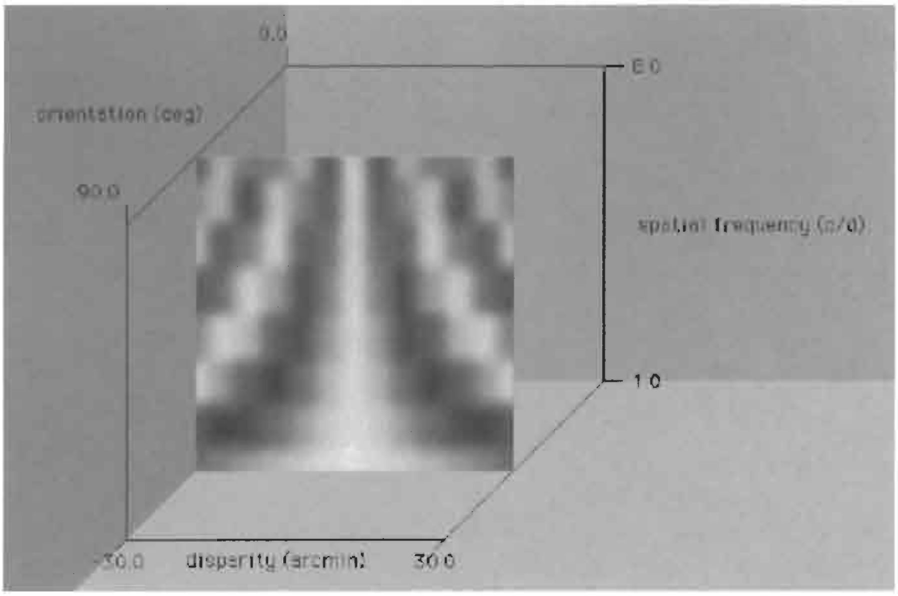
[5] Many recent studies have not measured the absolute receptive field position in the two eyes, as it is very difficult to do. Thus, the notion that absolute monocular receptive field position plays a role in stereopsis cannot be rejected.

Yet, because these cells are tuned to a certain phase disparity of a given *spatial frequency*, there remains an ambiguity concerning the absolute disparity of a stimulus. This can be seen in Fig. 10, which plots the output (as brightness) of a hypothetical collection of cells tuned to various values of phase disparity, orientation, and spatial frequency. The tuning of the cell is given by its position in the volume; in Fig. 10(a) orientation is ignored, and only a single spatial frequency/disparity surface is shown. In Fig.10(a), note that the output of cells tuned to a single spatial frequency contains multiple peaks along the dimension of disparity, indicating the phase ambiguity of the output. It has been suggested that this ambiguity could be resolved by units that sum the outputs of disparity units across spatial frequency and orientation [e.g., 36]. Such units would solve the phase ambiguity in a manner very analogous to the intersection-of-constraints solution to motion ambiguity described above. In the case of disparity, as a broadband stimulus is shifted along the disparity axis, it yields a sinusoidal variation in output at all spatial frequencies, but the frequency of modulation is proportional to the spatial frequency to which the cells are tuned. The resolution to the ambiguity lies in the fact that there is only *one* disparity at which peak output is obtained at *all* spatial frequencies, and that is the true disparity of the stimulus. This is shown in Fig. 10(a) by the white ridge running down the spatial frequency — disparity plane.

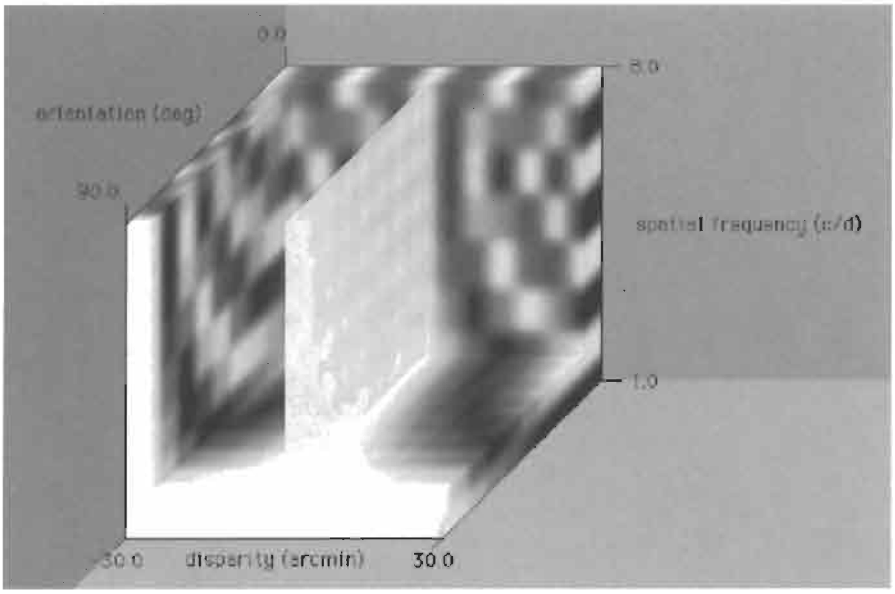
The pattern of outputs of cells tuned to a single spatial frequency but to a variety of orientations as a function of disparity is shown on the floor of Fig. 10(b). Summing across cells tuned to different orientations will also disambiguate disparity information because a Fourier component at an oblique orientation will behave as a vertical component with a *horizontal* frequency proportional to the cosine of the angle of its orientation from the vertical.

Figure 10(b) is best thought of as a volume of cells whose sensitivity is given by their position in the volume (for visualization convenience, the phase information is repeated for the higher spatial frequencies, so the phase tuning is giving by the position on the disparity axis modulo 2π). The combined spatial frequency and disparity information results in a surface of maximum activity at the true disparity of a broadband stimulus, so a cell that sums across surfaces in this space will encode for physical disparity independent of spatial frequency and orientation.

Very recent work indicates that cells in MT might perform just such a task [38]. Recall from above that cells in MT decouple velocity information from the spatial frequency and orientation sensitivity of motion selective cells. DeAngelis *et al.* [38] have discovered a patterned arrangement of disparity sensitive cells in the same area and have demonstrated their consequence in perceptual judgments. Given the conceptually identical nature of the ambiguities to be resolved the domains of motion and disparity, it would seem likely that the disparity-sensitive cells in MT perform role in stereopsis analogous to that which the velocity-sensitive cells play in motion perception.



(a)



(b)

FIGURE 10 (a) Output of cortical cells on the spatial-frequency/disparity plane. The output of any one cell uniquely specifies only a phase disparity, but summation across spatial frequencies at the appropriate phase disparities uniquely recovers absolute disparity. (b) Orientation is added to this representation.

5 Concluding Remarks

Models are wonderful tools and have an indispensable role in vision science. Neuroscientists must reverse-engineer the brain, and for this the methods of engineering are required. But the tools themselves can lead to biases (when all you have is a hammer, everything looks like a nail). There is always a danger of carrying too much theory, often implicitly, into an analysis of the visual system. This is particularly true in the case of modeling, because a model must have a quantitative output and thus must be specified, whether intentionally or not, at what Marr called the level of computational theory [7]. Tools, like categories, make wonderful servants but horrible masters.

Yet without quantitative models, it would be almost impossible to compare psychophysics (human behavior) and physiology except in trivial ways.[6] This may seem like a strong statement, but there are subtle flaws in simple comparisons between the results of human experiments and single-cell response profiles. Consider an example taken from [39]. The experiment was designed to reveal the underlying mechanisms of disparity processing. A "mechanism" is assumed to comprise many neurons with similar tuning properties (peak location and bandwidth) on the dimension of interest working in parallel to encode that dimension. The tuning of the mechanism then reflects the tuning of the underlying neurons. This experiment used the typical psychophysical technique of *adaptation*. In this technique, one first measures the sensitivity of human observers along a dimension; in this case, we measured the sensitivity to the interocular correlation of binocular white noise signals as a function of binocular disparity. Following this, the subjects adapted to a signal at a given disparity. This adaptation fatigues the neurons sensitive to this disparity and therefore reduces the sensitivity of any mechanism comprising these neurons. Retesting sensitivity, we found that it was systematically elevated in the region of the adaptation, and a difference between the pre- and postadaptation sensitivity yielded a "tuning profile" of the adaptation, for which a peak location, bandwidth, etc. can be defined.

But what *is* this tuning profile? In these types of experiments, it is tempting to assume that it directly reflects the sensitivity profile of an underlying mechanism, but this would be a dangerous and generally wrong assumption. The tuning profile actually reflects the combined outputs of numerous mechanisms in response to the adaptation. The degree to which the tuning profile itself resembles any one of the individual underlying mechanisms depends on a number of factors involving the nature of the mechanisms themselves, their interaction, and how they are combined at subsequent levels to determine overall sensitivity.

If one cannot get a direct glimpse of the underlying mechanisms using psychophysics, how does one reveal them? This is where computational models assert their value. We constructed various models incorporating different numbers of mechanisms, different mechanism characteristics, and different methods of combining the outputs of mechanisms. We found that with a small number of disparity-sensitive mechanisms (e.g., three, as had been proposed by earlier theories of disparity processing) we were unable to simulate our psychophysical data. With a larger number of mechanisms, however, we able to reproduce our data rather precisely, and the model became much less sensitive to the manner in which the outputs of the mechanisms were combined.

So although we are unable to get a *direct* glimpse at underlying mechanisms using psychophysics, models can guide us in determining what kinds of mechanisms can and cannot be used to produce sets of psychophysical data. As more physiological data become available, more precise models of the neurons themselves can be constructed, and these can be used, in turn, within models of psychophysical behavior. It is thus that models sew together psychophysics and physiology, and I would argue that without them the link could never be but tenuously established.

[6] Psychophysicists, such as myself, attempt to quantify the performance of human sensory and perceptual systems. Psychophysics encompasses a host of experimental techniques used to determine the ability of sensory systems (e.g., the visual system) to detect, discriminate, and/or identify well-defined and tightly-controlled input stimuli. These techniques share a general grounding in signal detection theory, which itself grew out of electronic communication theory and statistical decision theory.

References

[1] I. Newton, *Opticks* (G. Bell & Sons, London, 1931).

[2] T. Young, "On the theory of light and color," *Phil. Trans. Roy. Soc.* **73**, 12–48 (1802).

[3] H. V. Helmholtz, *Treatise on Physiological Optics* (Dover, New York, 1962).

[4] G. Westheimer, "The eye as an optical instrument," in K. R. Boff, L. Kaufman, and J. P. Thomas, eds., *Handbook of Human Perception and Performance* (Wiley, New York, 1986).

[5] A. Roorda and D. R. Williams, "The arrangement of the three cone classes in the living human eye, *Nature* **397**, 520–522 (1999).

[6] W. S. Geisler, "Sequential ideal-observer analysis of visual discriminations," *Psychol. Rev.* **96**, 267–314 (1989).

[7] D. R. Williams and N. J. Coletta, "Cone spacing and the visual resolution limit," *J. Opt. Soc. Am.* A **4**, 1514–1523 (1987).

[8] G. Wyszecki and W. S. Stiles, *Color Vision* (Wiley, New York, 1982).

[9] W. S. Geisler and M. S. Banks, "Visual performance," in M. Bass, ed., *Handbook of Optics* (McGraw-Hill, New York, 1995).

[10] H. B. Barlow, "Measurements of the quantum efficiency of descrimination in human scotopic vision," *J. Physiol.* **150**, 169–188 (1962).

[11] D. G. Pelli, "The quantum efficiency of vision," in C. Blakemore, ed., *Vision: Coding and Efficiency* (Cambridge U. Press, Cambridge, 1990).

[12] W. S. Geisler and K. Chou, "Separation of low-level and high-level factors in complex tasks: visual search," *Psychol. Rev.* **102**, 356–378 (1995).

[13] D. Marr, *Vision* (Freeman, New York, 1982).

[14] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vis. Res.* **5**, 583–601 (1965).

[15] C. Enroth-Cugel and J. G. Robson, "The contrast sensitivity of retinal ganglion cells in the cat," *J. Physiol.* **187**, 517–522 (1966).

[16] A. B. Watson, "Temporal sensitivity," in K. R. Boff, L. Kauffman, and J. P. Thomas, eds., *Handbook of Perception and Human Performance* (Wiley, New York, 1986).

[17] A. M. Derrington and P. Lennie, "Spatial and temporal contrast sensitivities of neurons in the lateral geniculate nucleus of Macaque," *J. Physiol.* **357**, 2219–240 (1984).

[18] P. Lennie, "Roles of M and P pathways," in R. Shapley and D. M. K. Lam, eds., *Contrast Sensitivity* (MIT, Cambridge, 1993).

[19] J. B. Troy, "Modeling the receptive fields of mammalian retinal ganglion cells," in R. Shapley and D. M. K. Lam, eds., *Contrast Sensitivity* (MIT Press, Cambridge, 1993).

[20] K. Donner and S. Hemila, "Modeling the spatio-temporal modulation response of ganglion cells with difference-of-Gaussians receptive fields: relation to photoreceptor response kinetics," *Visual Neurosci.* **13**, 173–186 (1996).

[21] D. H. Hubel and T. N. Weisel, "Receptive fields, binocular interaction and functional archetecture in the cat's visual cortex," *J. Physiol.* **160**, 106–154 (1962).

[22] B. C. Skottun, R. L. DeValois, D. H. Grosof, J. A. Movshon, D. G. Albrecht, and A. B. Bonds, "Classifying simple and complex cells on the basis of response modulation," *Vis. Res.* **31**, 1079–1086 (1991).

[23] D. B. Hamilton, D. G. Albrecht, and W. S. Geisler, "Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function," *Vis. Res.* **29**, 1285–1308 (1989).

[24] D. J. Field and D. J. Tolhurst, "The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex," *Proc. Roy. Soc. London* **228**, 379–400 (1986).

[25] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am.* A **2**, 284–299 (1985).

[26] A. B. Watson and A. J. Ahumada, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am.* A **2**, 322–341 (1985).

[27] W. S. Geisler and D. G. Albrecht, "Visual cortex neurons in monkeys and cats: detection, discrimination, and stimulus certainty," *Visual Neurosci.* **14**, 897–919 (1997).

[28] D. J. Heeger, "Nonlinear model of neural responses in cat visual cortex," in M. S. Landy and J. A. Movshon, eds., *Computational Models of Visual Processing* (MIT Press, Cambridge, 1991).

[29] W. S. Geisler and D. G. Albrecht, "Bayesian analysis of identification performance in monkey visual cortex: nonlinear mechanisms and stimulus certainty," *Vis. Res.* **35**, 2723–2730 (1995).

[30] A. B. Watson, "The cortex transform: rapid computation of simulated neural images," *Comput. Vis. Graph. Image Process.* **39**, 311–327 (1987).

[31] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vis. Res.* **38**, 743–761 (1998).

[32] D. J. Heeger, E. P. Simoncelli, and J. A. Movshon, "Computational models of cortical visual processing," *Proc. Nat. Acad. Sci.* **93**, 623–627 (1996).

[33] E. H. Adelson and J. A. Movshon, "Phenomenal coherence of visual moving pattens," *Nature* **300**, 523–525 (1982).

[34] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, "Depth is encoded in the visual cortex by a specialized receptive field structure," *Nature* **352**, 156–159 (1991).

[35] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, "Encoding of binocular disparity by simple cells in the cat's visual cortex," *J. Neurophysiol.* **75**, 1779–1805 (1996).

[36] D. J. Fleet, H. Wagner, and D. J. Heeger, "Neural encoding of binocular disparity: energy models, position shifts, and phase shifts," *Vis. Res.* **36**, 1839–1858 (1996).

[37] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, "Encoding of binocular disparity by complex cells in the cat's visual cortex," *J. Neurophysiol.* **76**, 2879–2909 (1997).

[38] G. C. DeAngelis, B. G. Cumming, and W. T. Newsome, "Cortical area MT and the perception of stereoscopic depth," Nature **394**, 677–680 (1998).

[39] S. B. Stevenson, L. K. Cormack, C. M. Schor, and C. W. Tyler, "Disparity-tuned mechanisms of human stereopsis," *Vis. Res.* **32**, 1685–1689 (1992).

[40] R. W. Rodiek, *The First Steps in Seeing* (Sunderland, Sinauer Associates, 1998).

Index

D

E

F

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Electrical Engineering/Image and Signal Processing/Communications Computer Science/Computer Graphics, Internet and Multimedia

Handbook of Image & Video Processing

Editor **AL BOVIK**
University of Texas, Austin

A VOLUME IN THE ACADEMIC PRESS SERIES IN COMMUNICATIONS, NETWORKING, AND MULTIMEDIA

SERIES *EDITOR-IN-CHIEF* JERRY D. GIBSON

Handbook of Image and Video Processing presents a comprehensive and highly accessible presentation of the basic and most up-to-date methods and algorithms for digital image and video processing. This timely volume will provide both the novice and the seasoned practitioner the necessary information and skills to be able to develop algorithms and applications for the burgeoning Multimedia, Digital Imaging, Digital Video, Telecommunications, and World-Wide Web (internet) industries. *Handbook of Image and Video Processing* is an indispensible resource for researchers in telecommunications, internet applications, multimedia, and nearly every branch of science. No other resource contains the same breadth of up-to-date coverage.

This handbook is arranged into highly focused chapters that represent the collective efforts of the leading educators and researchers working in the areas of image and video processing. Beginning with a series of tutorial chapters on basic gray-level image processing, binary image processing, image Fourier analysis and convolution, the Handbook then describes the latest and most effective techniques for:

- Linear, non-linear, morphological, and wavelet-based image enhancement
- Basic, regularized, multi-channel, multi-frame, and iterative image restoration
- Motion detection and estimation
- Video enhancement and restoration
- Scene reconstruction, image stabilization, and mosaicking
- Models of human vision and their impact on image processing
- Wavelet, color, and multispectral image representations
- Models for image noise, image modulations, and random fields
- Image and video segmentation, classification, and edge detection
- Review of available image processing development environments and software
- Lossless image compression
- Lossy image compression using BTC, vector quantization, and wavelets
- Image compression standards, including JPEG
- Modern video compression, including DCT, object-, and wavelet-based methods
- Video compression standards, including H.261, and MPEG I, II, IV, and VII
- Image and video acquisition, sampling, and interpolation
- Image quantization, halftoning, and printing
- Perceptual quality assessment of compressed images and video
- Image and video databases, indexing, and retrieval
- Image and video networks, security, and watermarking

The Handbook concludes with a set of carefully selected, instructive, and exemplary image processing applications in diverse areas such as; radar imaging, computed tomography, cardiac imaging, digital mammography, fingerprint classification and recognition, human face recognition, confocal microscopy, and automatic target recognition. Developers of these applications as well as those seeking applications that parallel their own will find these chapters to be indispensable guides.

PRINTED IN CANADA

ACADEMIC PRESS
A Harcourt Science and Technology Company

UPC
6 08628 97905 2

EAN
ISBN 0-12-119790-5
90119
9 780121 197902