Technical Report 1154

The Computation of Color

Anya C. Hurlbert

MIT Artificial Intelligence Laboratory

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AI-TR 1154	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): The Computation of Color	5. TYPE OF REPORT & PERIOD COVERED: technical report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Anya C. Hurlbert	8. CONTRACT OR GRANT NUMBER(s): DACA76-85-C-0010 N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209	12. REPORT DATE: September 1989 13. NUMBER OF PAGES: 208
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Office of Naval Research Information Systems Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

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

color vision: review of computational models
image segmentation
lightness algorithms
color constancy
learning from examples
color vision: psychophysics

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

See reverse side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

THE COMPUTATION OF COLOR

by

Anya C. Hurlbert

Submitted to the Department of
Brain and Cognitive Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

at the

Massachusetts Institute of Technology
September, 1989

This paper describes research done within the Center for Biological Information Processing, in the Department of Brain and Cognitive Sciences, and at the Artificial Intelligence Laboratory. This research is sponsored by a grant from the Office of Naval Research (ONR), Cognitive and Neural Sciences Division; by the Artificial Intelligence Center of Hughes Aircraft Corporation; by the Alfred P. Sloan Foundation; by the National Science Foundation; by the Artificial Intelligence Center of Hughes Aircraft Corporation (S1-801534-2); and by the NATO Scientific Affairs Division (0403/87). Support for the A. I. Laboratory's artificial intelligence research is provided by the Advanced Research Projects Agency of the Department of Defense under Army contract DACA76-85-C-0010, and in part by ONR contract N00014-85-K-0124. A.C.H. was sponsored by a Whitaker Health Sciences Fund M.D./Ph.D. Fellowship.

Abstract

Color vision has traditionally been the domain of many sciences: physics, physiology, psychology, and philosophy. This thesis maintains that interdisciplinary tradition and looks at color vision from several points of view. It focusses on the phenomenon of color constancy and uses its formulation as a computational problem to link the different viewpoints.

The primary contributions of the thesis are (1) the demonstration of a formal framework for lightness algorithms, which represent one class of solution to the problem of color constancy; (2) the derivation of a new lightness algorithm based on regularization theory; (3) the synthesis of an adaptive lightness algorithm using "learning" techniques; (4) the development of a segmentation algorithm that uses color information to mark material boundaries, with guidance from luminance edges; and (5) an experimental investigation into the cues that human observers use to judge the color of the illuminant, demonstrating that under certain conditions, observers ignore correct information from specular reflections in favor of incorrect information from other cues. Other computational approaches to the problem of color constancy are reviewed and some of their links to psychophysics and physiology are explored.

Acknowledgements

First and foremost thanks go to Tomaso Poggio, my thesis advisor, who was magnanimous enough to treat me always as a colleague and friend. His boundless energy, enthusiasm, and optimism are matched only by his scientific skill and inventiveness. His intellectual generosity nurtured many of the ideas in this thesis.

Thanks also to the other members of my thesis committee: to Ellen Hildreth for her careful criticisms; to Peter Schiller for patiently introducing me to neurophysiology; and especially to Hsien-Che Lee, for his meticulous reading of thesis drafts, insights into color science, and inspiring collaboration on the "Mondrian-sphere" experiments. They may take credit for many of the improvements on earlier drafts but none of the blame for the fuzziness that still remains. Thanks also to Shimon Ullman, who provided guidance in the early stages of my research. Jeremy Wolfe and Whitman Richards also helped to prod me in the right direction.

For their computer expertise and help in deciphering color graphics systems, thanks to Heinrich Bülthoff and Nikos Logothetis. Thanks also to Heinrich for his essential contributions to the "Mondrian-spheres" experiments. Special thanks to Nikos for the hours and hours he spent helping me learn "C," neurophysiology laboratory techniques, and a host of other basics of experimental neuroscience. His unceasing willingness to help and to teach are truly remarkable.

For stimulating my interest in color, providing generous hospitality on numerous visits to the Rowland Institute, and maintaining an inimitable and admirable style in science, invention, and business, thanks to Edwin Land.

For indefinite loans of equipment, many provocative discussions, and downtown parties with an international flair, thanks to John McCann and his group at Polaroid.

Thanks also to Robert Savoy and Mike Burns of the Rowland Institute. Special thanks to Robert for sharing his wise experience in doing psychophysics on computer-driven CRT color displays. Thanks to Larry Arend of the Retina Foundation and Dan Ts'o of the Rockefeller Foundation for sharing similar experiences, and especially to Larry for illuminating discussions on lightness. Horace Barlow, David Hubel, Margaret Livingstone, Frank Wilczek, Alan Yuille and many others were also kind enough to engage in helpful discussions.

For the sharpness of his mind, the quickness and often devastating accuracy of his criticisms, and for allowing me to spend time with one of the greatest scientists of our time, I thank Francis Crick.

Many thanks to the Department of Brain and Cognitive Sciences for providing a stimulating and supportive environment, especially to Emilio Bizzi, Janice Ellertsen and the others who run the department. Thanks also to the Health Sciences and Technology Program for its consistent support. The Whitaker Health Sciences Fund provided financial support for the last two years of my research; I thank the Whitaker

family for their generosity and Dean Irwin Sizer for administering the fellowship.

Heartfelt thanks to all the people at the Center for Biological Information Processing (especially to Elizabeth Willey, Polly McGahan, Daphna Weinshall, Norberto Grzywacz, and Shimon Edelman) in general for being helpful and convivial, as expert at debugging and debunking as at having fun, and in particular, for bearing with me, especially during the printer-hogging, nose-burying, social-niceties-neglecting final stages of completing this document. Special thanks to perhaps the most congenial office-mate ever to grace a research center, Lyle Borg-Graham. Thanks also to my peers at the Artificial Intelligence Lab, who keep computers, spirits, and monthly Chinese dinners up – especially Davi Geiger and Mike Drumheller – and to my colleagues in the Schiller lab. Thanks also to other faithful friends, especially Eric McFarland and Philippe Sands.

Finally, for their emotional, intellectual, nutritional and financial support during these uncountable years of higher education, thanks to my parents. And to Matt, wordless gratitude.

Contents

Chapter 1

Introduction

Color vision has traditionally been the domain of many sciences. Physicists, physiologists, psychologists, physicians and philosophers all have found its features amenable to analysis using the techniques of their specialities. This thesis maintains that interdisciplinary tradition and looks at color vision from several points of view. It focusses on the phenomenon of color constancy and uses its formulation as a computational problem to link the different viewpoints.

Color constancy is the tendency of objects to stay the same perceived color under changing illumination. The computational approach to color vision in general, and to color constancy in particular, starts with the question: what is the goal of color vision? Why do humans exhibit color constancy? By stating clearly the goal of color constancy, and by specifying the input data and physical limitations with which the visual system must work, computational theorists hope to arrive at a problem that can be solved by computational methods. The computational solution, in turn, may be used not only to equip machines with color vision but also to shed light on the mechanisms underlying human color vision.

From a computational viewpoint, color vision serves two goals: first, to help segment an image into distinct materials – that is, to help discriminate objects in a given scene; and second, to assign objects labels that stay constant under changes in the illumination and the scene composition – that is, to help recognize objects

whatever their context. A ripe, yellow banana should always be distinguishable from other fruit in the bowl, and its yellow color should remain a reliable indicator of its edibility under almost any natural light. Ideally, the visual system could meet both goals by labeling each object with its surface spectral reflectance properties. These stay constant under changing illumination, scene composition and geometry, while the light that the object reflects to the eye varies. The fact that humans do show approximate color constancy implies that our visual system does attempt to recover a description of the invariant spectral reflectance properties of objects. The fact that color constancy is *only* approximate – colors of objects do change measurably under different light sources – implies that the human visual system does not meet its goals perfectly.

The computational problem that the human visual system must solve in order to attain color constancy, and the limitations that prevent it from attaining perfect color constancy, become clear when one considers the input data with which it is supplied. The human retina has three types of photoreceptor – the cones – responsive to medium and high intensities of light (the fourth type of photoreceptor – the rod – is responsive only at low light levels and is thus said to mediate "nighttime" vision). Each cone is selectively responsive to a different portion of the visible spectrum of light; the short-wavelength-sensitive (S) cone gives a peak response to light at 420 nm and essentially no response to light beyond approximately 550 nm; the middle-wavelength-sensitive (M) cone peaks at 534 nm and does not respond beyond approximately 650 nm; and the long-wavelength-sensitive (L) cone peaks at 564 nm and does not respond beyond approximately 700 nm [10]. Because there are only three types of cone, and because their spectral sensitivities are broad and overlapping, the human visual system cannot register exactly the spectral energy distribution of light incident on the retina. It records instead only the energy of light integrated over the spectral sensitivity range of each cone. Thus the input data is a triplet of cone responses, each response a measure of the incident light in either the long-, middle- or short-wavelength band of the spectrum, with one such triplet for each location in the image (in fact, the input

data is even further restricted by the fact that even in the fovea, the area of densest concentration of cones, the cones are not distributed uniformly enough to provide three types at each location). The light reflected from an object's surface into the retina is proportional to the product of the illumination incident on the surface and the invariant spectral reflectance properties of the surface. Thus the visual system cannot directly extract from each cone response a measure of the surface spectral reflectance in that wavelength band. The problem that the visual system faces is to transform the triplet of cone responses, which vary with the type of illumination on a given object's surface, into a set of responses that remain fixed for a given surface.[1] This set of values can at best approximate a measure of surface reflectance in each of several distinct wavelength bands, which may or may not be the same as the cones'.

Computational models of color vision, then, if they purport to explain human color vision, must provide methods for transforming cone responses into colors that are the same as those we perceive under all conditions. Some computational theories do attempt to reproduce, as closely as possible, known characteristics of human color vision. Others attempt only to construct algorithms that can achieve useful color vision for machines. These do not explicitly attempt to model human performance, although human performance is often the standard that vision machines strive to reach. Most computational models, though, do not fall neatly into one of these two types. This is perhaps because most assume that the goals of human and machine color vision are the same and that therefore a model that meets (or attempts to meet) that goal for a machine must be considered as an explanation of human color vision.

[1]Throughout this thesis, the terms "perception" and "sensation" will be used to distinguish between two levels of color vision. Color sensation refers to the triplet of cone responses, whereas color perception, or perceived color, refers to the transformed set of values that remain roughly constant for a given surface under changing illumination. Unless otherwise noted, the term color will refer to perceived color. Arend et.al. [4] demonstrate this distinction experimentally. Their results show that we can make a distinction between the color *sensation* of an object and the *perception* of its surface color or the judgment of the illuminant color. Their results suggest that we are aware of both the input and output of the color computation – that we can consciously identify both the input sensory stimulus (the retinal response to the physical stimulus) and the output surface color judgment.

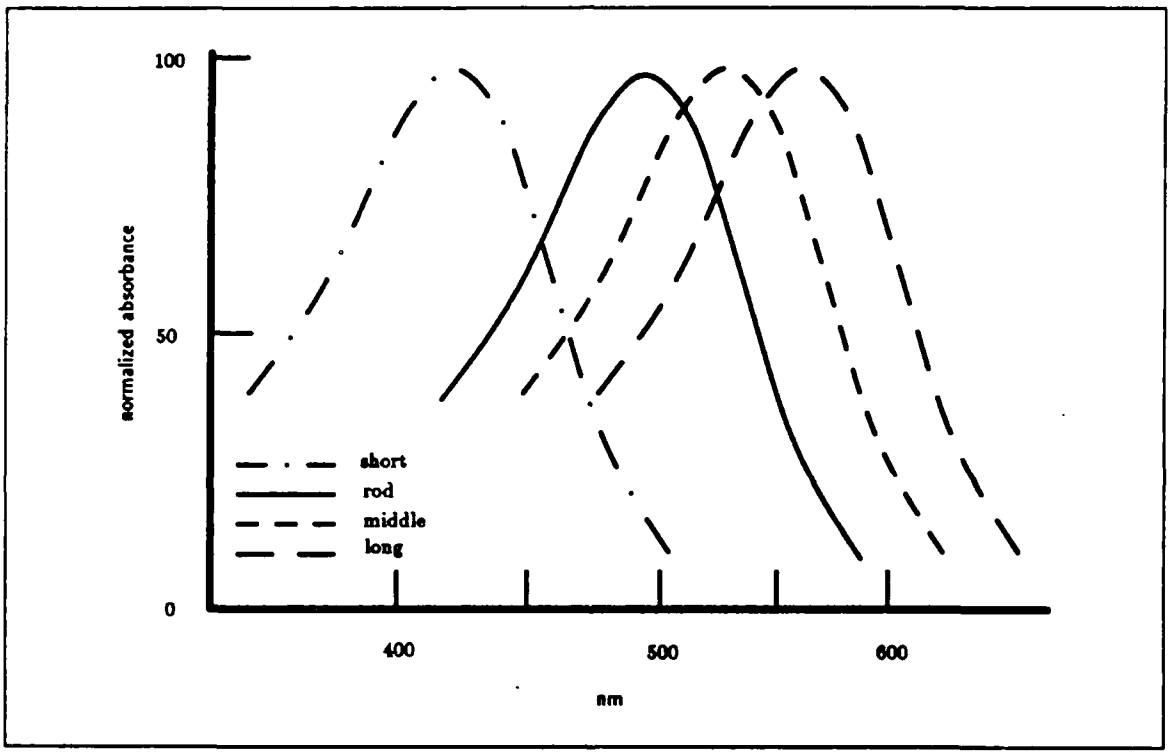

Figure 1.1: The human photoreceptor spectral sensitivity curves: the rod and the three types of cone: short-, middle- and long-wavelength-sensitive. Redrawn from Dartnall et. al. in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L.T. Sharpe, eds. (Academic, London, 1983).

In this thesis, the computational approach provides a framework for a study of color vision within which the techniques and results of several disciplines may be integrated. Three broad questions are addressed: (1) What perceptual features must a computational model reproduce if it is to elucidate the mechanisms underlying human color vision? (2) Do existing models accurately reproduce known features of color vision? and (3) How might the human visual system implement the steps required by existing models?

Chapter 2 addresses the first question by describing some well-known psychophysical results on color constancy, simultaneous color contrast, segmentation and filling-in. It briefly discusses the color vision deficits due to certain cerebral cortical lesions, which, together with other accompanying deficits, point toward the role that color perception plays in the whole of visual perception. The importance of color for object recognition, for example, is suggested by the poor performance on certain recognition tasks of some humans with cortical lesions.

Chapters 3 and 7 address the second question. Chapter 3 classifies and reviews existing computational models of color vision, in terms of the stated problems they attempt to solve, how well they solve them, and how well they simulate human color constancy. The classes of color algorithms introduced here are (1) *lightness algorithms*, which recover an estimate of surface reflectance in each of the three wavelength bands represented by the cone spectral sensitivities, for the restricted world of Mondrians (2) *spectral basis algorithms*, which exploit the redundancy in the responses of different cone types at different retinal locations to recover an estimate of surface reflectance in terms of a small number of basis functions that describe most naturally occurring reflectances (3) *segmentation algorithms*, which exploit interaction between cone responses across space to distinguish material boundaries from other causes of change in the image irradiance signal and (4) algorithms which use specular reflections as cues to the illuminant color, for example, the *chromaticity convergence algorithm*. It begins with a discussion of reflectance models (founded on the physics of light reflection and image formation), which provide a basis for the image irradiance equation. This

equation appears in various forms as the starting-point of all computational models. This section illustrates how the assumptions under which the equation is simplified for certain computational models affect their applicability for both human and machine vision. The chapter also draws new connections between existing models, makes new predictions based on assumptions underlying the models, and briefly reviews existing psychophysical results that bear on these predictions.

Chapter 7 describes a new psychophysical experiment that tests some of the methods described in Chapter 3 against human perception, in terms of how they directly or indirectly recover the color of the illuminant. Lightness algorithms that normalize to white correctly recover the color of the illuminant only when the reflectances of the brightest patches in each lightness channel are the same (when the "whites" in each channel are the same). Those that average to grey correctly recover the illuminant color when the average reflectance in each lightness channel is the same. Both methods are forms of Von Kries adaptation [112]. The chromaticity convergence algorithm and other similar algorithms correctly recover the illuminant color when there is sufficient information from specular reflections in the scene. The experiment reported here addresses the questions of how accurately humans perceive the color of the illuminant and whether human perception agrees with the predictions of the computational methods. Observers were asked to judge the illuminant color in computer-generated images of Mondrian-textured spheres with and without simulated specular reflections. The data suggest that the human visual system relies on a form of Von Kries adaptation even when specularities provide conflicting cues.

Chapter 3 also lays the foundation for the new computational models presented in Chapters 4, 5, and 6. Chapter 4[2] explores the formal connections between several distinct lightness algorithms. It establishes a common framework for lightness algorithms in the form of a single mathematical formula from which each algorithm may be derived under different conditions. It introduces a second framework for lightness algorithms based on regularization theory, from which a linear filter is derived that

[2]This chapter has been published in slightly altered form in Ref. [50].

embodies the constraints of lightness algorithms and transforms image irradiance (the data) into surface reflectance (the solution).

Chapter 5 builds on the regularization framework discussed in Chapter 4 to describe a new method for synthesizing lightness algorithms that can adapt to their 'environment'[3]. Here the assumption is made that a linear filter that transforms irradiance into reflectance exists, but no constraints on its form are imposed. Instead it is synthesized directly from pairs of input and output examples, and thus adapts to them as they change, 'discovering' the constraints employed by analytically-derived lightness algorithms. The examples used are one-dimensional Mondrian reflectance patterns under varying types of illumination gradients. The simplest linear synthesizing technique, optimal linear estimation, is compared with the nonlinear techniques of backpropagation (BP) and optimal polynomial estimation. Optimal linear estimation produces a lightness operator that is approximately equivalent to a center-surround, or bandpass, filter and which resembles the filter derived from regularization theory, which in turn resembles a new lightness algorithm recently proposed by Land. Requiring only one assumption, that the operator that transforms input into output is linear, the technique finds the linear estimator that best maps input into desired output, in the least squares sense. This assumption is true for a certain class of early vision algorithms that may therefore be synthesized in a similar way. Although BP performs slightly better on new input data than the estimated linear operator, the optimal polynomial operator of order two performs marginally better than both. The techniques developed here might be useful for other vision algorithms whose constraints might not be known in advance or for vision algorithms that must work in several different environments under different constraints.

The lightness operators synthesized in Chapter 5 (whether by optimal linear estimation, BP, or optimal polynomial estimation) accurately reproduce certain features of human lightness and color perception – lightness constancy under spatial changes in the illumination and simultaneous color contrast, for example – but fail to repro-

[3]This chapter has been published in slightly altered form in Refs. [51] and [52].

duce other features, in particular those that involve interaction between luminance and chrominance information, such as filling-in. This interaction is essential to the problem of image segmentation (how to segment an image into distinct materials) which is the first goal of color vision. Although humans effortlessly segment scenes into distinct objects, vision machines still can not. Chapter 6 presents a group of new segmentation algorithms, implemented on simulated neural networks, which mark boundaries between image regions of different surface reflectance and fill in uniform color labels. By analogy with the color filling-in phenomenon of human vision, the algorithms use luminance edges to help localize discontinuities in surface reflectance and guide the filling-in operators. They perform well on synthetic and natural images even in the presence of image irradiance changes due to shading, shadows and specularities.

Chapter 8 addresses the third question: how might the human visual system implement the algorithms proposed by computational models? It discusses what sort of cells would be required by certain lightness algorithms and discusses what computational uses known cells might serve. In particular, it poses the puzzle of single-opponent cells, which at first glance appear unsuited for any computational purpose, and outlines a new argument for their existence.

In summary, the primary contributions of this thesis are (1) the demonstration of a formal framework for lightness algorithms, which represent one class of solution to the problem of color constancy; (2) the derivation of a new lightness algorithm based on regularization theory; (3) the synthesis of an adaptive lightness algorithm using and comparing two "learning" techniques, optimal linear estimation and back-propagation; (4) the development of a segmentation algorithm that uses color information to mark material boundaries, with guidance from luminance edges; and (5) an experimental investigation into the cues that human observers use to judge the color of the illuminant, demonstrating that under certain conditions, observers ignore correct information from specular reflections in favor of incorrect information from other cues.

Chapter 2

Features of Color Vision

2.1 Color Constancy

A fundamental feature of human color vision is color constancy: our ability to assign roughly constant colors to objects even though the light they reflect to the eye varies over time, space, and spectrum. A red apple is red at morning or evening, under skylight or candlelight. The phenomenon of color constancy has long been appreciated by color vision scientists but still has not been explained by a complete model of color vision. In 1789, the French scientist Monge observed that red objects appeared white when viewed against a richly colored background through a piece of red glass, yet appeared red when observed individually through the same lens [84]. Helmholtz (1866) [108] concluded that the perception of color must involve an act of judgment, in which the effects of the illuminant were discounted. Helson and Judd (1936) [40] chronicled the subtle or striking changes in chromaticity of colored papers under different illuminants, calling the changes "color conversion". Buchsbaum (1980) [15], focussing on the strong shifts in colors of objects under certain illuminants, such as fluorescent ones, wrote that the phenomenon should be renamed "color inconstancy". Land (1971) [61] revived interest in color constancy with his elaborate "Mondrian" displays, reminiscent of Monge's simpler demonstrations. Today color constancy is studied intensively not only as a fundamental feature of human vision but also as the ultimate goal for image processors and intelligent machines, essential not only

for accurate image production in, for example, the photographic industry, but also for object recognition tasks in a range of applications from car manufacturing to surveillance.

2.1.1 The Mondrian Demonstrations: Spectral Normalization and Spatial Decomposition

Land aroused new interest in color constancy by performing striking experiments with two-dimensional surfaces covered with rectangular patches of random colors, named Mondrians after the twentieth-century Dutch painter Piet Mondrian. In one demonstration, Land illuminates a Mondrian with three differently-colored light beams, created by putting a different narrow-band filter on each of three light sources. He then adjusts the intensity of the sources so that the light reflected from patch A in the first viewing is the same in each of the three chromatic channels as the light reflected from patch B in the second viewing. Yet the perceived colors of the two patches are different from each other and stay roughly constant from one viewing to the next. Conversely, the color of patch A changes when the light reflected from it is held constant while the light from the surrounding patches is changed. From these results, Land draws the forceful conclusion that "the color of an object is not solely determined by the light it reflects" [61].

In another demonstration, Land adjusts an illumination gradient across a black-and-white Mondrian so that a dark grey patch at the top of the Mondrian reflects the same amount of light as a light grey patch at the bottom. Yet the top patch appears darker than the bottom patch, and the greys of both patches stay constant when the illumination is made uniform across the Mondrian. Other results [80, 33] indicate that we do not see slow illumination gradients[1] and that we interpret the slow irradiance gradients that we do see as due to the illumination rather than the

[1]The visibility of a linear gradient of illumination appears to depend only on its contrast ($L_{max} - L_{min}/L_{max}$ where L_{max} is the maximum illumination value and L_{min} the minimum) and not on the retinal gradient it produces (contrast divided by the retinal angle subtended by the illumination gradient). Threshold visibility is about 10%. [80, 33]

surface reflectance properties of the object [3].

Land's Mondrian demonstrations illustrate two features of color constancy. The first is *spectral normalization*, our ability to compensate for temporal changes in the color of the illuminant. The second is *spatial decomposition*, our ability to compensate for spatial changes in the illuminant energy.

2.2 Simultaneous Color Contrast: A Close Cousin of Constancy

Land's Mondrian demonstrations imply that in computing constant colors we can compensate for changing illumination but not for changing backgrounds. In Land's demonstrations, when the illuminant color changes, the light that a patch reflects changes in the same way as the light that its surround reflects. The ratio of patch to surround reflection does not change and that ratio, computed in a particular way, correlates with constant colors. Conversely, when the light from the surround changes, the ratio changes, and the patch color changes. Explained in this way, color constancy implies color contrast. [2]

Simultaneous brightness contrast (or simultaneous darkness induction, as the phenomenon is more precisely termed by Creutzfeldt et. al. [18]) is the tendency of objects to look darker when against a bright background than when against a dark background. Simultaneous color contrast is the tendency of objects to change color when the background color changes. Although brightness and color contrast, like color constancy, have been studied for well over a hundred years, their relationships to each other and to color constancy are not fully understood.

The Gestalt psychologist Koffka [54] argued that color constancy and color contrast produced opposing effects, using the following experiment. An object casts a

[2]Yet Land's original model of color computation – the original Retinex – assumed that the environments to which it applied were sufficiently Mondrian-like that the surrounds of all objects or patches would average to roughly the same color. Thus his model did not predict color contrast effects. See the discussion of the grey-world assumption in Chapter 3.

shadow on a piece of white paper, with the whole scene illuminated by yellow light. The paper, which reflects yellow light, looks white and the shadow, which reflects grey light, looks blue. If yellow paper is now put in place of the white paper, but without covering the shadowed area, the paper looks yellow and the shadow looks grey. Koffka argued that color contrast would tend to make the shadow look even more blue when its surroundings look yellow, but that color constancy, by normalizing the yellow illuminant color to white, only makes the shadow less yellow than its surroundings.

On the other hand, Koffka argued that color contrast and color constancy could not be distinct phenomena, but that both must involve central, rather than simply retinal, mechanisms. Color contrast, for example, did not depend just on the absolute reflectances and shapes of objects, but rather on the "forces of unit formation" (see the description of the Koffka Ring, section 2.3). Color constancy also depended on higher-order mechanisms that first segregated scenes into different objects using, for example, depth information before assigning colors.

Lettvin [14, 65] emphasizes that the color of an object depends on the colors of objects in its immediate surround, and that therefore the constancy of the colors in Land's Mondrians is an exceptional result due to their richly random arrangement. In [14], Lettvin and his colleagues create a CRT display of a regular array of hexagons of random colors. Five grey hexagons sit equidistant from each other, one close to each of the four corners of the display, and one in the center. When the hexagons are put in order according to color, so that, for example, red intensity values increase from left to right, and green values increase from right to left, the fixed grey hexagons no longer look grey, but take on different colors depending on the colors of the hexagons in their immediate surrounds. Thus, the hexagon-Mondrian demonstrations imply that the color of an object depends on a weighted local average of the colors in its surround, not on an unweighted global average.

Shapley (pers. comm.), following the tradition established by Hering [41], argues that lightness and color constancy depend solely on local contrast mechanisms.

Shapley claims that retinal lateral inhibition and gain control mechanisms mediate the calculation of the average local contrast at the border of an object and its background, and that the average local contrast determines lightness. Gilchrist et. al. [34] and Arend [4] argue on the basis of convincing psychophysical results that local contrast effects are not nearly strong enough to account for color constancy.

These results, together with many others [38], illustrate that although there is almost certainly an underlying connection between color contrast and color constancy, its details are still unknown.

2.2.1 Segmentation and Filling-In

Koffka [54] emphasized that simultaneous color contrast depends not merely on the absolute reflectances and shapes of parts of an image but also on the way the parts are segmented into wholes. The Koffka Ring is an image of a uniform grey annulus placed against a bipartite rectangular background, one half red, the other half green. When the midline dividing the background is drawn across the annulus, the annulus splits into two pieces. The semi-annulus on the red half appears greenish, and the semi-annulus on green appears reddish. If the midline is erased, the annulus appears uniformly grey once more. The effect is stronger for an achromatic ring. A grey annulus against a rectangular background, one half black and one half white (see Figure 2.1), more strikingly splits into two semi-annuli, one light and one dark, when the midline is drawn across it (see Figure 2.2).

The Koffka Ring shows that the presence of a luminance edge between regions influences the perception of color or lightness within the regions. Koffka interpreted his illusion as evidence for a Gestaltist theory of form perception: the "forces of unit formation" are strong enough to hold the annulus together when it is not split by the midline, but not when it is. In more modern terms, the Koffka Ring might be said to show that surfaces are segmented into regions by luminance edges before their attributes of lightness or color are computed. Recent experiments by Gilchrist [33] show that lightness is computed separately for surfaces in separate depth planes,

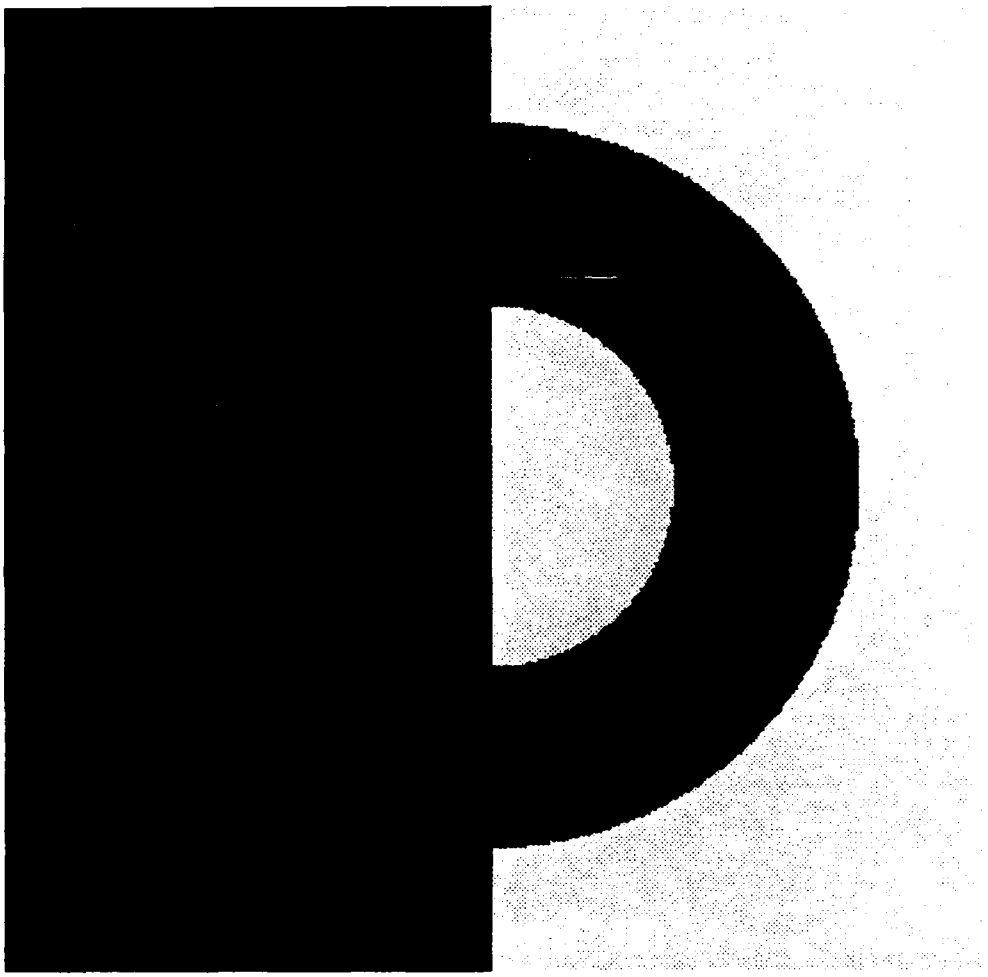

Figure 2.1: The Koffka Ring: a uniform grey annulus against bipartite background, here with no midline drawn through the annulus. Annulus appears to be uniform grey throughout.

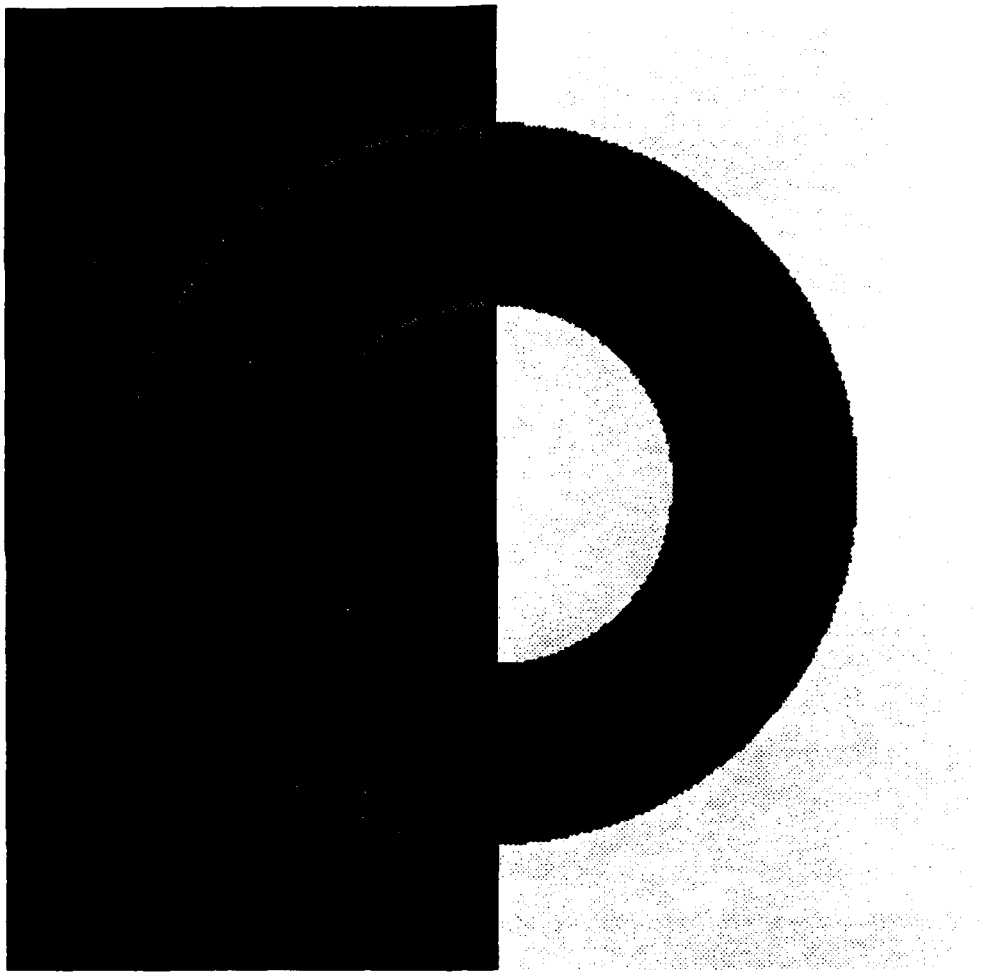

Figure 2.2: With the midline drawn down the center of the annulus, the left half appears lighter than the right.

supporting the idea that segmentation of the image based on form and depth cues precedes lightness or color computation and therefore precedes segmentation based on lightness or color cues.

Yarbus [113] illustrated that in some sense, lightness or color is computed at edges and spreads across, or "fills in", regions within edges. In his classic stabilized-edge experiment, two red disks are set against a bipartite background, the left-hand disk on white, the right-hand disk on black. When the edge between the backgrounds is stabilized, and therefore "disappears," the left-hand disk turns dark red, the right-hand disk light red. Crane and Piantinada [17] demonstrate a similar filling-in effect that results in the striking perception of reddish green.

Gilchrist, et. al. [34] demonstrate two crucial features of the importance of edges for lightness perception: (1) distant edges, not just those at the boundary of the region in question, influence the region's lightness; and (2) reflectance and illumination edges influence lightness perception in different ways. Gilchrist et. al. create a display which, under one condition, appears to be the traditional simultaneous contrast display: a bipartite rectangular background, the left half white, the right half black, with two grey square targets, one on each half. Under this condition, the edge between the two halves of the background appears to be the edge between two papers of different reflectances, and the two targets look grey, the left one slightly darker than the right. Under another condition, the edge between the two background halves appears to be the edge between the lit and shadowed halves of a single paper with a uniform reflectance. The left target then looks black and the right looks white, although the luminance ratios across all the edges in the display have not changed.

The Gilchrist et. al. experiment illustrates the strength of the spatial decomposition computation in color constancy. Not only slow gradients but also sharp changes of illumination can be disentangled from reflectance changes. The results also emphasize a crucial distinction between color contrast and color constancy: the former is predominantly a local phenomenon, whereas the latter is global. Blackwell and Buchsbaum [6], for example, demonstrate that regions greater than two degrees of

visual field distant from a central patch do not induce contrast effects there.

2.3 Color Vision from a Neurological Perspective

One way to discover how, where and why the human brain produces this variety of phenomena in seeing color is to examine what goes wrong when the brain's capacity to see color is impaired. In experimental biology, lesion studies are often fruitful. That is, often the quickest, although perhaps not the most elegant, way to determine the function of an identifiable structure is to determine which functions are lost when that structure is ablated. In humans, brain lesions are not normally made deliberately to determine function. Yet they do occur often enough to provide data for natural lesion studies, as the result of natural accidents or as inadvertent byproducts of neurosurgical attempts to control neurological disease, as in the famous case of the epileptic H.M. whose amnesia following bilateral temporal lobectomy revealed that certain temporal lobe structures were essential to memory.

The human color vision pathway extends from the retina, where the three types of cone respond differentially to different wavelengths of light, to visual association cortex. Lesions anywhere along this pathway may cause disturbances of color vision. The most commonly known is retinal color blindness, in which either the long- or middle- wavelength-sensitive type of cone is missing (dichromatic vision) or its spectral sensitivity is displaced from normal (anomalous trichromatic vision). The post-receptoral color vision pathway in the retinally color-blind is generally thought to be intact although it has not been extensively studied psychophysically in humans or physiologically in animal models (see, e.g., [86]). The fact that the retinally color-blind do not have blatant deficits in the phenomena of color vision discussed above (e.g., color constancy, color contrast, and image segmentation using color) within the limits defined by their wavelength discrimination abilities, suggests that these phenomena are indeed mediated by higher-level cortical mechanisms.

Lesion studies have not yet proved this conjecture because they have yet to prove

conclusively that a cortical center specialized in any way for color can be segregated from other regions of visual cortex. That is, lesion studies have first to establish the separability of the color vision pathway from other visual pathways at the cortical level before attempting to isolate stages in color processing. The problems in doing the former are at least two: the color vision pathway from retina to cortex appears particularly vulnerable to damage, so that color vision deficits occur most often in conjunction with other visual deficits [81]; and tightly circumscribed cerebral lesions are rare. In the case of cerebrovascular accidents, it is almost too much to ask that if an anatomically distinct cerebral color center exists it will be supplied by a branch of an artery that may occlude or rupture causing complete damage to the center without affecting supply to functionally distinct regions. Largely for these reasons, evidence for a cortical color center from lesion studies has been controversial. Now, though, based largely on physiological data from monkeys (see Chapter 8), it is widely accepted that at least for several stages of visual processing color information is processed separately from other types of visual information.

Meadows [81] reviews 14 cases of achromatopsia due to cerebral lesions, the earliest being MacKay and Dunlop's of 1899 [69], the latest one of his own (1974). All but one of the patients had suffered a cerebrovascular accident; one had undergone trauma to the occipital region. All complained of a complete or near-complete loss of color vision, most describing the world as being "black and white" or "grey." The majority of patients had uni- or bilateral upper quadrant visual field deficits, suggesting that lesions of the inferior striate cortex or optic radiation were responsible. In three patients on whom autopsies were done, lesions were found in the fusiform (now called 'occipitotemporal') and/or lingual gyri. This led neurologists of the late nineteenth century to conclude that there is a human colour center and that it lies in these gyri.

Most of these patients did not undergo detailed testing of visual function. Thus it is not possible to determine exactly which capacities are subserved by the lesioned areas or which capacities dependent on intact color vision were compromised. Twelve out of 14 of the patients also suffered prosopagnosia. But from this, one cannot

conclude, attractive as the hypothesis might be, that color vision is a prerequisite for face recognition. The more likely conclusion is that the putative color center destroyed by infarction of the lingual and fusiform gyri lies close to a distinct area that is crucial in face recognition and that is also damaged. As both Meadows [81] and Damasio et. al. [20] note, this is supported by the fact that prosopagnosia and achromatopsia may occur independently of each other. It is nonetheless interesting to note that the painter J. I. (discussed below) found that faces were "unidentifiable until they were close" because they had "lost color and tonal contrast" [100]. His complaints illustrate that although color may not be the essential cue for object recognition, it is an important one.

More recently, Damasio et. al. [20] have reported a case of pure hemiachromatopsia, which provides the strongest support for the existence of an anatomically distinct area specialized for color computation. This patient (C.M.) suffered a cerebrovascular accident that selectively lesioned the lingual and fusiform gyri in his right hemisphere, leaving him unable to perceive colors in his left visual field. Except for a small scotoma in his left upper quadrant, he had no other visual deficits, in particular no prosopagnosia and no loss of visual acuity.

Damasio et. al. [20], together with other authors, speculate that this lesion-defined color center may be the human analogue of the monkey area V4, subregions of which contain cells selectively responsive to color (see Chapter 8). Yet as both Damasio et. al. and Meadows [81] point out, V4 is differently located in the monkey. Monkey V4 is on the lateral surface of the occipital lobe in the banks of the lunate sulcus, superior and lateral to the lingual and fusiform gyri. But since the primary visual cortex in humans is medial relative to monkey V1, it may be that the analogue of V4 is similarly translated.

Heywood et. al. [42] emphasize that monkey V4 appears to be responsible for aspects of form as well as color perception. They report a case of partial achromatopsia in a patient (C. B.) who suffered occipital lobe damage in an auto accident. C.B.'s color perception is "grossly impaired," as evidenced by grossly abnormal scores on

the Farnsworth-Munsell 100-Hue test, although not completely destroyed. (He can discriminate between widely but not narrowly separated hues.) As in Damasio's patient, C.B.'s form perception, as evidenced by normal performance on shape and line orientation discrimination tests, is intact. Each patient's pattern of deficits supports the notion of separability of form and color perception at a high level in the visual system. Yet because the site of C.B.'s lesion is not precisely known, it is difficult to draw specific conclusions about the human analogue of V4.

Although almost all achromatopsic patients complain that the normally vivid colors of objects are usurped by black, white and shades of grey, which colors turn white and which turn black vary from case to case. Sacks et. al. report the case of the "color-blind painter," a successful abstract artist (J. I.) who after suffering a concussion in a minor car accident lost all color vision. To him, objects suddenly seemed not only grey but also "disgusting." Curiously, though, while saturated reds and greens appeared black, saturated blues appeared pale, almost white. This is odd, because if the luminance pathways remained perfectly intact greens and reds should have been lighter than blues, since the short-wavelength-sensitive cone contributes little to the luminance channel. In other respects, J. I.'s luminance channel appeared relatively intact, in that he ordered other colors on a luminance scale that was accurate but much more coarsely divided than normal. This anomaly suggests that the luminance channel interacts with the chrominance channels beyond the retina, and that determinations of brightness and whiteness hinge on high-level cortical mechanisms. Perhaps in J.I. the blue-yellow chrominance channel was less or more damaged than the red-green channel. One must postulate some such differential damage to explain why other achromatopsics see different blacks and white. Like J. I., two of the patients Meadows reviews perceived "relatively unsaturated" blues and greens as "bright," and dark browns, reds and blues as black. Unlike J.I., one saw blues as dark and reds as light. Another saw blues and greens as brown or grey.

For at least some achromatopsics, the loss of color vision is not merely an inconvenience, a minor deprivation in an otherwise luxurious visual world. For some

patients, it seriously hampers object recognition, at least for those objects, as Pallis's patient explained, for which "colour [was] the main clue as to what it [was]" [89]. This patient said "... when I open a jar I never know if I'll find jam or pickles in it." Similarly, J.I. found that he would mistake "mustard for mayonnaise" or "ketchup for jam" [100].

Perhaps the most interesting and simultaneously the most anecdotal lesion studies bearing on the why and how of cortical color mechanisms are those described by Critchley [19]. He relates provocative reports that support the notion that the filling-in of color occurs separately from the determination of luminance and depth boundaries. He describes the phenomenon of "illusory visual spread" that results from some (unspecified) cerebral diseases. ".. Colours may seem to extend far beyond the confines of the object to which they belong, especially when a bold pattern of vivid contrasting colours is concerned. Thus the design of a frock may appear to spread over the face and bare arms of the wearer" [19]. He describes another, similar phenomenon in which colours stray outside normal object boundaries, sometimes settling in a different depth plane from the object to which they normally belong. "As colours no longer appear to be integral to the objects in question, they may seem to occupy a plane somewhere between the subject and the object. When the patient wishes to touch something, he has an odd impression as though he had to plunge his hand through a translucent sheet" [19]. Critchley writes, "Hoff ... has stated that with occipital lobe stimulation, colours may be seem to be 'loosened' from the confines of their objects so as to constitute a 'separation of colour from the object,' sometimes in the form of coloured balls [19]." These reports suggest that color cannot be the only cue used to segment an image into distinct materials, as objects still appear distinct even when their colors are detached.

So far, there has been no report of a neurological deficit in which hue discrimination is preserved but color constancy is lost, nor of a psychophysical study of an achromatopsic in which, for example, quantitative measurements are made of contrast sensitivity functions at varying temporal frequencies and with varying amounts

of chrominance information. The latter study might reveal a great deal about the mechanisms in the lesioned areas and whether they have the characteristics of the "parvo" pathway described in monkeys (see e.g. [25, 26, 68]). Nonetheless, lesion studies do provide both crude support for the modularity of human vision, and a glimpse of the way color vision interacts with other modules.

Chapter 3

Computational Models of Color Vision

The computation of color is a prime example of the difficult problems of inverse optics. Inverse optics is the science of recovering unique three-dimensional scenes from ambiguous two-dimensional images. In the computation of color, the goal is to recover the invariant spectral reflectance properties of object surfaces from the image irradiance signal. In the image irradiance equation, both the spatial and spectral variations of surface spectral reflectance are entangled with those of the surface illumination. Just as one spatial dimension is lost in the projection of a three-dimensional scene onto a two-dimensional imaging plane, information about the surface reflectance is irrevocably lost in the ambiguous image irradiance signal. Most of the computational approaches to the problem discussed here simplify the image irradiance equation by assuming that all surfaces are perfectly matte and that each is illuminated by only one light source. These assumptions reduce the image irradiance equation to the product of two terms: surface reflectance and illumination. The goal of most color algorithms is to separate these two terms.

The problem of separating surface reflectance from illumination itself has two parts: spatial decomposition and spectral normalization. The first is the problem of disentangling the spatial variations of the two components; the second, that of disentangling their spectral variations. *Lightness* algorithms specifically address the

first problem by exploiting natural constraints on the spatial form of reflectance and illumination. The lightness algorithms and *segmentation* algorithms discussed here employ one of two methods to perform spectral normalization: averaging to grey or normalizing to white. *Spectral basis* algorithms rely on similar methods. The differences between algorithms lie largely in the assumptions they make about the physical properties of surfaces and illuminants and in the assumptions they rely on to achieve spectral normalization.

The simplified image irradiance equation introduced here is not adequate for many materials in the natural world, which reflect light with significant *specular* components. The first two sections of this chapter discuss how the simplified image irradiance equation may be derived from its more complicated correct form. The more complicated form is used as the basis for the class of algorithms that achieve spectral normalization by recovering the color of the illuminant using information from specular reflections, characterized by the *chromaticity convergence* algorithm.

3.1 The Image Irradiance Equation

The starting-point for computational models of color vision is the image irradiance equation. This describes the relationship between the light falling on the image plane (the *image irradiance*) and the physical properties of the scene being imaged. In general, the image irradiance due to one surface in the scene is proportional to the product of the amount of light falling on the surface (the *surface irradiance* or *illumination*) and the proportion of incident light that the surface reflects (the *surface reflectance*).[1] The *surface spectral reflectance* is the surface reflectance as a function of wavelength. This is an invariant property of a surface, dependent only on the material from which it is made; it stays constant under changes in illumination. Thus, in order to compute constant colors, some measure of surface spectral reflectance must be

[1]Image irradiance (the light falling *onto* the image) is proportional to surface radiance (the light exiting *from* the surface, by factors dependent on the imaging geometry, as discussed in slightly more detail in the following section. Irradiance is specified in units of Watts per meter-squared (W/m^2). Radiance is specified in units of Watts per steradian per meter-squared.

extracted from the image irradiance. The fact that our perception of constant colors correlates strongly with certain measures of surface spectral reflectance suggests that the human visual system performs such a computation.

Most computational models use a simplified form of the image irradiance equation that relies on several restrictive assumptions about the world being imaged. The equation may be written:

$$I(\lambda, \mathbf{r}_i) = R(\lambda, \mathbf{r}_s, \mathbf{v}, \mathbf{n}, \mathbf{s}) E^*(\lambda, \mathbf{n}, \mathbf{s}), \tag{3.1}$$

where λ is wavelength; $\mathbf{r}_i$ is the spatial coordinate in the image onto which $\mathbf{r}_s$, the surface coordinate, projects; $\mathbf{n}$ is the unit vector in the direction of the surface normal at point $\mathbf{r}_s$; $\mathbf{v}$ is the unit vector in the direction of the viewer; and $\mathbf{s}$ is the unit vector in the direction of the light source. $E^*(\lambda, \mathbf{n}, \mathbf{s})$ is the surface illumination[2] on point $\mathbf{r}_s$, and $R(\lambda, \mathbf{r}_s, \mathbf{v}, \mathbf{n}, \mathbf{s})$ is the surface reflectivity function[46, 45] (see Figure 3.1).

This equation assumes that the way in which surfaces reflect light is governed by a single reflectivity function. In particular, most computational models further simplify the equation by separating the reflectivity function into two independent spatial and spectral terms:

$$R(\lambda, \mathbf{r}_s, \mathbf{v}, \mathbf{n}, \mathbf{s}) = \rho(\lambda, \mathbf{r}_s) F(\mathbf{v}, \mathbf{n}, \mathbf{s})$$

where $\rho(\lambda, \mathbf{r}_s)$, here called *surface reflectance* or *reflectance*, is the component of the reflectivity function that depends only on the material properties of the surface[3]; and $F(\mathbf{v}, \mathbf{n}, \mathbf{s})$ is the component that depends on the viewing geometry. $\rho(\lambda, \mathbf{r}_s)$ is written as a function of the spatial coordinate of the surface to illustrate that the material of which it is made may vary across space.

This separation is valid for surfaces that reflect light equally in all directions, that is, for perfectly matte surfaces. Such surfaces are said to be modified *Lambertian*

[2]The surface illumination, in general, includes the direct source illumination and the diffusely reflected illumination from all surfaces in the scene. Here the mutual surface reflections will be ignored.

[3]Often $\rho(\lambda, \mathbf{r}_s)$ is called the *albedo*.

surfaces. (Ideal Lambertian reflection is not only diffuse but also white.) The light reflected from any one point on a Lambertian surface has the same spectral composition and radiance in each direction. This implies that $F(\mathbf{v}, \mathbf{n}, \mathbf{s})$ is a constant, or, in other words, that the surface reflectance function is independent of the angle at which light enters or exits the surface. That is, the surface reflectance depends solely on the material of which the surface is made and not on its orientation, shape, or position relative to the light source.

For many materials, $F(\mathbf{v}, \mathbf{n}, \mathbf{s})$ is not constant but varies with $\mathbf{v}$ and $\mathbf{n}$, as discussed in Section 3.2. For these materials, the factors in Equation 3.1 may be regrouped by defining the *effective irradiance* $E(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s})$, where $E(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s}) = F(\mathbf{v}, \mathbf{n}, \mathbf{s})E^*(\lambda, \mathbf{n}, \mathbf{s})$. The effective irradiance is the surface irradiance modified by the orientation, shape, and location of the reflecting surface and by the geometry of the imaging system.

For many types of materials, then, Equation 3.1 may be simplified to:

$$I(\lambda, \mathbf{r}_i) = \rho(\lambda, \mathbf{r}_s)E(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s}). \tag{3.2}$$

The effective irradiance is the intensity that the surface would reflect if it were painted perfect white, that is, if $\rho(\lambda, \mathbf{r}_s) = 1$. If the surface were also normal to both the source and viewer direction, that is, if $F(\mathbf{v}, \mathbf{n}, \mathbf{s}) = 1$, then the effective irradiance would equal the surface irradiance. If the surface irradiance only were white and $F(\mathbf{v}, \mathbf{n}, \mathbf{s}) = 1$, then the reflected radiance would be directly proportional to the surface reflectance everywhere by the same factor.

In most biological and artificial visual systems, the image irradiance is captured by an array of light sensors. The signal transmitted by a sensor is a function of the light intensity integrated over the sensor's spectral sensitivity range. For the cones in the human retina, this function is nonlinear and is commonly approximated by a logarithm as shown: [4]

[4] Baylor, et.al [5] have shown that although the cone response is logarithmic with respect to large changes in retinal irradiance, it is linear within small operating ranges.

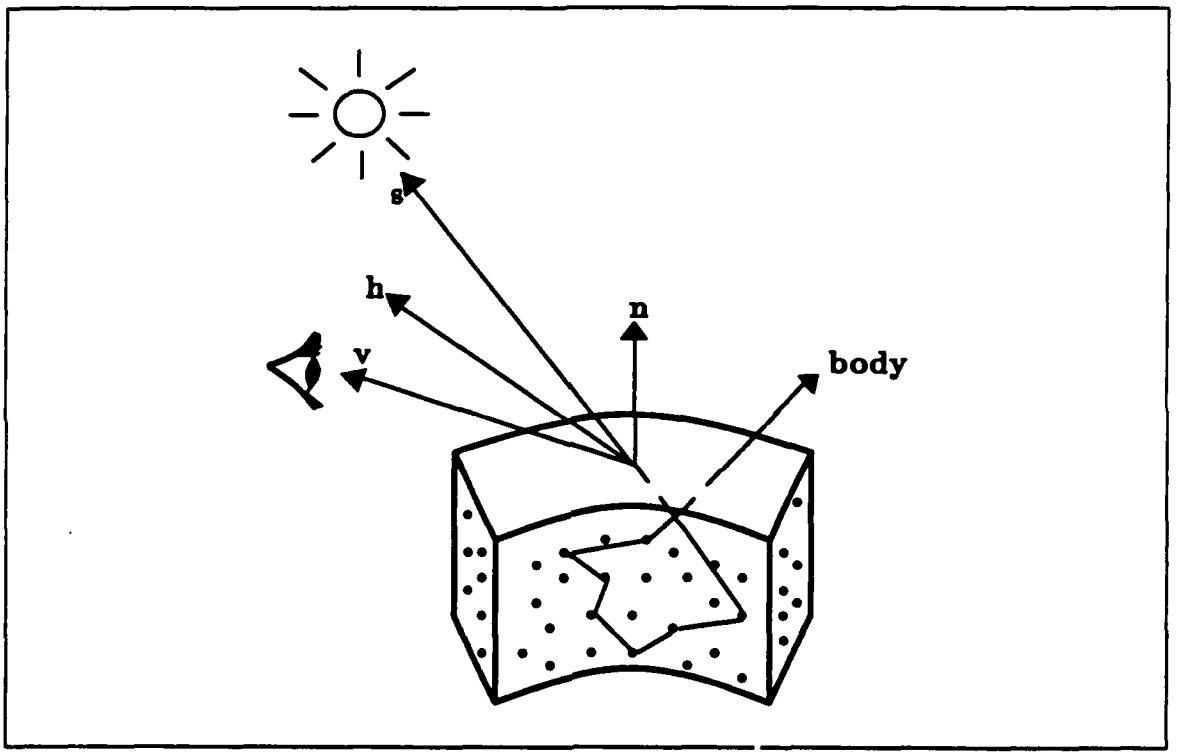

Figure 3.1: The imaging geometry.

$$S^{\nu}(\mathbf{r}_i) \;=\; \log \int a^{\nu}(\lambda)\rho(\lambda, \mathbf{r}_s)E(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s})d\lambda, \tag{3.3}$$

where ν labels the type of receptor (the long-, middle- or short- wavelength-sensitive cone; see Figure 1.1), $S^{\nu}(\mathbf{r}_i)$ is the signal sent by the sensor and $a^{\nu}(\lambda)$ is its spectral sensitivity function.

From Equation 3.3, it is apparent that the computational problem of color constancy, to extract surface spectral reflectance from image irradiance, is underconstrained. At every location in the image, there are two unknown variables – surface reflectance and illumination – and only one known – the image irradiance. To solve it, one must find constraints on the behavior of the reflectance and illumination that enable them to be disentangled.

The problem may be stated in a more general way for the human visual system, without assuming that its goal is explicitly to recover surface spectral reflectance. The problem is: transform the triplet of cone activities S^{ν} at each location on the retina into a triplet of values C^{k} that remain constant for a given surface under changes in illumination.

3.2 Reflectance Models

Equation 3.2 makes the implicit assumption that a surface reflects only one type of light. In fact, the light that a surface reflects is, in general, composed of two types: *body reflection*, light whose color is determined solely by the material properties of the surface; and *interface reflection*, which for certain materials, takes on the illuminant color (as termed by Shafer [102]). Body reflection, as its name implies, emerges from the body of the material and is diffuse in direction. Interface reflection arises from the air-surface interface and, for smooth surfaces, is concentrated in a single direction. (For smooth surfaces, the interface reflection is called *specular*.) Each type of reflection is governed by a distinct reflectivity function. In other words, the term $\rho(\lambda, \mathbf{r})F(\mathbf{v}, \mathbf{n}, \mathbf{s})$ in equation 3.1 should strictly be written as the sum of two terms,

each with a separate dependence on wavelength and direction. Equation 3.2, which serves as the starting-point for many of the computational models reviewed here, is based on a modified Lambertian reflectance model in which the interface reflection term is neglected; that is, surfaces are assumed to be perfectly diffusely reflecting, with no specularities.

The Lambertian reflectance model vastly simplifies reflection in the real world, which consists of a variety of materials, illuminants and geometries. Several authors (Healey and Binford [39], Lee [63, 62], and Shafer [102]) make a useful distinction between the reflectance properties of two types of material. An optically homogeneous material, such as a metal or crystal, consists of a single opaque material. The light it reflects is dominated by interface reflection. The energy of the reflected light is a function both of the complex refractive index of the material, which is itself a function of wavelength, and of the imaging geometry. Thus, for homogeneous materials, the color of the interface reflection is not the same as the illuminant color, nor is it constant under changing illumination conditions.

An optically inhomogeneous material, such as a paint or plastic, consists of a carrier material, which is largely transparent, and of pigment particles imbedded in the carrier. The light it reflects is dominated by body reflection, which emerges in all directions after being refracted into the carrier, where it is scattered and absorbed by the pigment particles. For inhomogeneous materials, the index of refraction of the carrier is virtually constant with respect to wavelength, so the interface reflection takes on the color of the illuminant. The color of the body reflection, on the other hand, is determined solely by the internal pigment particles – thus, it is to the body reflection that we attribute the constant color of the object.

Many computer graphics models assume a reflectance model based on the properties of optically inhomogeneous materials (see [62]): diffuse reflections are assigned the color of the object, while specular reflections take on the color of the light source. Lee et. al. [62] term this the *neutral interface reflection* (NIR) model.

Lee et. al.[62] derive a more exact form of equation 3.2 for inhomogeneous materi-

als, roughly as follows. The precise equation relating the surface irradiance $E(\lambda, \mathbf{n}, \mathbf{s})$ to the image irradiance $I(\lambda, \mathbf{r}_i)$ is:

$$I(\lambda, \mathbf{r}_i) = \alpha \int_{\omega_s} f(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s}) E(\lambda, \mathbf{n}, \mathbf{s}) cos\theta_s d\omega_s \tag{3.4}$$

where θ_s is the angle between the surface normal $\mathbf{n}$ and the source direction $\mathbf{s}$, and ω_s is the solid angle in the direction $\mathbf{s}$. $f(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s})$ is the bidirectional spectral reflectance distribution function. Formally, f is defined as the ratio of the reflected radiance in the viewing direction $\mathbf{v}$ to the incident irradiance in the source direction $\mathbf{s}$. Integrating the product of f over all solid angles of incidence of the source gives the total reflected radiance of the surface in the direction $\mathbf{v}$. The factor α governs the conversion of surface radiance to image irradiance. It is dependent on parameters of the imaging system: the focal length, diameter of the entrance pupil, and the off-axis angle, the latter of which varies with position in the image.

Under the assumption that there is a single light source (see Section 6.1), the surface irradiance E may be separated into the product of two terms:

$$E(\lambda, \mathbf{n}, \mathbf{s}) = E(\lambda)\epsilon(\mathbf{n}, \mathbf{s}). \tag{3.5}$$

That is, the single source assumption says that the spatial dependence of the surface irradiance is the same for each wavelength. Under the assumption that the pigment particles are uniformly distributed throughout the carrier material and none protrude through the surface [102], the function f may also be decomposed into distinct spectral and spatial terms:

$$f(\lambda, \mathbf{v}, \mathbf{n}, \mathbf{s}) = \rho_b(\lambda) g(\mathbf{v}, \mathbf{n}, \mathbf{s}) + \rho_s(\lambda) h(\mathbf{v}, \mathbf{n}, \mathbf{s}). \tag{3.6}$$

Equation 3.4 then becomes:

$$I(\lambda, \mathbf{r}_i) =$$

$$\alpha E(\lambda)\left[\rho_b(\lambda)\int_{\omega_s} g(\mathbf{v},\mathbf{n},\mathbf{s})F(\mathbf{n},\mathbf{s})cos\theta_s d\omega_s \; + \rho_s(\lambda)\int_{\omega_s} h(\mathbf{v},\mathbf{n},\mathbf{s})F(\mathbf{n},\mathbf{s})cos\theta_s d\omega_s\right]. \tag{3.7}$$

Setting

$$F_b(\mathbf{v},\mathbf{n},\mathbf{s}) \; = \alpha\int_{\omega_s} g(\mathbf{v},\mathbf{n},\mathbf{s})\epsilon(\mathbf{n},\mathbf{s})cos\theta_s d\omega_s$$

and

$$F_s(\mathbf{v},\mathbf{n},\mathbf{s}) \; = \alpha\int_{\omega_s} h(\mathbf{v},\mathbf{n},\mathbf{s})\epsilon(\mathbf{n},\mathbf{s})cos\theta_s d\omega_s,$$

equation 3.7 simplifies to the form of equation 3.2, with the two components of the reflectivity function made explicit:

$$I(\lambda,\mathbf{r}_i) \; = E(\lambda)\left[\rho_b(\lambda,\mathbf{r}_s)F_b(\mathbf{v},\mathbf{n},\mathbf{s}) + \rho_s(\lambda,\mathbf{r}_s)F_s(\mathbf{v},\mathbf{n},\mathbf{s})\right]. \tag{3.8}$$

Each component of the image irradiance may be split into two terms: an effective illumination term (which absorbs different geometric factors for each type of reflection) and a reflectance term. Here ρ_b is the body spectral reflectance factor (*surface reflectance* or *reflectance*, as defined earlier) and ρ_s the specular spectral reflectance factor.[5] (The reflectance factor technically is defined as the ratio of reflected light in a given direction to that reflected in the same direction by a perfectly reflecting diffuser identically illuminated. It differs from the reflectance factor as defined here by a factor of π.)

Healey and Binford [39] illustrate how the NIR model may be derived from Equation 3.8 for certain types of inhomogeneous materials. For these materials, the refractive index is constant across the spectrum,

$$\rho_s(\lambda) = k_s \tag{3.9}$$

or, in other words, the specular reflection is neutral, constant with respect to wavelength. For materials with perfectly smooth surfaces,

[5]The precise form of g depends not only on the vectors $\mathbf{n},\mathbf{s},\mathbf{v}$, and $\mathbf{h}$, but also on the fraction of microfacets at point p_n oriented in the direction of $\mathbf{h}$ and on a geometrical attenuation factor accounting for masking between facets.

$$F_s(\mathbf{v}, \mathbf{n}, \mathbf{s}) = F_s(\mathbf{s}). \tag{3.10}$$

That is, the interface reflection is concentrated in the perfect specular direction, which depends only on **s**. Thus, for a smooth-surfaced inhomogeneous material, the surface reflection will be concentrated in one direction and will take on the illuminant color.

The body reflectance term also simplifies under certain assumptions. Healey and Binford [39] use the Kubelka-Munk theory for reflection from inhomogeneous materials to show that body reflection is roughly constant with respect to the source direction, and that therefore F_b is in general a function only of **v** and **n**. (This result also justifies the separation of the body reflection into spatial and spectral terms.) For an ideally diffusely reflecting surface, F_b is constant with respect to **v**, **s** and **n**.

Lee et. al. [62] tested the validity of these assumptions for certain real materials by measuring the spectral power distribution of reflected light from a variety of inhomogeneous materials at several viewing angles, relative to the reflected light from an ideal diffuser (pressed barium sulphate powder) identically illuminated. They find that for many materials – plants, paints, oils, and plastics – but not for paper or ceramics, the spectral power distribution changes in magnitude but not in shape. That is, the spectral power distribution curves are linearly related to one another at the different viewing angles, or, the brightness but not the hue of body reflection from inhomogeneous materials changes with viewing direction. This result indicates that for these materials, $\rho_b(\lambda)$ is indeed independent of the imaging geometry, and that it is separable from F_b. A separate plot of the chromaticities of the reflected light at different angles indicates that the NIR model holds even for those materials (e.g., yellow paper) for which the spectral power distribution changes with viewing angle. This can occur because the chromaticities are the result of integrating the reflected light over the broad range of a sensor signal. Thus, in the integral, changes in spectral power distribution with viewing angle at one wavelength may be canceled by opposite changes at another wavelength.

In summary, the modified Lambertian reflectance model used in equation 3.2 is

accurate for body reflection from inhomogeneous materials with little or no specular component under a single light source. Color algorithms based on it can be expected to fail where the image irradiance signal is dominated by interface reflection. Equation 3.8 is more generally correct for inhomogeneous materials with or without significant specularities and can be simplified to the NIR model for many materials, but it also relies on the single source assumption. Neither equation accurately describes light reflection from homogeneous materials such as metals, and neither takes into account mutual reflections between objects.

3.3 Lightness Algorithms

Lightness algorithms [57, 61, 58, 43, 7, 50], pioneered by Land, assume that the color of an object can be specified by its lightness, or relative surface reflectance, in each of three independent chromatic channels, and that lightness is computed in the same way in each channel. (Lightness is defined by Evans [30] as "the visually apparent reflectance of a surface under a given set of conditions." Lightness algorithms attempt to recover an estimate of the "visually apparent" integrated reflectance in each of three distinct chromatic channels. The channels are usually assumed to be represented by the photoreceptor spectral sensitivities.) Computing color is thereby reduced to extracting surface reflectance from the image irradiance in a single chromatic channel.

The computation is simplified by several assumptions. The first is that there are no specular reflections. $\rho(\lambda, \mathbf{r}_s)$ is therefore the body reflectance factor. The second is that the viewing direction and surface normal are everywhere the same (true for a perfectly flat two-dimensional surface always viewed from the perpendicular angle). The effective irradiance in equation 3.2 therefore becomes a function only of wavelength and of source direction, which is written here as a dependence on the surface coordinate $\mathbf{r}_s$. That is, $E(\lambda)F_b(\mathbf{v}, \mathbf{n}, \mathbf{s}) = E(\lambda, \mathbf{r}_s)$. The third is that the light sensor response may be written as the logarithm of the product of terms approximating illumination and reflectance and may thereby be converted to a sum

of logarithms. The monochromatic image irradiance equation therefore becomes:

$$S^{\nu}(\lambda, \mathbf{r}) \;=\; \log a^{\nu}(\lambda)\rho(\lambda, \mathbf{r_s}) \;+ \log E(\lambda, \mathbf{r_s}). \tag{3.11}$$

For the signal integrated over the sensor sensitivity range, the product is not transformed so easily into a sum. It is clear that if $a^{\nu}(\lambda)$ were a delta function, $\delta(\lambda - \lambda_{\nu})$ (that is, if the sensor range were a very narrow band), or if the illumination consisted of very narrow-band lights, each falling in the range of only one sensor, Equation (3.3) would become:

$$S^{\nu}(\mathbf{r}) = \; \log\left[\rho^{\nu}(\mathbf{r})E^{\nu}(\mathbf{r})\right] \tag{3.12}$$

where $\rho^{\nu}(\mathbf{r})$ is the surface reflectance and $E^{\nu}(\mathbf{r})$ is the effective irradiance integrated over the narrow νth spectral band (dropping the subscripts on the spatial coordinates $\mathbf{r}$). For human photoreceptors and most filters and cameras used in artificial visual systems, the spectral sensitivities are not delta functions, so Equation (3.12) does not immediately follow. Other descriptions of lightness algorithms have neglected this problem.

The desired decomposition may be obtained exactly by first writing ρ and E as weighted sums of a fixed set of basis functions, chosen to describe fully most naturally-occurring illuminants and surface albedos and to satisfy certain constraints (see Section 3.4). By making appropriate transformations, the sensor signals may then also be written in terms of the basis functions, which therefore define new lightness channels, as illustrated in Appendix A. The image irradiance equation becomes

$$S^{i}(\mathbf{r}) = E^{i}(\mathbf{r})\rho^{i}(\mathbf{r}) \tag{3.13}$$

where the index i labels the type of new lightness channel formed by a linear combination of the original sensor channels. Taking logs of both sides yields:

$$\tilde{S}^{i}(\mathbf{r}) = \tilde{E}^{i}(\mathbf{r}) + \tilde{\rho}^{i}(\mathbf{r}), \tag{3.14}$$

where $\tilde{S}^i(\mathbf{r}) = \log S^i(\mathbf{r})$, and so forth.

This derivation relies on several assumptions, the most restrictive being that the basis functions into which the illuminant and reflectance are decomposed form a bi-orthogonal set with respect to integration with the original sensor channels. The visual system might nonetheless perform a similar, less exact transformation, of which the biological opponent-color channels, for example, may be the result.

Yet a more general way to achieve the decomposition is given by Lee et. al. [62]. This method highlights the fact that, given a finite number of sensors, each with a broad spectral sensitivity function, the visual system can recover only an approximation to surface reflectance. This fact is central to the problem of color constancy.

Under the assumptions and definitions given in the preceding section, the image irradiance equation given in equation 3.3 becomes:

$$S^\nu(\mathbf{r}) = \log\left[F_b(\mathbf{v},\mathbf{n},\mathbf{s})\int a^\nu(\lambda)\rho_b(\lambda,\mathbf{r})E(\lambda,\mathbf{r})d\lambda\right] \tag{3.15}$$

where $F_b(\mathbf{v},\mathbf{n},\mathbf{s})$ is defined as in equation 3.8 and $E(\lambda,\mathbf{r})F_b(\mathbf{v},\mathbf{n},\mathbf{s})$ is the effective irradiance. Defining the *integrated reflectance* $\rho^\nu(\mathbf{r})$ as:

$$\rho^\nu(\mathbf{r}) = \frac{\int \rho(\lambda,\mathbf{r})E(\lambda,\mathbf{r})a^\nu(\lambda)d\lambda}{\int E(\lambda,\mathbf{r})a^\nu(\lambda)d\lambda} \tag{3.16}$$

and the *integrated effective irradiance* E^ν as:

$$E^\nu(\mathbf{r}) = F_b(\mathbf{v},\mathbf{n},\mathbf{s})\int E(\lambda,\mathbf{r})a^\nu(\lambda)d\lambda, \tag{3.17}$$

Equation 3.15 becomes

$$S^\nu(\mathbf{r}) = \log\left[\rho^\nu(\mathbf{r})E^\nu(\mathbf{r})\right], \tag{3.18}$$

which may be decomposed into a sum of logs as above. The integrated reflectance is no longer perfectly constant under changes in the illumination, except in the limiting cases of narrow-band sensor spectral sensitivities or illuminants. Thus, even if the

visual system could recover ρ^{ν} exactly, it would demonstrate only approximate color constancy, unless ρ^{ν} is derived by the basis transformation in Appendix A.

3.3.1 The Lightness Problem

The assumption underlying the use of lightness algorithms in color computation [43, 7, 50] [and Crick, pers. comm.] is that Equation (3.14) may be solved for $\rho^{i}(\mathbf{r})$ independently and identically for each of the three chromatic channels, and the resulting triplet of lightness values specifies color. Thus at each location in the sensor array, there are three equations and six unknowns, so the problem is under-constrained. In the absence of direct measurements of either the effective irradiance or the surface reflectance, physical constraints must be imposed in order to solve for the unknowns.

The lightness problem may be broken down in terms of two sub-problems of color computation, *spatial decomposition* of the intensity signal and *spectral normalization* of the surface reflectance and the effective irradiance.

Spatial Decomposition

The first step in lightness algorithms is to split the intensity signal into its two components at each spatial location – spatial decomposition. This step is performed by spatial differentiation of the intensity signal under the following assumptions:

Lightness Assumption 1: The scene is a two-dimensional Mondrian, that is, a flat surface covered with irregularly shaped patches of uniform surface reflectance. Therefore, $\rho^{i}(\mathbf{r})$ is uniform within patches, but changes sharply at edges between patches.

Lightness Assumption 2: The effective irradiance varies slowly and smoothly across the entire scene and is everywhere independent of the viewer's position. $E(\lambda, \mathbf{n}, \mathbf{s})$ may therefore be written as a function of λ and $\mathbf{r}$, only, as in Equation 3.11.

Under these constraints, the two components are disentangled by:

(1) Differentiating the sensor signal, $S^{i}(\mathbf{r})$, over space.

(2) Thresholding the derivative, $d[S^i(\mathbf{r})]$, to eliminate small values that are due to smooth changes in the effective irradiance and to retain large values that are due to abrupt changes in the surface reflectance at borders between patches.

(3) Integrating the thresholded derivative, $Td[S^i(\mathbf{r})]$ over space to recover surface reflectance. (Because only the large changes due to reflectance remain in $d[S(\mathbf{r})]$ after thresholding, $Td[S(\mathbf{r})]$ is approximately equal to $d[\rho(\mathbf{r})]$.)

In summary, the first two lightness assumptions ensure that spatial changes in the effective irradiance may be disentangled from spatial changes in the spectral reflectance function of a surface, under the given conditions.

Spectral Normalization

Spatial decomposition does not recover surface reflectance exactly because the constant of integration, representing the absolute scale of $\rho^i(\mathbf{r})$ relative to $S^i(\mathbf{r})$, is lost in the final step and must be reset. (Furthermore, as discussed below, $\rho^i(\mathbf{r})$ is itself only an estimate of the integrated surface reflectance in channel i. The threshold operation is also inherently inaccurate, unless it is made flexible enough to recognize variations in the size of irradiance and surface reflectance spatial changes.) The result of the computation, lightness, is therefore at most only proportional to surface reflectance. We may write $[k^i]^{-1}[TS(\mathbf{r})]^i = L^i(\mathbf{r}) = c^i(\mathbf{r})\rho^i(\mathbf{r})$ where $[TS(\mathbf{r})]^i$ is the result of integration of the thresholded intensity before the constant of integration, $[k^i]^{-1}$ has been set, L^i is the lightness in the ith wavelength band, and $c^i(\mathbf{r})$ is a multiplicative factor that may vary across space.

The value of k^i depends on the spectral normalization that the algorithm performs. Most versions of the retinex algorithm and, by extension, other lightness algorithms that do not explicitly address the normalization problem, impose the final constraint:

Lightness Assumption 3: **The average surface reflectance of each scene in each wavelength band is the same: grey, or the average of the lightest and darkest naturally occurring surface reflectance values.**

This assumption, the "grey-world" assumption, appears in similar guise in other

theories.[6] Lightness algorithms that rely on it are here said to 'average to grey.'

k^i may then be set to the average value of the thresholded intensity of the scene, so that the computed lightness of a patch approximately equals its surface reflectance relative to the average reflectance of its surround. If lightness assumption 3 holds, then the lightness of a patch is an accurate and invariant measure of its surface reflectance. The triplet of lightness values in the three wavelength channels then defines the color of a patch.

Lightness Assumption 3 implies that any change of scale in S is always interpreted as a change of scale in E. To illustrate, consider a test patch against a background of random colors. Lightness algorithms compute the integrated effective irradiance on the scene, $E_C^i(\mathbf{r})$, by dividing $S^i(\mathbf{r})$ by the lightness $L^i(\mathbf{r})$: $E_C^i(\mathbf{r}) = E^i(\mathbf{r})\rho^i(\mathbf{r})/c^i(\mathbf{r})\rho^i(\mathbf{r}) = E^i(\mathbf{r})/c^i(\mathbf{r})$. If $E^i(\mathbf{r})$ is increased, $k^i(\mathbf{r})$ and $[TI]^i(\mathbf{r})$ are each increased by the same factor and the computed lightness of the test patch does not change. Effectively, the average surface reflectance in the ith-type channel is automatically reset to grey. $E_C^i(\mathbf{r})$ is increased by the same factor as $E^i(\mathbf{r})$, so the spectral skew is correctly interpreted. The closeness of E_C to E depends on the value of $c^i(\mathbf{r})$, which depends partly on the value of k^i and which may vary across space if, for example, a spatial gradient of surface reflectance has been discarded by the threshold operator.

Alternatively, if the background reflectances are skewed toward one part of the spectrum, the computed lightness of the test patch changes, but the spectral skew in the intensity signal is incorrectly interpreted as a skew in $E^i(\mathbf{r})$: unless $c^j(\mathbf{r})$ is skewed in the same way as $c^i(\mathbf{r})$, $E_C^i/E_C^j \neq E^i/E^j$. That is, lightness algorithms will 'see' a dull red patch against a range of green patches as lighter than when against a range of red patches under the same illumination, and will interpret the skew towards green as a lack of red in the illuminant. The integrated effective irradiance

[6] Evans was one of the first to formulate it. R.W.G. Hunt[49] (p. 294) quotes Evans: "The more pleasing effect is often produced in colour prints if they are so made that instead of the colour balance being correct, in which grey is printed as grey, it is so adjusted that the whole picture integrates to grey." Quote attributed to R. M. Evans, U. S. Patent 2571697, 1946.

is correctly recovered only if the spatial variation and the average of the integrated surface reflectance is the same in each channel (i.e., $k^i = k^j$ for all i, j).

In summary, the third lightness assumption provides a normalization scheme that will compensate for temporal changes in the spectral energy distribution of the illuminant on a given scene but cannot distinguish such changes from skews in reflectance distributions between scenes.

The original retinex algorithm (performed by a hypothetical neural system involving elements of both retina and cortex) [61] performs the step of spatial decomposition by taking the difference in $S(\mathbf{r})$ between adjacent points in many pairs along many paths radiating from $\mathbf{r}_o$, the point at which lightness is to be computed. The differences are thresholded and summed for each path, giving roughly the ratio between the reflectances at the start and end points, $\rho(\mathbf{r}_o)$ and $\rho(\mathbf{r}_e)$. Spectral normalization is performed by averaging the ratios over all end points to give roughly the ratio between $\rho(\mathbf{r}_o)$ and $\overline{\rho(\mathbf{r})}$, the average reflectance of the scene (see Section 4.1.1).

Other versions of the retinex algorithm use a different value for k^i. They assume not that the average surface reflectance in each channel is grey, but that the highest surface reflectance in each channel is white. That is, such algorithms assume that the brightest patch in each channel reflects all incident light in that channel. k^i may then be set to the intensity of the brightest patch in the ith channel. These methods are here said to 'normalize to white.'

3.3.2 Von Kries Adaptation

The spectral normalization methods that lightness algorithms employ are forms of Von Kries adaptation, as Worthey [112] observes. In Von Kries adaptation, long considered to be one of the primary mechanisms underlying color constancy, the receptor sensitivities $a^\nu(\lambda)$ are scaled by normalizing factors that vary with the illumination in order to keep perceived color constant. In the notation introduced in section 3.1, the Von Kries adaptation model says that under a change in illumination from $E_1(\lambda)$ to $E_2(\lambda)$, the receptor sensitivities are scaled

$$k_1^\nu a^\nu(\lambda) \mapsto k_2^\nu a^\nu(\lambda)$$

so that

$$C_1^k = C_2^k$$

where the C^k are the constant color values into which the receptor responses S^ν are transformed. For example, let $C^\nu = S^\nu$. Then Von Kries adaptation requires:

$$k_1^\nu \log\left[\int a^\nu(\lambda)\rho(\lambda,\mathbf{r})E_1(\lambda)d\lambda\right] = k_2^\nu \log\left[\int a^\nu(\lambda)\rho(\lambda,\mathbf{r})E_2(\lambda)d\lambda\right].$$

Averaging to grey achieves exactly this, by setting $(k^\nu)^{-1}$ to the average of S^ν over the scene. Normalizing to white sets $(k^\nu)^{-1}$ to the highest value of S^ν in the scene. In the first case, $k^\nu S^\nu$ will stay roughly constant provided that the average surface reflectance is the same for every scene, and in the second case, $k^\nu S^\nu$ will stay roughly constant provided that the highest surface reflectance in the νth channel is the same in each scene.

3.3.3 Predictions and Tests

Land demonstrates that color perception in a Mondrian world correlates with the triplet of lightness values as computed by the retinex algorithm with averaging to grey. As Land himself is careful to note, his demonstration does not show that the Mondrian colors stay constant under changing illumination, only that they are the same as those predicted by the retinex algorithm. He even emphasizes that the lightness values recovered by the retinex algorithm may have an arbitrary relationship to the physically correct surface spectral reflectance function.[7] Neither do Land's

[7] Land illustrates the arbitrary relationship between lightness and reflectance with a photograph of a Mondrian under a simple illumination gradient: although the surface reflectance of a photographed patch is a product of the Mondrian patch reflectance and the Mondrian illumination, and so is non-uniform, it is interpreted as uniform by lightness algorithms. This effect follows from the lightness assumption that all slow changes in the intensity signal are due to the source illumination, and our perception of the photographed Mondrian indicates that the assumption holds true to some extent.

demonstrations provide an adequate test of the hypothesis that, in the real world, color is computed by a retinex-type algorithm. The Mondrian world satisfies perfectly the three lightness assumptions: reflectance is uniform within patches and changes sharply between patches; the illumination varies smoothly across the scene; and, most importantly, the average surface reflectance of the Mondrian is the same from one viewing to the next, and indeed is roughly the same for all Mondrians with a sufficiently random collection of colors. If these assumptions are not satisfied in the real world, lightness algorithms will fail to recover constant colors there. The true test of lightness algorithms as models of human color computation is, then, to test whether lightness algorithms and human color perception fail in the same way when these assumptions are not met.

Spatial Decomposition

The first step is to test the lightness assumptions in the real world. Does the illumination always vary slowly and smoothly across the scene? In a world in which shape, shading, specularities and shadows confound the effective irradiance, the answer is clearly no. Lightness algorithms will fail to distinguish between sharp changes in the image irradiance due to object contours, specularities and shadows and those due to true reflectance edges. This inadequacy of lightness operators is graphically illustrated in Chapter 5. Segmentation algorithms, introduced in Section 3.5 below, address this problem by comparing the image irradiance signal in different wavelength channels across space.

As Land has recently demonstrated using still-life scenes, lightness algorithms may, nevertheless, approximate human perception even in the three-dimensional world ([59] and pers. comm.).

Spectral Normalization

But the critical question is one of spectral normalization. Is the grey-world assumption satisfied in most natural images, and, if not, does human color perception vary

with the average reflectance of the scene? Or does human color perception vary predictably with the presence or absence of true "whites" in the scene?

Ironically, the object color that the retinex algorithm computes has been criticized both for varying too much with its surround and for varying too little. That is, the retinex algorithm has been criticized for being unable to predict either color constancy or simultaneous color contrast in the context of changing backgrounds. The second criticism arises from the fact that if the normalizing factor is obtained by averaging over the entire scene, giving equal weight to areas close and far from the point at which lightness is computed, the color of a grey patch, say, will not change when the patches around it are merely shuffled in position. In other words, the local effect of simultaneous color contrast will disappear in the global computation of constant colors. But, as Brou et. al.[14] have demonstrated (see Section 2.2), a grey patch *will* change color strikingly as its surround is shuffled: it looks grey when surrounded by a random array of colors, yellowish when surrounded by bluish colors, and pinkish when surrounded by bluish colors.

The basis of the first criticism is clear: use of the grey-world assumption in lightness algorithms predicts that the computed color of a patch will change with the average reflectance of its surround. Brainard and Wandell [12] computed the change in lightness triplets for patches in simulated Mondrians with changing composition. The simulated Mondrians consisted of 9 patches in a 3x3 array whose spectral reflectance functions were chosen to match those of certain Munsell chips under CIE standard daylight D65. Lightness triplets for the relevant patches were computed by integrating the product of reflectance (using Cohen's basis functions) and illumination (using Judd and Wyszecki's description of D65) with the Smith-Pokorny estimates of the human cone spectral sensitivities. As expected, the lightness values for patches in the top two rows, which were held constant throughout, varied with the changing composition of the bottom row. To estimate the corresponding change in color, the lightness triplets were matched to "standard" triplets computed for 462 Munsell chips arranged in a simulated Mondrian under CIE D65. The matching Munsell value of

one of the constant patches changed from R8/4 to P6/6 (a relatively large change in color from red to purple) under certain substitutions in the bottom row. Brainard and Wandell state that "a human observer viewing these Mondrians on a black background illuminated by daylight perceives virtually no change in the appearance of the upper two rows of chips [with changing composition of the bottom row]."

This claim implies that humans do not rely on the grey-world assumption for spectral normalization. But it requires additional buttressing. To test whether humans do perform an instantaneous spatial average over the scene, one must restrict the human observer's view to the Mondrian alone and limit the duration of viewing. Brainard and Wandell make no mention of the field of view or whether it was carefully controlled. This leaves open the possibility that the human observer could obtain the normalizing factor by averaging over a much larger area than the Mondrian alone, which the simulated lightness algorithm was not allowed to do. The fact that the view was not limited to a flash in time leaves open the possibility that the average could be computed over time, as each receptor samples separate locations in space, increasing the likelihood that the normalizing factor is the same for each location in the scene and decreasing the effects of simultaneous contrast. Land [59] does, though, emphasize that color constancy can occur in a "fraction of a second" although he does not report details of such "flash" demonstrations.

McCann [78] reports a controlled test of the grey world assumption in which observers viewed Mondrians under diffuse illumination in a box that eliminated all other objects from view. One paper in each of five different Mondrians was chosen so that under different illuminants (composed of different ratios of three narrow-band lights) each of the five different papers emitted the same triplet of radiances in the three illuminant bands. The surrounds of the Mondrians were constructed so that the total average radiance from each display under the chosen illuminants was the same. Although lightness algorithms that relied on the grey-world assumption would now assign each of the five different papers the same color, human observers perceived the papers as having different colors, close to what lightness algorithms would calculate if

the papers appeared under neutral illuminants against unbiased backgrounds. Local contrast effects were observed but could not account for the extent of deviation from the grey-world predictions. On the other hand, lightness algorithms that normalize to white would correctly predict the observer's responses in this experiment.

Both experiments indicate that although the grey world assumption might play a role in spectral normalization, it cannot be the only responsible factor. In extreme cases of biased surrounds, the grey world assumption predicts larger color shifts than those observed. Normalizing to white, which is closely related to the grey world assumption, might be the more dominant factor, but it has not been fully explored.

Let us assume for the sake of argument that humans do normalize via a form of Von Kries adaptation, say normalizing to white, or by inversely scaling the sensitivity in each lightness channel by the highest value of the image irradiance in that channel. The question then arises: which is the lightness channel scaled? That is, are the receptor sensitivities themselves scaled or are the responses in channels formed by transformations of the receptor channels scaled? For example, if the decomposition of the image irradiance signal into the sum of logs is done as in the appendix, the set of channels indexed by i in equation 3.14 do not represent the photoreceptor types. Rather they represent channels formed by a linear combination of the photoreceptor activities, and these channels correspond to a fixed set of basis functions in terms of which the source irradiance and surface reflectance are decomposed. McCann et.al. [79] specifically state that lightness is computed independently in each of the channels defined by the cone spectral sensitivities. The proposal of a global computation that occurs independently in each cone channel is, as McCann et.al. write, "a significant departure from Young's theory, which proposes intercomparisons of long-, middle- and short-wave receptors at a point."

In fact, specific predictions about the strength of color constancy under certain illuminant changes may be tested to distinguish between normalization in photoreceptor channels versus opponent-color channels. If normalization takes place in opponent-color channels, one would expect color constancy to be weakest under illuminant

changes along the blue-yellow axis. The proof of this statement for the illuminants used by McCann et. al. [79] is illustrated in Appendix B. McCann et. al. [79] measure the changes in color of Mondrian patches under varying illumination provided by three narrow-band lights. Their results indicate that color constancy is strongest under shifts along the blue-yellow axis (as demonstrated in detail by Worthey [112]), suggesting that normalization indeed takes place in the original cone channels.

3.4 Spectral Basis Algorithms

Spectral basis algorithms [15, 13, 70, 101, 115] are founded on the assumption that most naturally occurring illuminants and reflectances can be described fully by linear combinations of a small number of fixed basis functions. Spectral basis algorithms start with a simplified form of the image irradiance equation similar to Equation 3.15 in which the sensor responses are written as a linear function of the sensitivity integral:

$$S^{\nu x} = \int a^{\nu}(\lambda) R^{x}(\lambda) E^{x}(\lambda) d\lambda$$

Here $a^{\nu}(\lambda)$ is the sensor sensitivity, $R^{x}(\lambda)$ is the surface reflectivity function, in which the dependence on the viewing geometry is incorporated into the image spatial coordinate x, and $E^{x}(\lambda)$ is the illumination. The central constraint on this equation is:

- $R^{x}(\lambda)$ and $E^{x}(\lambda)$ may be written in terms of a fixed set of basis functions: $R^{x}(\lambda) = \sum_{j=1}^{n} \rho_j^x R_j(\lambda)$ and $E^{x}(\lambda) = \sum_{i=1}^{m} \epsilon_i^x E_i(\lambda)$, where $R_j(\lambda)$ and $E_i(\lambda)$ are basis functions and ρ_j^x and ϵ_i^x are spatially varying coefficients.

Additional constraints are then imposed on $R^{x}(\lambda)$ and $E^{x}(\lambda)$:

- $E^{x}(\lambda)$ is uniform across space (or slowly varying enough that its spatial dependence can be ignored). That is, $E^{x}(\lambda) = E(\lambda)$ and $\epsilon_i^x = \epsilon_i$.

- $R^x(\lambda)$ is implicitly assumed to be invariant under changes in the viewing geometry, which can only be true for the body reflection component of an ideally diffusely reflecting surface. In other words, spectral basis algorithms assume a Lambertian reflection model.

Under these constraints the image irradiance equation becomes:

$$S^{\nu x} = T^{\nu j}(\epsilon)\rho_j^x$$

where $T^{\nu j}(\epsilon) = \sum_{i=1}^{m} \epsilon_i \tau_{ij\nu}$ and $\tau_{ij\nu} = \int a^\nu(\lambda) R_j(\lambda) E_i(\lambda) d\lambda$. The variables have finite ranges: $\nu = 1, 2, ...k$ and $j = 1, 2, ...n$. The matrix $T^{\nu j}(\epsilon)$ depends on the illuminant and reflectance basis functions, on the sensor spectral sensitivities, which are fixed, and on the illuminant, which is variable. If $T^{\nu j}(\epsilon)$, which Maloney [70] terms the *light transformation matrix*, is known and is nonsingular, then the coefficients ρ_j^x may easily be determined by inverting it (see below).

The above equation may be written in vector form:

$$\mathbf{S}^x = T^{\nu j}(\vec{\epsilon})\vec{\rho}^x$$

where $\mathbf{S}^x = (S^\nu, S^\mu, S^\omega)^x$, the photoreceptor responses at the image location x, and $\vec{\rho}^x = (\rho^k, \rho^l, \rho^m)^x$, the coefficients of the surface spectral reflectance function at x.

If $T^{\nu j}(\epsilon)$ is not known, then for each location x there are $(m + n)$ unknowns (n components of $R[= R^x(\lambda)]$ and m components of $E[= E(\lambda)]$) and k equations. The unknowns cannot be found unless the equations are further constrained.

Buchsbaum [15] solves for the coefficients ρ_j^x by imposing a fourth constraint equivalent to the grey-world assumption. He writes:

$$\overline{\mathbf{S}} = Q^{\nu i}(\mathbf{a})\vec{\epsilon}$$

where $\overline{\mathbf{S}}$ is the average sensor response vector over the entire field,

$$\overline{\mathbf{S}} = \sum_x b^x \int a^\nu(\lambda) R^x(\lambda) E(\lambda) d\lambda,$$

with b^x the weighting factor for the reflectance at x. $\vec{\epsilon}$ is the vector formed by the illumination basis function coefficents. $Q^{\nu i}(\mathbf{a})$ depends on the average surface spectral reflectance of the scene, **a**. That is,

$$Q^{\nu i}(\mathbf{a}) = \sum_{j=1}^{n} \tau_{ij\nu} \left(\sum_{x} b^x \rho_j^x \right).$$

Buchsbaum then assumes that

$$Q^{\nu i}(\mathbf{a}) = Q^{\nu i}(\mathbf{h}) = \sum_{j=1}^{n} \tau_{ij\nu} \rho_j^h.$$

where ρ_j^h is the surface spectral reflectance of a standard material, e.g., a reference grey. That is, Buchsbaum imposes the following constraint:

- The weighted average of all reflectances in a given scene is known.

Under this final constraint, the basis function coefficients of R and E may be computed by inverting $Q^{\nu i}(\mathbf{h})$ to find $\vec{\epsilon}$.

(Because Buchsbaum chooses approximations to the normalized sensor spectral sensitivities as basis functions for both illumination and reflectance, $a^i(\lambda) = R^i(\lambda) = E^i(\lambda)$, and because the assumptions he employs are equivalent to the Retinex assumptions, the coefficients he recovers are directly related to the lightness triplets the Retinex recovers. But they are not identical because the $a^i(\lambda)$ are not an orthogonal set of basis functions.)

Wandell and Maloney [70, 71] and Yuille [115] find $T^{\nu j}(\epsilon)$, and therefore E, by a different method. They assume that there is at least one more sensor type than there are components of R. This obviates the need for the grey-world assumption and exploits the redundancy in signals carried by different spectral channels at different image locations. If $k = n + 1$ then the number of equations obtained by taking s samples of the scene is $(n + 1)s$. Assuming that each sample is taken from an area of different surface spectral reflectance, the number of unknowns is $(sn + m)$, since the illumination is the same at each sampled location. Therefore if $s > m$, the number

of equations exceeds the number of unknowns and, in principle, R and E may be computed. In practice, since the set of equations is nonlinear, a solution does not necessarily exist but must be demonstrated. Wandell and Maloney (W-M) find the m components of E by finding the equation of the geometric surface on which the samples S^ν fall, computing $T^{\nu j}(\epsilon)$ from E, and computing R with the inverse of $T^{\nu j}(\epsilon)$. Yuille proposes a similar solution.

These algorithms thus eliminate the need for assumptions about the average reflectance by comparing the irradiance signals sampled by different sensors at different image locations. Given only three photoreceptors, these algorithms would recover constant colors only for materials that can be adequately characterized by two basis functions.

Note that whereas in lightness algorithms the scale of R relative to E for each individual channel cannot be recovered, in spectral basis algorithms only one normalizing factor is lost.

3.4.1 Predictions and Tests

The W-M algorithm predicts that color constancy will be possible only for (1) scenes in which there are at least as many distinct surfaces as degrees of freedom in reflectance and (2) a certain class of reflectances fully describable by only 2 degrees of freedom, assuming we have only 3 receptor types operative in color vision. If there is such a class of reflectances then the first prediction says that there must be at least 2 such distinct surfaces for color constancy to hold. Lightness algorithms that rely on the grey-world assumption may be tested directly against spectral basis algorithms by finding a subset of such surfaces that under neutral illumination averages to a color other than grey. If these surfaces are the only objects in the field of view, then the W-M algorithm as well as lightness algorithms predict that they will display constant colors under changing illumination. But if one surface is removed and placed among another subset of surfaces describable by 2 degrees of freedom, which now averages to grey under neutral illumination, the W-M algorithm predicts that its color will stay

the same, whereas the grey-world assumption predicts that it will change.

Spectral basis algorithms have not been tested adequately either as models of human perception or for machine vision applications. As H.-C. Lee notes (pers. comm.), a set of basis functions with fixed phase with respect to wavelength cannot fit all naturally occurring reflectances and illuminants, and, unless the basis functions are chosen to be orthogonal to each other, will not allow a numerically stable solution.

3.5 Segmentation by Material Boundaries

A third class of algorithms, "segmentation algorithms", (see [98] and Chapter 6) take on a slightly different goal from lightness and spectral-basis algorithms. The primary objective of segmentation algorithms is to use color information to segment the scene into distinct materials. An ancillary objective is to assign to the distinct materials color labels that are invariant under illumination changes. This goal requires that segmentation algorithms recover only a coarse description of the surface spectral reflectance.

Segmentation algorithms thus directly address the first goal of color vision as put forward in Chapter 1: to distinguish objects from one another in a given scene. To meet this goal, segmentation algorithms must successfully distinguish image irradiance changes due to object boundaries from those due to, e.g., shadows, specularities and surface orientation changes, which is exactly the task on which lightness algorithms most often fail. Appropriately, segmentation algorithms rely on one critical feature of color vision that lightness algorithms ignore: the interaction between image irradiance signals in different wavelength channels.

The advantage gained over simply computing lightness in independent chromatic channels becomes clear in the real world. For 2-D Mondrians illuminated by a single light source, lightness algorithms successfully identify sharp irradiance changes with boundaries between different materials. In the real 3-D world, shadows, shading, specularities and surface orientation changes also cause irradiance changes that

can be confused with true reflectance boundaries. For example, lightness algorithms cannot distinguish the sharp decrease from bright to dark across a shadow edge from a reflectance change between patches of different reflectance. The segmentation algorithm of Rubin and Richards [98], for example, compares the signals in different chromatic channels and rules out a shadow edge if the irradiance in one channel decreases sharply as the irradiance in a second channel increases sharply across the same edge. Segmentation algorithms similarly distinguish other irradiance edges from those caused by reflectance changes.

Strictly speaking, the existence of a visible shadow implies the existence of more than one light source. That is, if an object is illuminated by a single, localized light source, the shadow it casts will be totally dark unless the shadowed area is illuminated by a second source not occluded by the object. Segmentation algorithms must therefore make three assumptions about the world: (1) there is a single light source (the single source assumption; see Chapter 6) (2) all surface reflectances average to grey (the grey-world assumption) and (3) the diffuse interrreflections between object surfaces therefore average to the color of the light source and serve to illuminate the shadowed area. The "shadow" source therefore has the same color as the primary light source.

Note that because of the inherent ambiguity of the irradiance signal, the grey-world assumption, or its equivalent, is needed by every type of algorithm, whether it provides normalization on one (as for spectral-basis algorithms) or three (as for lightness algorithms) spectral channels. In requiring the single source assumption, segmentation algorithms are more restricted than lightness algorithms, but this is because the former attempt to solve a different and more difficult problem.

In the following section the algorithm proposed by Rubin and Richards [98] is briefly discussed. In Chapter 6, a new segmentation algorithm is discussed in detail and the results of its implementation on real images are illustrated.

3.5.1 Rubin's and Richards' Algorithm

In Rubin's and Richards' algorithm [98], edges in the three image irradiance signals (S^i, S^j and S^k) are first detected by a separate algorithm. Two conditions are then tested for each edge to determine whether it corresponds to a material boundary.

1. *Spectral Crosspoint*: Two points ($\mathbf{r}_1$) and ($\mathbf{r}_2$) are chosen on opposite sides of the edge. The crosspoint product is calculated:

$$[S^i(\mathbf{r}_1) - S^i(\mathbf{r}_2)][S^j(\mathbf{r}_1) - S^j(\mathbf{r}_2)]$$

   If the product is negative, that is, the irradiance in one channel increases across an edge while the irradiance in another channel decreases, a spectral crosspoint is said to occur. Under the single source assumption, neither shadows nor surface orientation changes on a uniform material can produce a spectral crosspoint, although specularities can.

2. *Ordinality Violation*: The regions marked by edges which produce spectral crosspoints are then characterized by their *ordinality*, which (liberally interpreting the procedure described in [98]) is the doublet of $+/-$ signs given by calculating the sign of the following ratio for two distinct pairwise combinations of unlike channels:

$$sign\left[\frac{S^i(\mathbf{r}_1) - S^j(\mathbf{r}_1)}{S^i(\mathbf{r}_1) + S^j(\mathbf{r}_1)}\right]$$

   where $\mathbf{r}_1$ is chosen on one side of the edge and $i \neq j$. Ordinality is normalized by scaling each S^i and S^j by α and β respectively, where α and β are selected so that the median of the ordinality ratio (not its sign), taken over the entire image, is zero:

$$median\left[\frac{\alpha S^i(\mathbf{r}_1) - S^j(\mathbf{r}_1)}{\alpha S^i(\mathbf{r}_1) + S^j(\mathbf{r}_1)_{all(\mathbf{r})}}\right] = median\left[\frac{\beta S^j(\mathbf{r}_1) - S^k(\mathbf{r}_1)}{\beta S^j(\mathbf{r}_1) + S^k(\mathbf{r}_1)_{all(\mathbf{r})}}\right] = 0$$

This step relies on the grey-world assumption: that for a given scene the average reflectance in each spectral channel is the same "grey." Ordinality gives a crude spectral characterization of a material marked by the preceding segmentation step. When ordinality changes across an image irradiance edge, ordinality violation is said to occur.

Rubin and Richards suggest that either the spectral crosspoint or the ordinality violation condition may be used to identify material boundaries, and that the regions so defined may then be characterized by their ordinality. It should be noted, though, that neither the spectral crosspoint condition nor the ordinality violation condition is sufficient to identify material boundaries in the absence of the single source assumption or in the presence of specularities. For example, if two spatially and spectrally distinct light sources illuminate a uniform surface, their energies can be adjusted to produce a spectral crosspoint at an edge. Similarly, a strong specularity produced by a spectrally biased light source can give rise to both a spectral crosspoint and a violation of ordinality (see Figure 6.2 for illustration of a similar point). Furthermore, neither condition is necessary for a material boundary. Since there are no reports on the implementation of this algorithm on real images, it is not possible to judge how robust it would be given these caveats. Chapter 6 illustrates how a similar segmentation algorithm performs on real images.

3.6 Recovering the Illuminant Color using Specular Reflections

Recovering reflectance from the image irradiance in Equation 1 would be simple and straightforward if the illuminant spectral energy distribution were known. In fact, in some situations, one can obtain direct knowledge of the illuminant color from the image irradiance. A fourth class of color algorithms find the illuminant color from specular reflections and use it to extract invariant descriptors of reflectance from image irradiance.

Klinker, Shafer and Kanade [102] use the *dichromatic reflection model* as the basis for an algorithm that separates specular from diffuse reflections for objects made of inhomogeneous materials. As described in section 3.2, the light that an inhomogeneous material reflects is the sum of specular reflections from its surface and diffuse reflections from the pigment particles imbedded in it. Like the NIR model, the dichromatic reflection model therefore assumes that the color of the reflected light is the vector sum of the color of the illuminant and the color of the pigment particles (or the color of the object), in color three-space. The color vector for each pixel in an image thus has two components: one vector that has the same direction as the illuminant color vector, with a magnitude depending on the imaging geometry; and another vector in the direction of the object color vector, whose magnitude also depends on geometry. The algorithm based on this model finds the illuminant color vector by finding characteristic cluster shapes in the three-dimensional plot of pixel color values from an image.

Lee [63] and Healey and Binford [39] use similar techniques to find illuminant color based on the same assumption, that for many materials the reflected light color is the sum of the body and illuminant colors. In his *chromaticity convergence* algorithm, Lee plots pixel color values in the two-dimensional CIE chromaticity space, or in an equivalent two-space with coordinates S^i/S^j and S^i/S^k. (Taking ratios of the image irradiance values in different color channels factors out the absolute illumination level.) The pixel values for any one object lie on a line connecting the coordinates of the illuminant color and the object color (see Figure 3.2(a)). (That is, if the light source color is considered the "white" point, the effect of a specularity is to move the color of the reflected light away from the color of the object and toward "white," or in other words, to "desaturate" it. Of course, if the light source is strongly saturated itself, this will not be destauration in the usual sense.) When the pixel values from many objects are plotted on one chromaticity diagram, the result is a set of lines, all of which point roughly to the single point that represents the illuminant color (see Figure 3.2(b)). Lee has proposed several methods ([63], pers. comm.) for finding the

convergence point that identifies the source color.

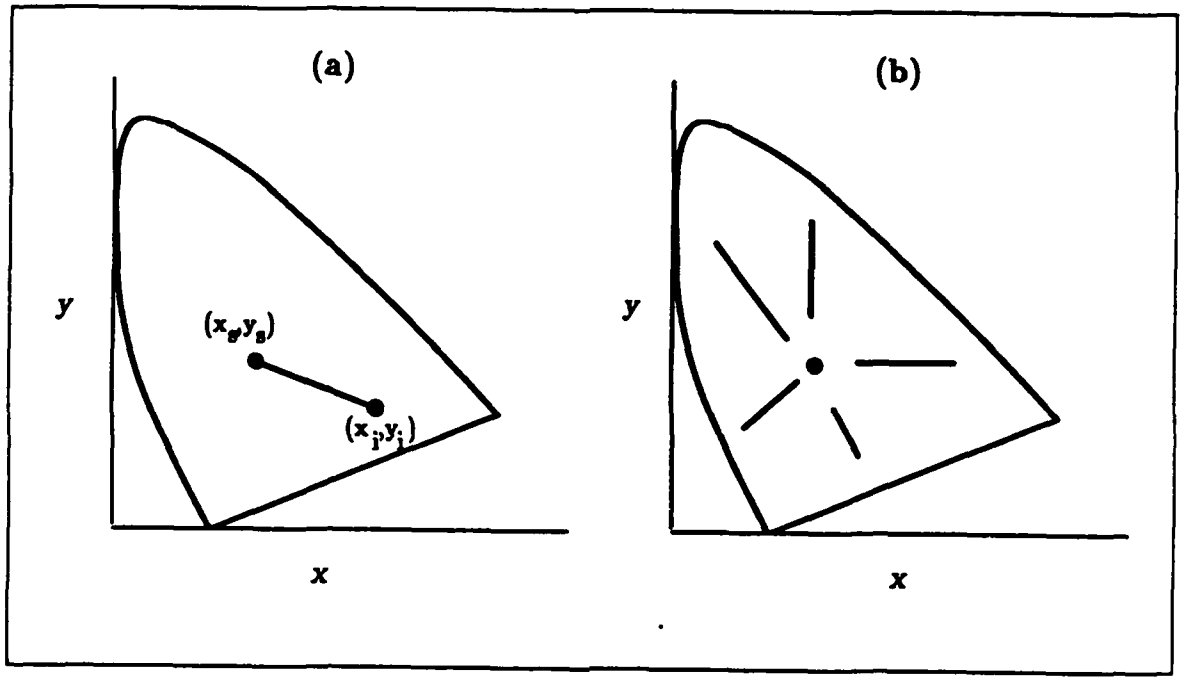

Figure 3.2: (a) The pixel color values from an inhomogeneous material fall on a line in the CIE chromaticity plane. The endpoints are the light source color (x_s, y_s) and the pigment or body color (x_i, y_i). (b) Pixel color values plotted from several different inhomogeneous materials under the same illuminant form a set of lines which converge on the source color. Redrawn from [63].

Both algorithms described above work under the single source assumption – both require that the illuminant color be the same for all objects in the scene, so that it may be identified as the color common to all pixels. As Lee has demonstrated using images of natural scenes (pers. comm.), variations in pigment concentration, surface orientation, etc., insure that the mix of specular and diffuse reflections varies enough so that even without overt highlights there is enough information to recover the illuminant color.

These algorithms demonstrate how to exploit information in a complicated, realistic image irradiance equation. The ostensibly simplifying assumptions underlying lightness algorithms, for example, might actually make the problem harder. It might

be easier for our visual system to recover the illuminant color given shading, 3-D curvature and specularities. The experiment described in Chapter 7 tests whether the human visual system uses information from specularities in estimating the color of the illuminant.

3.7 Conclusions

Most of the algorithms described here rely on similar assumptions about the natural world and hence may be expected to fail in the same way when the assumptions are not met. For example, all classes of algorithm except the last will give inaccurate results when confronted with surface reflectances with significant specular components. Whereas lightness algorithms are designed to compensate for illumination that varies smoothly across space, spectral basis algorithms (in their present form) can compensate only for virtually uniform illumination. Yet both require a Mondrian world in which boundaries between surfaces of different reflectances are marked by step changes in image irradiance. Some forms of lightness algorithms, segmentation algorithms, and spectral basis algorithms require that the collection of surface reflectances average to the same "grey" in every scene. Segmentation algorithms further require that there be only a single light source illuminating the scene. Another form of spectral basis algorithm relinquishes the grey-world assumption by assuming instead that there is one more sensor type than there are independent components of surface reflectance. To distinguish better between these algorithms as candidates for explanations of human color vision, more quantitative tests of color phenomena are needed.

Chapter 4

Lightness Algorithms

Lightness algorithms, which recover an approximation to surface reflectance in independent wavelength channels, were introduced in Chapter 3 as one method to compute color. In this chapter, specific lightness algorithms are reviewed in detail and a new extension (the multiple scales algorithm) of one of them is proposed. Two new results are discussed. The first is the fact that each of the lightness algorithms may be derived from a single mathematical formula under different conditions, which, in turn, imply different limitations for their implementations by man or machine. In particular, the algorithms share certain limitations in their implementation that follow from the physical constraints imposed on the statement of the problem and the boundary conditions applied in its solution. This analysis leads to a new rigorous statement of Land's original retinex algorithm. The second is the demonstration of a framework based on regularization theory that incorporates several of the assumptions on which most lightness algorithms are based and leads to a form of lightness algorithm similar to that most recently proposed by Land [60].

4.1 Review of Lightness Algorithms

The starting-point is again the simplified form of image irradiance equation derived in Section 3.3 (Equation 3.18):

$$S^i(\mathbf{r}) = \log I^i(\mathbf{r}) = \log \left[E^i(\mathbf{r})\rho^i(\mathbf{r})\right] = \tilde{E}^i(\mathbf{r}) + \tilde{\rho}^i(\mathbf{r}). \quad (4.1)$$

where $S^i(\mathbf{r})$ is the logarithm of the image irradiance $I^i(\mathbf{r})$, $E^i(\mathbf{r})$ is the integrated effective irradiance, and $\rho^i(\mathbf{r})$ is the integrated surface reflectance, each given in the ith-type lightness channel. $\tilde{\rho}$ is the logarithm of ρ; $\tilde{E}$ is the logarithm of E.

4.1.1 The Retinex

Land's retinex algorithm [58] uses the three lightness assumptions described in section 3.3.1 and assumes that the image irradiance signal is recorded by a discrete array of sensors, indexed by a one-dimensional variable.

Spatial decomposition of the intensity signal is performed by one-dimensional first-order spatial differentiation in the following way. Sensor x is connected by random paths to each of many different sensors w ($w = 1, 2, ...N$; w labels the startpoint and the path). Following a single direction in each path, at each junction (k^w, $k^w + 1$) the difference of activities ($S^i_{k^w+1} - S^i_{k^w}$) is calculated and thresholded, for each lightness channel i. The sum of differences up to sensor x is collected for each path that crosses it. For one path, this sum of differences approximates the difference between the signal at x and the signal at the start-point of the path. Under the three lightness assumptions, this difference, $\tilde{L}^i(x, w)$, closely approximates $\log(\rho^i_x/\rho^i_w)$, the log of the ratio of surface reflectances at the end- and start-points.

Spectral normalization is then achieved by averaging $\tilde{L}^i(x, w)$ over all startpoints w, which, by lightness assumption 3, fall in a random assortment of colors, to yield the lightness $L^i(x)$:

$$L^i(x) = \frac{\sum_{w=1}^N \tilde{L}^i(x, w)}{N} \simeq \log \rho^i_x - \frac{\sum_{w=1}^N \log \rho^i_w}{N} = \log \rho^i_x - \overline{\log \rho^i_w}.$$

As observed by Brainard and Wandell [12], the latter expression is equivalent to:

$$\log \left[\rho^i_x / \hat{\rho}^i_w\right]$$

where $\hat{\rho}^i_w$ is the geometric mean of surface reflectances in the scene. Lightness is therefore normalized by the average surface reflectance of the scene, as described in Chapter 3.

In the original retinex algorithm [61], spectral normalization is achieved by a different method: normalizing to white. Paths are drawn across the sensor array as above. At each junction in each path, the ratio of image irradiances $I^i_{k^w+1}/I^i_{k^w}$ is taken. (Note that since $I = \log S$, the ratio is equivalent to the difference above). The sequential product of ratios up to sensor x is taken along each path that crosses x. The threshold step is implemented by setting each ratio close to 1 equal to 1. Normalizing to white is achieved by resetting to 1 each ratio greater than 1 before it enters the sequential product. This reset operation insures that the sequential product at sensor x for one path, $\tilde{L}^i(x,w)$ is equal to the ratio of the irradiance at x to the highest irradiance encountered along that path. Under the lightness assumptions, this ratio is approximately equal to the ratio of surface reflectance at x to the highest reflectance along the path. Assuming that each path reaches the area of highest reflectance in the scene, $\tilde{L}^i(x,w)$ is a measurement of the reflectance at x normalized to "white." The lightness at sensor x, $L^i(x)$ is then set to the geometric mean of $\tilde{L}^i(x,w)$ over all paths. In other words, lightness is normalized by assigning 100% reflectance to the brightest patch in each lightness channel.

4.1.2 Horn's Algorithm

Horn's algorithm [44] makes the first two lightness assumptions listed in Section 3.3.1. It differs from the retinex algorithm in that the rotationally symmetric two-dimensional Laplacian operator, ∇^2, performs the first step, spatial differentiation of the intensity signal. The lightness obtained by this method is therefore a solution to the Poisson equation inside the Mondrian:

$$\nabla^2 L(x,y) \;=\; T[\nabla^2 S(x,y)], \qquad (4.2)$$

where T represents the threshold operation which is performed on the output of the

Laplacian.[1]

In its continuous form (that is, when (x, y) are continuous spatial variables), Equation (4.2) has a known solution under certain boundary conditions. If (1) the region on which S is defined (the sensor array or retina) is infinite *or* (2) the sensor array is finite, but the image is wholly contained in it and is surrounded by a border of constant reflectance for which $T[\nabla^2 S] = 0$, the inverse Laplacian is then performed by convolution with the Green's function $g = \frac{1}{2\pi}\ln(r)$:[2]

$$L(x,y) = \int\int (1/2\pi)\ln(r) \ * \ T[\nabla^2 S(\zeta,\eta)]\, d\zeta d\eta \tag{4.3}$$

where $r^2 = (x-\zeta)^2 + (y-\eta)^2$ and (x, y) are the coordinates of the point in the sensor array at which lightness is evaluated.

When either condition (1) or (2) is met, the reconstructed lightness is unique up to a constant, which is set by normalizing as prescribed by Lightness Assumption 3. Condition 1 is not satisfied by the photoreceptor array in the human retina. For a finite sensor array or a retina, condition (2) is crucial to Horn's solution; its importance is discussed later.

When the intensity function S is not continuous but discrete, for example, sampled by a discrete sensor array, and boundary condition (2) is met, the inverse solution may be expressed as a convolution with a similar function.

An alternative method of solving Equation (4.2) in the discrete case is by iteration. The Laplacian is approximated by a discrete differencing operator:

$$T(S_{ij} - 1/N \sum_{i=1}^{N} S_{kl}^{i}) \ = \ T[\nabla^2 S_{ij}] \tag{4.4}$$

where S_{ij} is the discretized version of $S(x, y)$; (ij) are coordinates of retinal cells; and S_{kl}^{i} are the N closest neighbors of S_{ij}, where $N = 6$ a closely-packed array.

[1]The superscript i is from now on dropped in the expressions $L^i(x,y)$, $S^i(x,y)$ etc. although it is still there by implication.

[2]In Horn's original formulation, the Green's function used is $(1/2\pi)\ln(1/r)$, which lacks the minus sign that makes it equal to the one used here.

Because $T[\nabla^2 S_{ij}] = \nabla^2 L_{ij}$ (Eq. 4.2), the left-hand side of Equation (4.4) may be rewritten as:

$$L_{ij} - \frac{1}{N}\sum^{N} L_{kl} = \nabla^2 L_{ij} \tag{4.5}$$

where L_{ij} is the lightness to be solved for. To perform the inverse Laplacian, Gauss-Seidel iteration takes as its starting-point the exact solution of Equation (4.5):

$$L_{ij} = \nabla^2 L_{ij} + \frac{1}{N}\sum^{N} L_{kl}. \tag{4.6}$$

In the first step of the iteration, the L_{kl} are set to zero, so that the first solution is $L_{ij} = \nabla^2 L_{ij} = T[\nabla^2 S_{ij}]$. This solution is then substituted for the L_{kl} in the next step (that is, $L_{kl} = \nabla^2 L_{kl}$). At each subsequent step of the iteration, the solutions L_{ij} obtained in the previous step are substituted for the L_{kl}. The iteration converges slowly but stably to the correct solution.

4.1.3 The Multiple Scales Algorithm

The multiple scales algorithm is an alternative method of solution for the inverse Laplacian problem expressed in Equation (4.2). This solution is based on the equation

$$\begin{aligned} -\int_0^\infty dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-\zeta, y-\eta; t)\nabla^2 f(\zeta,\eta)d\zeta d\eta = \\ \int_0^\infty dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nabla^2 G(x-\zeta, y-\eta; t) f(\zeta,\eta)d\zeta d\eta \;= f(x,y) \end{aligned} \tag{4.7}$$

which states that the Laplacian operator may be inverted (modulus a harmonic function) by summing its convolution with a Gaussian (G) over a continuum of variances or scales (t). [3] It should be noted at the outset that although this equality is true for continuous integrals, it might degrade gracelessly in the discrete case. In other words, a finite sum of terms over a discrete range of scales might not converge to the desired function. This is possible because each term in the sum over scales is a bandpass

[3] A similar equation and result has been independently obtained by Zucker and Hummel[118], although not applied to lightness computation.

function in the spatial frequency domain; that is, $\nabla^2 G$ assumes very small values for zero or very high frequencies. These frequencies may therefore be irretrievably lost in the sum over scales, particularly if there is sufficient noise in the original signal.

The proof of Equation 4.7 is as follows. Any function $u(x,y;t)$ such that

$$u(x,y;t) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x \;-\; \zeta, y \;-\; \eta; t) f(\zeta,\eta) d\zeta d\eta, \tag{4.8}$$

where G is the Gaussian kernel and $f(\zeta,\eta)$ is a bounded function, satisfies the heat equation,

$$\nabla^2 u(x,y;t) \;= \frac{\partial}{\partial t} u(x,y;t). \tag{4.9}$$

Conversely, given any bounded function $f(x,y)$, $u(x,y;t)$ may be constructed as in equation 4.8. Then, given that

$$-\;\int_0^T \frac{\partial}{\partial t} u(x,y;t) dt \;=\; u(x,y;0) \;-\; u(x,y;T),$$

it follows that

$$-\lim_{T\mapsto\infty}\int_0^T \frac{\partial}{\partial t} u(x,y;t) dt \;=\; u(x,y;0) \;-\; \lim_{T\mapsto\infty} u(x,y;T) \;=\; u(x,y;0), \tag{4.10}$$

if we set $\lim_{T\mapsto\infty} u(x,y;T)$ to zero. (In doing so, we impose the boundary condition that any harmonic function added to the solution must equal zero everywhere.) The heat equation then implies that

$$-\lim_{T\mapsto\infty}\int_0^T \nabla^2 u(x,y;t) dt \;=\; u(x,y;0). \tag{4.11}$$

Because

$$\nabla^2 u(x,y;t) =$$

$$\nabla^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x-\zeta, y-\eta; t) f(\zeta,\eta) d\zeta d\eta \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nabla^2 G(x-\zeta, y-\eta; t) f(\zeta,\eta) d\zeta d\eta,$$

equation 4.11 implies

$$-\lim_{T\mapsto\infty}\int_0^T dt\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\nabla^2 G(x-\zeta,y-\eta;t)f(\zeta,\eta)d\zeta d\eta = u(x,y;0)$$

and therefore, because

$$u(x,y;0) =$$

$$\int_{-\infty}^{\infty} G(x-\zeta,y-\eta;0)f(\zeta,\eta)d\zeta = \int_{-\infty}^{\infty}\delta(x-\zeta,y-\eta)f(\zeta,\eta)d\zeta d\eta = f(x,y)$$

the statement in equation 4.7 is proved.

The last part of Equation (4.7) implies that performing the Laplacian and then taking its inverse on a function f is equivalent to summing the convolution of f with $-\nabla^2 G$ over a continuum of scales. Because the Gaussian itself satisfies the heat equation,

$$\nabla^2 G = \frac{\partial}{\partial t}G(x,y;t) \simeq \frac{G(x,y;t_1)-G(x,y;t_2)}{(t_1-t_2)}, \tag{4.12}$$

one may approximate the $\nabla^2 G$ filter as the difference of two Gaussians of different scales. The sum over $\nabla^2 G$ filters is therefore roughly equivalent to the sum over difference-of-Gaussians (DOG), in which all but the smallest- and largest-scale Gaussians are canceled (as illustrated in figure 4.1). Formally, the sum of $\nabla^2 G$ over all scales yields

$$-\lim_{t_1\mapsto 0}G(x,y;t_1) - \lim_{T\mapsto\infty}G(x,y;T) = \delta(x,y) - k$$

where $\delta(x,y)$ is a delta function, a Gaussian of infinitesimally small scale, and k is a constant that may be normalized to zero. Equation (4.7) therefore becomes:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x-\zeta,y-\eta)f(\zeta,\eta)d\zeta d\eta = f(x,y). \tag{4.13}$$

The multiple scales algorithm therefore filters the intensity signal $S(x,y)$ through $-\nabla^2 G$ (the DOG operator), thresholds the result, and sums it over a continuum of scales to recover $L(x,y)$:

$$\int_0^\infty dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T[\nabla^2 G(x-\zeta, y-\eta; t) S(\zeta,\eta)] d\zeta d\eta \; = L(x,y) \qquad (4.14)$$

Discrete sums over less than ten scales of $-\nabla^2 G$ yield good approximations to the several simple functions f that we have tested. We have not implemented the algorithm on the more complicated functions necessary to test it as a method for lightness recovery. Hence it is presented here largely for academic interest, but may yet prove useful in machine vision applications.

4.1.4 Crick's Edge-Operator Algorithm

Crick's solution[4] to the lightness problem is formulated for a Mondrian under uniform illumination surrounded by a border of constant reflectance. Under these special conditions, the edge-operator formula recovers the lightness function solely from information at the edges between patches of constant reflectance. The formula is the two-dimensional analog of Gauss' integral, which, for a function F that is constant within patches and zero on the boundary, makes the following statement:

$$F(x,y) = \frac{1}{2\pi}\int_\epsilon dF \frac{\mathbf{n}\cdot\mathbf{r}}{r^2} ds \qquad (4.15)$$

where ϵ indicates that the integral is performed over all edges in the image, dF is the difference of F across each edge (always taken in the decreasing direction), $\mathbf{n}$ is the normal to the contour s along which the integral is evaluated, and $\mathbf{r}$ is the distance from the point (x,y) to the contour.

Lightness is obtained using Equation (4.15) by setting $F = S$. This method of solution makes particularly clear the problem of accurate normalization of the computed lightness. Under the special condition that the illumination is uniform

[4] F. C. Crick, personal communication.

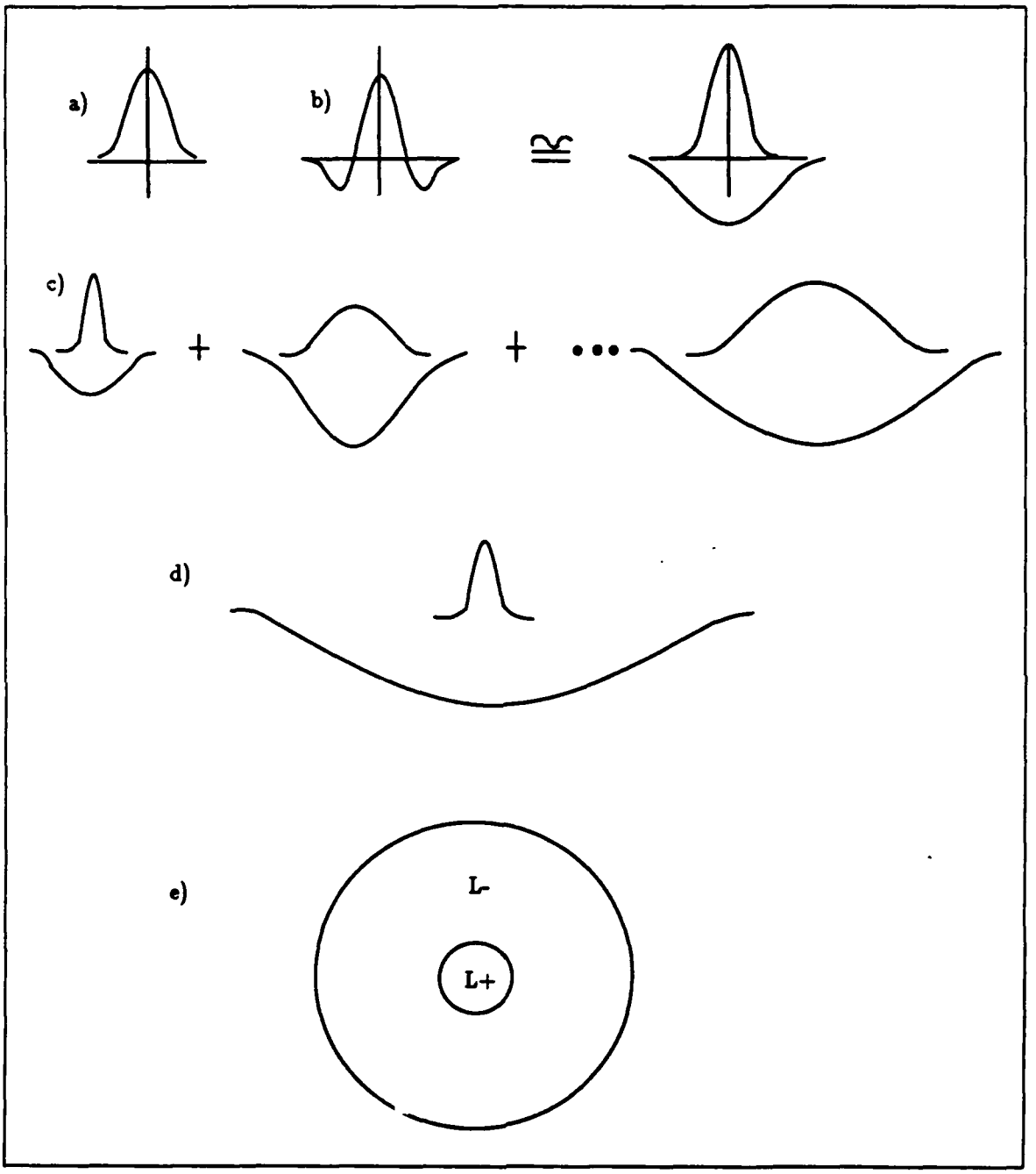

Figure 4.1: (a) The Gaussian distribution in one dimension. (b) The Laplacian of the Gaussian is approximately equal to the difference of two Gaussians of different scales. (c) The sum over multiple scales of DOG's yields (d) the difference of the smallest scale Gaussians and the largest scale Gaussian. Physiologically, the sum would yield a "lightness" neuron (e) with a narrow receptive field center and a large surround.

on the scene, its contribution to dS at edges cancels, and the integral over edges recovers $\tilde{\rho}$. Because the difference at an edge is ultimately referred to the value of $\tilde{\rho}$ on the boundary, this value must be added to the result of the integral in order to normalize it. Therefore, as in Horn's algorithm, the condition that $\tilde{\rho}$ be constant on the boundary is crucial in obtaining an accurate result. Yet because of the lightness ambiguity discussed in section 3, it is impossible to recover the boundary value of $\tilde{\rho}$ exactly: only the value of $S = \tilde{E} + \tilde{\rho}$ on the boundary is known. Because $\tilde{E}$ is assumed constant, the value of S on the boundary may be used as the normalization term, ensuring that lightness will be everywhere proportional to reflectance by the same factor.

When a gradient of illumination falls on the scene, if it is slow enough, it has no effect on values of dS at edges, but if it falls on the boundary region it violates the condition that S be constant there, and therefore $\tilde{\rho}$ will be inaccurately recovered.

The distinguishing feature of Crick's formula for lightness is the use of lightness information from the edges only. In summing the contributions only from contours across which the intensity signal changes sharply, the formula performs the differentiation and thresholding of lightness algorithms in one step rather than two. It is clear that in some sense, Land's retinex algorithm performs a similar integral, because when the threshold is appropriately set only the edges contribute to the lightness calculation. The similarity between these algorithms is made formally explicit in section 4.2.

4.1.5 Blake's Algorithm

Blake [7] proposes a modification of Horn's lightness algorithm based on the following criticism, hinted at in the previous section: The use of the Green's function convolution in solving for L in Equation (4.3) relies crucially on the condition that $\tilde{\rho}$ is constant on the boundary, which is not always met. Blake argues that many images (Mondrian or natural) are bordered by surfaces of varying reflectance. For these images, the solution obtained by using Equation (4.3) yields an inaccurate reconstruction

of lightness.

The inaccuracy results from the fact that the Poisson equation requires only that $\nabla^2 L = T[\nabla^2 \tilde{\rho}] = 0$ inside surface patches. Thus any harmonic function, e.g. a linear or sinusoidal function, in addition to a constant function, will solve the equation. A solution in which the reflectance function varies linearly across patches, for example, would be forced on data from a scene with a non-uniform boundary, in order to match it to inappropriate boundary conditions. If, instead, the condition is imposed only that $T[\nabla^2 S]$ is zero on the boundary, then the solution to the lightness problem will not be unique; that is, both accurate and inaccurate reconstructions may be obtained.

A stricter constraint restricting the lightness function to be constant within patches is more appropriate but would, in general, not admit any solution when Horn's boundary conditions do not apply to the data.

In Blake's lightness algorithm, the problem of the boundary conditions is obviated by using the gradient operator $\overline{\nabla}$ instead of the Laplacian to perform the spatial differentiation step. The reconstructed lightness function must therefore satisfy the equation:

$$\overline{\nabla} L = T[\overline{\nabla} S] \tag{4.16}$$

and is solved for by the path integral

$$L(x,y) = \int_P T[\overline{\nabla} S] ds, \tag{4.17}$$

where P is a path connecting (x_0, y_0), at which the value of L is arbitrarily defined, and (x, y). If

$$curl\ (T[\overline{\nabla} S]) = 0,$$

as for a Mondrian, then the value of the integral depends only on (x_0, y_0) and (x, y) and is independent of the path taken between them. This method represents a formalization of Land's algorithm for a Mondrian.

In the discrete case, an iterative solution may be performed that is similar to Horn's, but differs in the location of the threshold operation. It is clear that Blake's use of the gradient operator allows retention of information about linear gradients in the reflectance function, which in Horn's method is discarded.

4.2 Formal Connections Between Lightness Algorithms

The algorithms described in the above section employ different methods to follow similar steps in the computation of lightness. In this section, the algorithms are put into a single compact mathematical form that clarifies the similarities between them. The demonstration of formal equivalences between lightness algorithms allows their further evaluation to be based entirely on questions of implementation rather than computation.

4.2.1 Green's Theorem and Lightness Algorithms

The full solution to the lightness problem, as stated by the Poisson equation (Eq. 4.2), is given by Green's theorem, which expresses the relationship between the surface and line integrals of a scalar function. The symmetric Green's formula is:

$$\int\int_{\Sigma}(\phi\ \nabla^2 g\ -\ g\nabla^2\phi)d\Sigma\ =\ \oint_{C}(\phi\partial_n g\ -\ g\partial_n\phi)ds \tag{4.18}$$

where ϕ is a scalar function defined in a region Σ enclosed by a contour C, ds is an infinitesimal element of the contour, and g is the chosen Green's function [56].

This formula may be used to solve for lightness in Equation (4.2) by making appropriate substitutions, and by specifying appropriate boundary conditions: $\phi\ =\ L(\zeta,\eta)$ and $g\ =\ \frac{1}{2\pi}\ln r$, where $r^2 =\ (\zeta - x)^2 + (\eta - y)^2$.

A unique solution is obtained if one of two boundary conditions is met: in the region beyond Σ, *either* (1) L is uniform *or* (2)$\partial_n L\ =\ 0$ [56].

The following identities may now be derived: (a) Because $\nabla^2 g$ is zero except at the origin (x, y), where it equals 2π (proved by using Green's theorem),

$$\int\int_\Sigma (\phi \nabla^2 g) \;=\; \frac{1}{2\pi} \int\int_\Sigma L(\zeta,\eta)(\nabla^2 \ln r) dS \;=\; L(x,y).$$

(b) Because, under the boundary condition, $\partial_n L \;=\; 0$ outside S,

$$\oint_C g \partial_n \phi \; ds = \frac{1}{2\pi} \oint_C (\ln r) \partial_n L(\zeta,\eta) ds \;=\; 0.$$

Therefore the full solution to the lightness problem under the boundary conditions specified above is

$$L(x,y) \;=\; \frac{1}{2\pi} \int\int_\Sigma (\ln r) \; * \nabla^2 L(\zeta,\eta) d\zeta d\eta \;+\; \frac{1}{2\pi} \oint_C L(\zeta,\eta) \frac{\mathbf{n}\cdot\mathbf{r}}{r^2} ds \qquad (4.19)$$

$$\textbf{Term 1} \qquad + \qquad \textbf{Term 2}$$

where C is a closed contour chosen to lie in the boundary region beyond any edges in the image so that L is defined along it, $\mathbf{n}$ is the normal to C, and $*$ indicates a convolution.

The following discussion shows that Term 1, which sums a local spatial derivative over the image region, yields the three different methods of computing lightness found in Land's, Horn's and Crick's algorithms. Term 2, which depends on the value of lightness in the boundary region, normalizes the lightness computed by Term 1. If the lightness on the boundary is specified exactly, and $\nabla^2 L$ in the image region is provided as data, then the formula yields a unique solution to the lightness problem.

4.2.2 The Normalization Term

In any application of this formula, one of the two boundary conditions listed above must be met. If $L(\zeta,\eta)$ is uniform in the boundary region, then Term 2 is constant, and, if unknown, may be reset arbitrarily, simply to offset lightness. If L varies along C, so that $\partial_n L(\zeta,\eta) \;=\; 0$ on C, then Term 2 will vary for each (x, y). In this case, if

Term 2 is arbitrarily set to a constant, the reconstructed lightness function in general will not resemble the true reflectance function. This point is effectively the same that Blake makes, using a different argument (see section 4.1.5).

In Land's normalization scheme, discussed below, $L(\zeta, \eta)$ is allowed to vary along C, but in such a way that Term 2 is, nevertheless, approximately equal for each (x, y).

4.2.3 Horn's Method

Term 1 is exactly Horn's solution to the lightness problem, under boundary condition 1. In Horn's normalization scheme, Term 2 is set arbitrarily to a constant (zero).

4.2.4 Crick's Method

Crick's solution to the lightness problem is obtained by integrating Term 1 by parts, under boundary condition 1. For a general (non-Mondrian) two-dimensional image, this technique yields

$$\text{Term 1} = \frac{1}{2\pi} \int\int_{S-\epsilon} (\ln \frac{1}{r}) \nabla^2 L(\eta, \zeta) d\eta d\zeta + \frac{1}{2\pi} \oint_{\epsilon} \partial_n L(\zeta, \eta) \frac{\mathbf{n} \cdot \mathbf{r}}{r^2} ds \qquad (4.20)$$

where ϵ labels the contours that correspond to intensity edges, and $S - \epsilon$ labels the remaining surface on which the intensity function is smooth. On a Mondrian image, the integral over $S - \epsilon$ vanishes, because $\nabla^2 L$ is zero everywhere except at borders between patches. The integral over edges is exactly Crick's edge-operator formula. Term 2 of Equation (4.19) again provides the essential normalization constant. As discussed in section 4, this term effectively adds to the first term the constant value of lightness on the boundary to which the lightness values in the image have been referred.

For a single uniform patch on a uniform background, the integral over $S - \epsilon$ again vanishes, and the lightness solution becomes

$$\frac{1}{2\pi} \oint_C \partial_n L(\zeta, \eta) \frac{\mathbf{n} \cdot \mathbf{r}}{r^2} ds + \frac{1}{2\pi} \oint_C L_C(\zeta, \eta) \frac{\mathbf{n} \cdot \mathbf{r}}{r^2} ds \qquad (4.21)$$

where the integral is evaluated around the one contour that encloses the point (x, y), $\partial_n L(\zeta, \eta)$ is the difference in L taken from the inside to the outside across the contour, and $L_C(\zeta, \eta)$ is the value of L on the boundary. Thus the equation reduces to

$$L(x,y) = \frac{1}{2\pi} \oint_C L_I \frac{\mathbf{n} \cdot \mathbf{r}}{r^2} ds \tag{4.22}$$

where L_I is the value of L anywhere inside the contour. Equation 4.22 is Gauss's integral.

In chapter 8, the physiological implications of this scheme are discussed.

4.2.5 Land's Method

Land's retinex formula may also be derived from Term 1 of equation 4.19, by writing ∇^2 in polar coordinates and again integrating by parts. Term 1 then becomes:

$$\textbf{Term 1} = -\frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{R_\theta} \frac{dL}{dr} dr = \frac{1}{2\pi} \int_0^{2\pi} d\theta (L(x,y) - L_{R_\theta}) \tag{4.23}$$

where R_θ is the radius to the edge of the region S at angle θ, when (x, y) is take as the origin. In deriving equation 4.23, boundary condition 2 is used to make the assumption that as $r \mapsto \infty$, $\ln r$ approaches ∞ much more slowly than $\frac{dL}{dr}$ approaches zero.

Written in polar coordinates, Term 2 becomes:

$$L_N(x,y) = \frac{1}{2\pi} \oint_C L(r,\theta) d\theta = \frac{1}{2\pi} \oint_C L_{R_\theta} d\theta$$

The sum of terms 1 and 2 thus yields $L(x, y)$, trivially.

The first part of Equation 4.23 is a new formal expression of Land's method: $L(x, y)$ is given by integrating the radial gradient of L along a radial path, starting at (x, y) and ending at the edge of the image, and then summing the integral over all paths. In the actual implementation of the method, the gradient is sampled and thresholded at discrete intervals, summed over each path, and then averaged over a

finite number of paths. Each path starts at (x, y), but may end anywhere in the image. Thus the result is the lightness at (x, y) relative to the average value of lightness in the entire image.

The retinex algorithm does not normalize this result by adding to it the average value of lightness in the image but instead relies on lightness assumption 3, which states that the average value of lightness is invariant for all scenes. Lightness is therefore simply offset by a constant that is the same for each scene.

In effect, the retinex algorithm assumes that the lightness averaged over the endpoints of paths is equal to L_N, the lightness averaged over the contour enclosing the image. It is important to note that each of these averages is *weighted*: the contribution of L_{R_θ} to L_N is weighted by the angle θ subtended by its part of the contour at the origin (x, y), and the contribution of $L(r, \theta)$ to the retinex average is weighted by the number of endpoints falling on it. If the image contains a fully random collection of colors, and a large number of paths is chosen (> 200), then for most points in the image the two normalization terms should be approximately equal. (Land's "alternative" retinex method for computing lightness, recently published [60] specifically normalizes the flux at a given point with a similar weighted average over the entire field.)

4.2.6 Spatial Integration and Temporal Iteration

The multiple scales solution and Horn's iterative solution to the inverse Laplacian are formally equivalent. This statement is proved by putting each into the form required for Liouville-Neumann substitution:

$$f = F + \mathbf{H}f$$

which is solved by making successive substitutions of $(F + \mathbf{H}f)$ for f in the term $\mathbf{H}f$, yielding the series:

$$f^{(n)} = F + \mathbf{H}F + \ldots + \mathbf{H}^{n-1}F$$

where F (a function), and $\mathbf{H}$ (a matrix or operator) are known. If in the integral in Equation 4.14 we set apart the first term due to a Gaussian of scale 0 (the delta function), then we may write $F = \Delta t\ T[\nabla^2 S]$, $f = L$, and $\mathbf{H} = G(\Delta t)$ where $G(\Delta t)$ is the Gaussian kernel of scale Δt. The multiple scales integral may then be expressed as a series, the nth term of which is

$$L^n(x,y) =$$

$$\Delta t \left\{ L''(x,y) + G(\Delta t) * L''(\zeta,\eta) + G^2(\Delta t) * L''(\zeta,\eta) + \ldots G^n(\Delta t) * L''(\zeta,\eta) \right\} \quad (4.24)$$

where $L'' = T[\nabla^2 S]$ and $*$ indicates the convolution.

Horn's iterative solution may also be expressed as a series, the nth term of which is:

$$L^n_{ij} = \frac{1}{2} \left\{ L''_{ij} + \mathbf{G} * L''_{kl} + \mathbf{G}^2 * L''_{kl} + \ldots \mathbf{G}^n * L''_{kl} \right\} \quad (4.25)$$

where $\mathbf{G}$ is the discrete approximation to the Gaussian in matrix form, and L''_{ij} is $T[\nabla^2 S_{ij}]$ as defined in Equation (4.4).

The two series are therefore equal to within a multiplicative constant. In practice, they might each be noise-sensitive as discussed above, and might not converge quickly to the correct solution, although Horn [43] has successfully implemented his method. The operator G (or matrix $\mathbf{G}$) grows with each iteration to involve finally the entire field, and thereby mediates the long-range effects represented explicitly by Land's paths.

4.3 A Regularization Framework for Lightness Algorithms

Equation 4.1 is impossible to solve in the absence of additional constraints: there are twice as many unknowns as equations for each location $\mathbf{r}$. Formally, it falls into the class of inverse problems summarized by the equation:

$$y = Az \tag{4.26}$$

where A is a known operator. (Here S corresponds to y, z corresponds to $\tilde{\rho}$ and A corresponds to the linear operator that sums $\tilde{\rho}$ and $\tilde{E}$ to give S.) The direct problem is to determine y given z. The inverse problem is to find z, given y – the function y becomes the "data" and z the "solution". Although the direct problem is usually well-posed, the inverse problem is often *ill-posed.* Equation 4.1 represents one such ill-posed problem.

Standard regularization theories for "solving" ill-posed problems have been developed by Tikhonov [104] and others. The basic idea underlying regularization techniques is to restrict the space of acceptable solutions by choosing the one solution that minimizes an appropriate functional. Among the methods that can be employed (see [93]), the main one is to find z that minimizes

$$\|Az - y\|^2 + \lambda\|Pz\|^2. \tag{4.27}$$

The choice of the norm $\|\cdot\|$, usually quadratic as in Equation 4.27, and of the linear *stabilizing functional* $\|Pz\|$, is dictated by mathematical considerations, and most importantly, by a physical analysis of the generic constraints on the problem. The regularization parameter λ controls the compromise between the degree of regularization of the solution and its closeness to the data.

To employ standard regularization techniques, one must formulate the physical constraints on the recovery of $\tilde{\rho}^i$ from S^i in terms of a quadratic functional. The constraints we would like to enforce are: *the single source assumption* and the *spectral* and *spatial regularization* constraints.

Assume for simplicity that the signal S is given on an array indexed by a one-dimensional variable x. The single source assumption says that:

$$E^i(x) = k^i E(x) \tag{4.28}$$

or that

$$\tilde{E}^i(x) = \tilde{k}^i + \tilde{E}(x). \tag{4.29}$$

This assumption reduces the number of unknowns at each location x.

The spectral regularization constraint says that most illuminants and reflectances have a finite (and small) number of degrees of freedom. In other words, as in the primary assumption underlying spectral-basis algorithms (see Section 3.4), most illuminants and reflectances may be fully described as linear combinations of a fixed set of basis functions.

The spatial regularization constraint formalizes and extends the lightness assumptions that (a) $\rho^i(x)$ is either constant or changes sharply at boundaries between different materials, and (b) $E(x)$ is either constant or changes more smoothly than $\rho^i(x)$ across space. Whereas Horn's algorithm, for example, imposes the strong constraint (in two-dimensions) that all values of $\nabla^2 E(x,y)$ strictly below a fixed threshold are due to $E(x,y)$, the regularization assumption requires only that $E(x)$ vary less sharply across space than $\rho^i(x)$ and effectively allows the limit on the spatial variation of $E(x)$ to be reset for each scene.

The spatial regularization constraint may be enforced on Equation 4.1 by requiring that its solution minimize the following variational principle:

$$\sum_i [S^i - (\tilde{\rho}^i + \tilde{E} + \tilde{k}^i)]^2 + \lambda[\frac{d}{dx}E]^2 + \beta[G * \tilde{\rho}^i]^2 + \gamma[\frac{d^2}{dx^2}\tilde{\rho}^i]^2, \tag{4.30}$$

where G is a gaussian filter with standard deviation σ, and λ, γ and β are parameters controlling the degree of regularization and its closeness to the data. The first term requires that the solution $(\tilde{\rho}^i + \tilde{E})$ be close to S^i. The second term enforces the constraint that the illumination vary smoothly across space. The third and fourth terms enforce the constraint that the reflectance function consist only of spatial frequency components. Smoothing the reflectance function insures that only the low to intermediate spatial frequencies that are passed by the Gaussian are minimized. The high spatial frequencies do not contribute to the smoothed reflectance function

derivative so they are not minimized. The fourth term requires that very high spatial frequencies of $\tilde{\rho}^i$ are also minimized, in order to eliminate noise.

Solving Equation 4.30 for its minimum in the Fourier domain reveals how the constraints may be applied in the simple form of a linear filter. Solving the Euler-Lagrange equations for Equation 4.30 yields the following result for the Fourier transform of $\tilde{\rho}^i$:

$$\tilde{\rho}^i(\omega) = \frac{\lambda\omega^2}{\lambda\omega^2 + [1 + \lambda\omega^2][\beta e^{-\omega^2\sigma^2} + \gamma\omega^4]} S^i(\omega) \tag{4.31}$$

In other words, $\tilde{\rho}$ (and similarly $\tilde{E}$) may be obtained by filtering S through a linear filter. The above expression clarifies the need for the fourth term in the variational principle of Equation 4.30. If γ were zero, the filter would approach unity for very high frequencies (both numerator and denominator would tend to infinity as ω^2), whereas for nonzero γ, the filter would tend to zero for very high frequencies (the denominator would tend to infinity as ω^4). For any γ, as ω approaches zero, the filter approaches zero. Figure 4.2 illustrates the behavior of the filter for several values of the parameters γ, β and σ.

The filter is thus a bandpass filter that cuts out intermediate spatial frequencies due to the effective irradiance and very high spatial frequencies due to noise and retains high spatial frequencies due to reflectance. It filters out the d.c. component, which must be reset by the normalization constant. The filter is similar to Land's most recent retinex operator [60], which divides the image irradiance at each pixel by a weighted average of the irradiance at all pixels in a large surround and takes the logarithm of that ratio to obtain lightness. Such a center-surround operator becomes a bandpass filter in the Fourier domain. The two operators are not identical, though. The regularizing filter acts on the logarithm of irradiance. The lightness value it computes is scaled by (roughly) the average value of the logarithm of irradiance at all pixels, not by the logarithm of the average value. The regularizing filter is therefore linear in the logarithms, whereas Land's is not. Yet the numerical difference between the outputs of the two filters is small in most cases (E. H. Land, pers. comm.), and

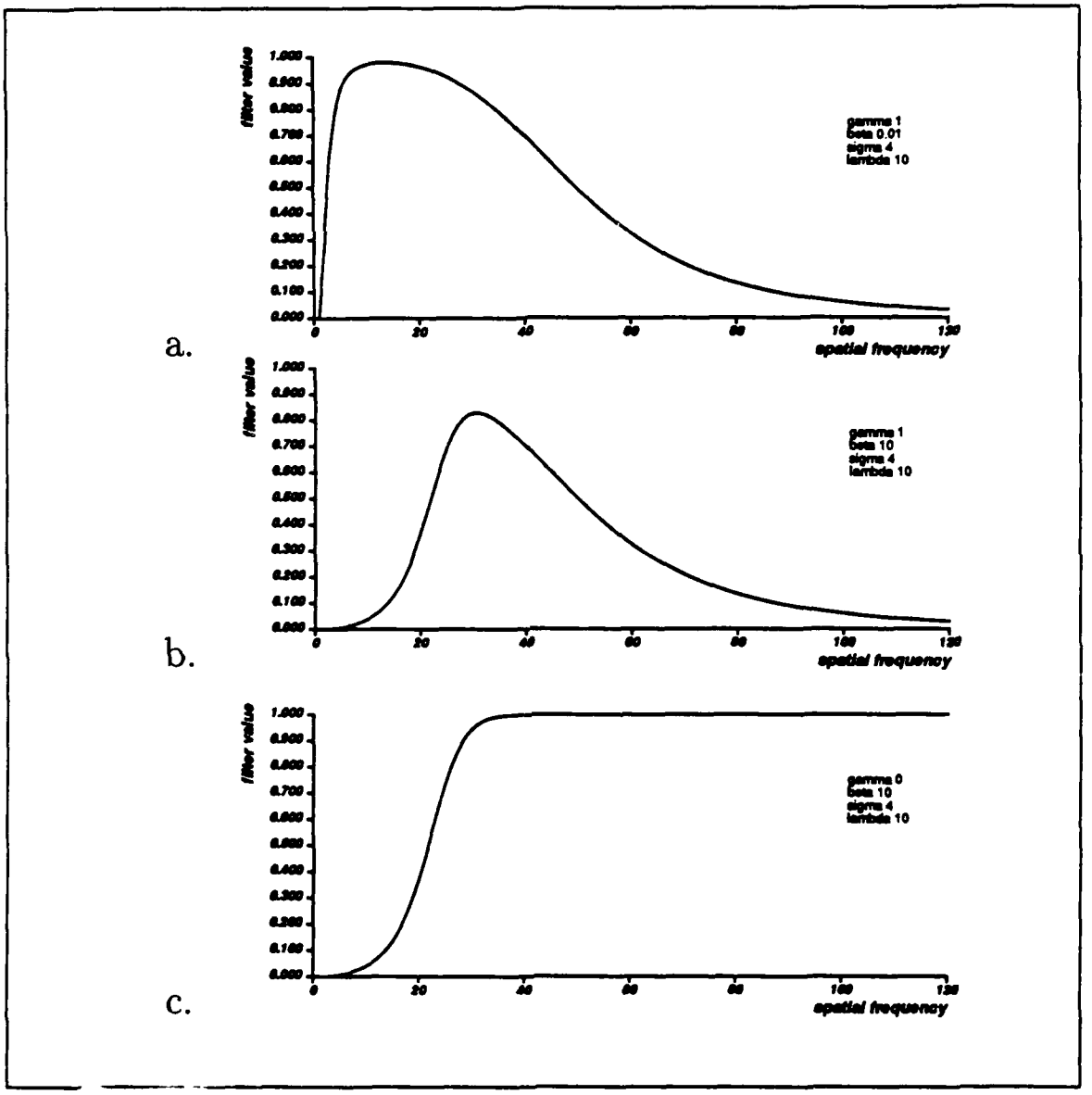

Figure 4.2: The filter that acts on $S^i(\omega)$ to yield $\tilde{\rho}^i(\omega)$, derived from the variational principle in Equation 4.30. (a) Parameters: $\gamma = 1$, $\beta = 0.01$, $\sigma = 4$, $\lambda = 10$. (b) Parameters: $\gamma = 1$, $\beta = 10$, $\sigma = 4$, $\lambda = 10$. (c) Parameters: $\gamma = 0$, $\beta = 10$, $\sigma = 4$, $\lambda = 10$. Note the behavior of the filter when γ goes to zero. β controls the extent to which low-frequency components of the irradiance signal are removed from the output reflectance.

both produce lightness values that correlate well with human perception.

The quadratic variational principle of standard regularization is a special case of a more general regularization scheme based on Markov random fields, sketched by Poggio and staff [95].

Chapter 5

Learning Lightness Algorithms

As demonstrated in Chapter 4, the lightness problem, to recover reflectance from image irradiance in a single chromatic channel, is ill-posed: the information supplied by the image is not sufficient in itself to specify a unique solution. The problem may be "solved" by using standard regularization techniques to enforce natural constraints. These constraints are explicitly crafted from an analysis of the physical properties of surfaces and lights. Yet how might a visual system learn and enforce similar constraints without *a priori* information? This question is important both for evolution of biological visual systems and for the construction of artificial ones. Biological visual systems evolved methods for extracting information from sensor signals without explicit knowledge of the constraints they could exploit, guided instead by how well the methods achieved their goals. If for color vision the goal was to find the cherry amongst the shadows and poisonous berries, then evolution selected the method that found the cherry, without "knowing" how it did it.[1]

With hindsight, we can discern the constraints that the optimal system, by ne-

[1] K.K. and R. L. DeValois write[24]: "Consider the problem of finding the ripe cherries in a cherry tree. Without color vision, one would find that the irrelevant contours produced by shadowing effectively mask the regularity of form and brightness of the fruit. With color vision, however, the problem is readily solved. The color contrast produced by red cherries against the green leaves is quite obvious, despite the irregular and imperfectly correlated luminance contours." Color vision in the sense of merely being able to discriminate between wavelengths of reflected light is not sufficient to solve the task described here – the task requires image segmentation into regions of constant color and so requires the steps discussed in Chapters 3 and 6.

cessity, but without instruction, incorporated. Artificial visual systems can then reap the benefits of evolution's hard work and be programmmed directly with the constraints that biological systems discovered by trial and error. But how more efficient it would be if we could program our artificial visual system with a fast, compressed version of *how* to discover constraints. Then the visual system could adapt to any new environment it encounters, and more quickly than a Darwinian animal. This goal, to make machines learn, is one of the spurs on the current neural networks movement.

The goal of the work in this chapter is to explore the simplest possible learning techniques for artificial visual systems, using the lightness problem as a testbed. The lightness problem serves well here, because it is a prototype for other problems in early vision, and because regularization theory immediately suggests one effective learning technique, optimal linear estimation. Here optimal linear estimation is compared with two additional simple techniques: backpropagation (BP) on a two-layer network, and optimal polynomial estimation.

5.1 Associative Learning of Standard Regularizing Operators

Minimization of the regularization functional in Equation (4.27) corresponds to determining a regularizing operator that acts on the input data y and produces as an output the regularized solution z. Suppose now that instead of solving for z, the task is: given y and its regularized solution z, find the operator that effects the transformation between them. This section demonstrates that the regularizing operator can be synthesized by associative learning from a set of examples. The argument consists of two claims. The first is that the regularizing operator corresponding to a quadratic variational principle is linear. The second is that any linear mapping between two vector spaces may be synthesized by an associative scheme based on the computation of the pseudoinverse of the data.

5.1.1 Linearity of the regularized solution

The discrete form of Equation (4.27) is:

$$\|A\mathbf{z} - \mathbf{y}\|^2 + \lambda\|P\mathbf{z}\|^2, \tag{5.1}$$

in which $\mathbf{z}$ and $\mathbf{y}$ are vectors, A and the Tikhonov stabilizer P are matrices, A does not depend on the data, and $\|\cdot\|$ is a norm.

The minimum of this functional will occur at its unique stationary point $\mathbf{z}$. To find $\mathbf{z}$, one sets to zero the gradient with respect to $\mathbf{z}$ of Equation 5.1. The solution to the resulting Euler-Lagrange equations is the minimum vector $\mathbf{z}$:

$$(A^TA + \lambda P^TP)\mathbf{z} = A^T\mathbf{y}. \tag{5.2}$$

It follows that the solution $\mathbf{z}$ is a linear transformation of the data $\mathbf{y}$:

$$\mathbf{z} = L\mathbf{y}, \tag{5.3}$$

where L is the linear regularizing operator. (If the problem were well-posed, the operator L would equal simply the inverse of A. It is important to note that L may depend on the given lattice of data points. In particular, if the data are sparse, then the components of the vectors $\mathbf{z}$ and $\mathbf{y}$ must be chosen carefully so that the same set of spatial locations is represented in each.) Thus, variational principles of the form of Equation (4.27) lead to a regularized solution that is a linear transformation of the data.

5.1.2 Learning a linear mapping

Given that the mapping between a set of input vectors $\mathbf{y}$ and their regularized solutions $\mathbf{z}$ is linear, how does one solve for it? To start, we arrange the sets of vectors $\mathbf{y}$ and $\mathbf{z}$ in two matrices Y and Z. The problem of synthesizing the regularizing operator L is then equivalent to "solving" the following equation for L:

$$Z = LY \tag{5.4}$$

A general solution to this problem is given by

$$L = ZY^{+}, \tag{5.5}$$

where Y^{+} is the Moore-Penrose pseudoinverse of Y [1]. This is the solution that is most robust against errors, if Equation 5.4 admits several solutions. It is the optimal solution in the least-squares sense, if no exact solution to Equation 5.4 exists. The latter case is the one of interest here. To avoid look-up table solutions (which would result if the set of input vectors **y** were orthogonal to one another and spanned the space of all subsequent test vectors) the problem must be overconstrained; that is, the number of examples (columns of Y) must be larger than the rank of the matrix L. In this case, there is no exact solution to Equation 5.4 and the matrix L must be found instead by minimizing the expression

$$M = \|LY - Z\|^2. \tag{5.6}$$

L is then:

$$L = ZY^{T}(YY^{T})^{-1} \tag{5.7}$$

These results show that the standard regularizing operator L (parametrized by the lattice of data points) may be synthesized without need of an explicit variational principle, if a sufficient set of correct input-output pairs is available to the system. Note that by supplying as examples the physically correct solutions **z**, one assumes that they are identical to the regularized solutions **z**, and thereby enforces both regularization and correctness on the linear operator obtained.

5.2 Constructing a Lightness Operator By Optimal Linear Estimation

In chapter 4 we derived a regularizing functional for the lightness problem and extracted a linear filter from it. To solve the lightness problem by "learning," we start instead with the single assumption that there exists a linear operator that transforms $S^i(\mathbf{r})$ into $\tilde{\rho}^i(\mathbf{r})$, where $\tilde{\rho}^i(\mathbf{r}) = \log \rho^i(\mathbf{r})$. Using optimal linear estimation, we may then synthesize a linear operator from examples and examine the constraints it embodies.

Let $\mathbf{p}$ be the vector representing the discrete version of the function $\tilde{\rho}(x)$ in one dimension, $\mathbf{e}$ be the vector representing the discrete version of the one-dimensional function $E(x)$, and $\mathbf{s}$ be the vector representing the sensor response $S(x)$, the sum of $\mathbf{p}$ and $\mathbf{e}$ (dropping the superscript i for now). The "learning lightness" problem is to construct an operator L that provides the best estimation $L\mathbf{s}$ of $\mathbf{p}$. The technique is to find the linear estimator that best maps the input $\mathbf{s}$ into the desired output $\mathbf{p}$, in the least squares sense.

$\mathbf{s}$ and $\mathbf{p}$ may be thought of as vertical scan lines across a pair of two-dimensional Mondrian images: an input image of a Mondrian under illumination that varies smoothly across space and its desired output image, the Mondrian reflectance pattern alone (see Figure 5.1). The input "training vectors" $\mathbf{s}$ are generated by adding together different random $\mathbf{p}$ and $\mathbf{e}$ vectors, according to Equation 4.1. Each vector $\mathbf{p}$ represents a pattern of step changes across space, corresponding to one column of a (log) reflectance image. The step changes occur at random pixels and are of random amplitude between set minimum and maximum values. Each vector $\mathbf{e}$ represents a smooth gradient across space with a random offset and slope, corresponding to one column of a (log) illumination image. The training vectors $\mathbf{s}$ and $\mathbf{p}$ are then arranged as the columns of two matrices S and R, respectively. The goal is then to compute the optimal solution L of

$$LS = R \tag{5.8}$$

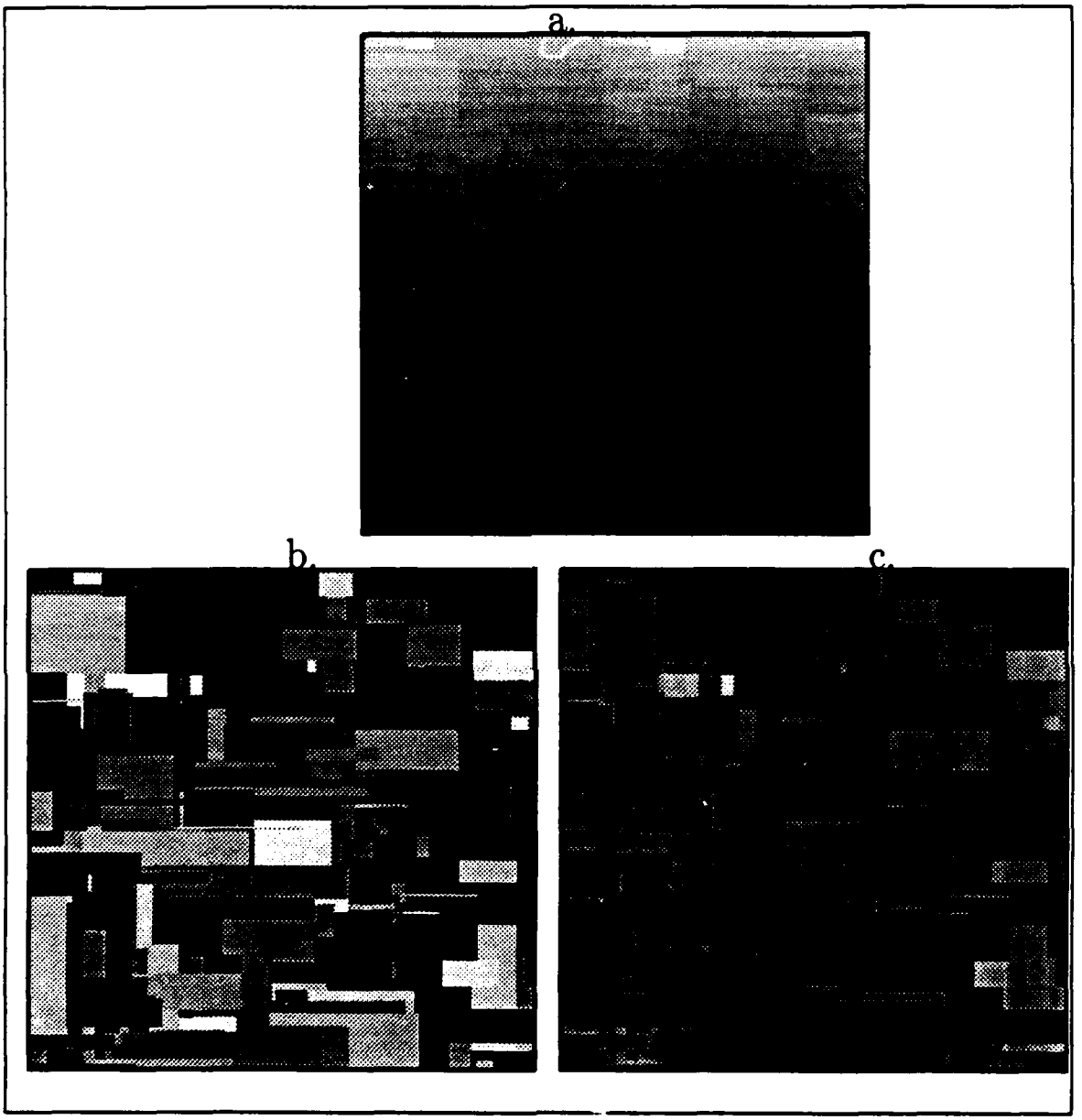

Figure 5.1: (a) A Mondrian under a linear illumination gradient, generated by adding together two 320x320 pixel images: one is the (log) reflectance image shown in (b), an array of rectangles each with a different, uniform grey-level; the other is the (log) illumination image, in which the pixel values increase linearly in the same way across each column. The Mondrian in (b) is equivalent to the Mondrian in (a) under uniform illumination. (c) The output of the linear operator when it acts on the Mondrian in (a).

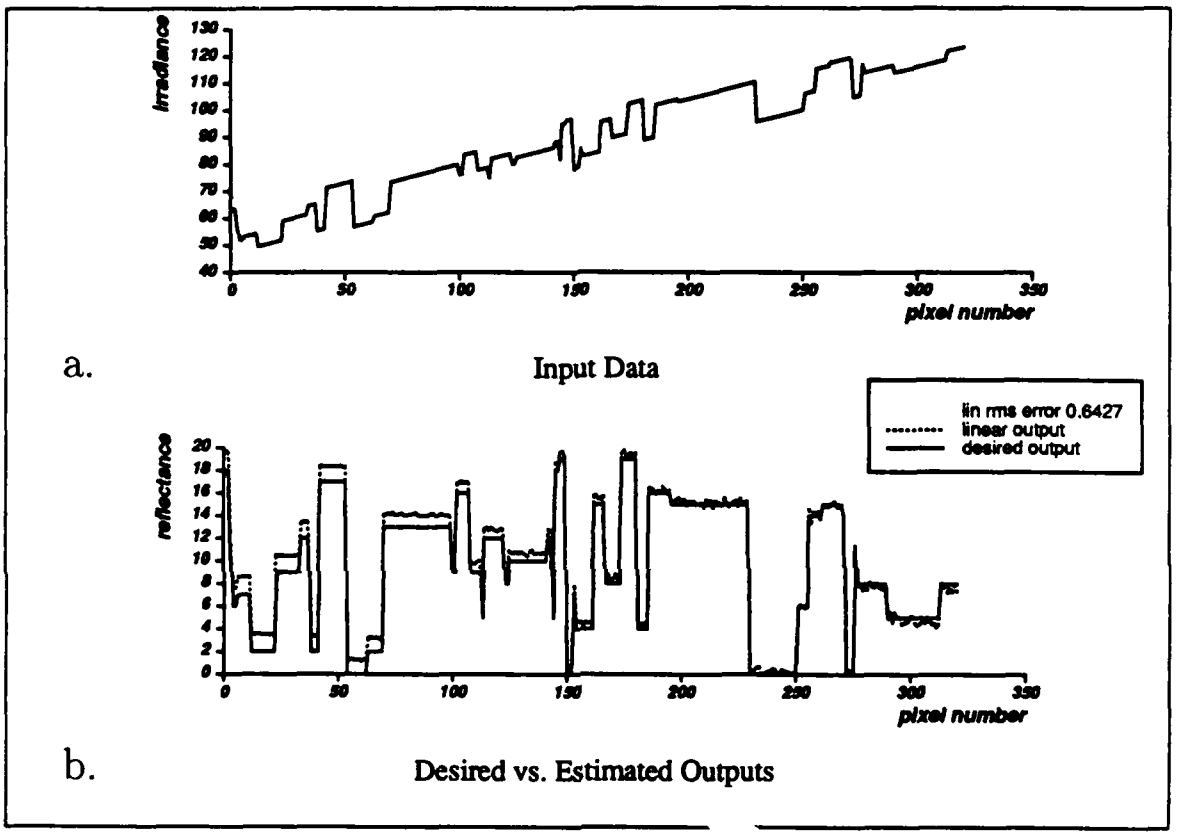

Figure 5.2: (a) Input data, a one-dimensional vector 320 pixels long. Its random Mondrian reflectance pattern is superimposed on a linear illumination gradient with a random slope and offset. (b) Corresponding output (dotted line) of linear operator trained on 1500 unrepeated examples with linear illumination gradients and 15 % edges (see text). Compare with correct solution (solid line), the reflectance pattern alone. The vector in (a) was not part of the training set. The rms error is the pixel-by-pixel root-mean-square error of estimated output compared to the desired output.

where L is a linear operator represented as a matrix.

Here we compute the pseudoinverse S^+ by overconstraining the problem – using many more training vectors than there are number of pixels in each vector – and using the formula that applies in the overconstrained case: $S^+ = S^T(SS^T)^{-1}$. To avoid numerical instability, we add a stabilizing functional to Equation 5.7:

$$L = RS^T(SS^T + \alpha\|SS^T\|)^{-1}$$

(The pseudoinverse may be computed by other methods based on singular value decomposition. It may also be computed by recursive techniques that improve its form as more data become available – see Appendix C. The latter procedure, although equivalent to our "one-shot" computing technique, may seem intuitively more like learning.)

The operator L computed in this way recovers a good approximation to the correct output vector $\mathbf{p}$ when given a new $\mathbf{s}$, not part of the training set, as input. Figure 5.2 shows the estimated output compared with the desired output, for a linear operator L trained on 1500 unrepeated examples of one-dimensional Mondrians under linear illumination gradients, in which the reflectance values are on average approximately 10 % of the total change in illumination across the vector, and at each pixel there is a 15 % chance of a reflectance edge occurring. For some test input vectors the prediction is even better, in others it is worse. A second operator, estimated in the same way, recovers the illumination e. Acting on a random two-dimensional Mondrian L also yields a satisfactory approximation to the correct output image (see Figure 5.1).[2]

What natural constraints does the linear operator embody? The answer lies in the structure of the matrix L. We assume that, although L is *not* a convolution operator, it should approximate one far from the boundaries of the image. That is, in its central part, the operator should be space-invariant, performing the same action

[2]Estimation of the operator on two-dimensional examples should be possible, but computationally very expensive if done in the same way. The present computer simulations require several hours when run on standard serial computers (Symbolics 3600 and 3650 LISP machines). The two-dimensional case will require probably much more time, unless run on a parallel machine.

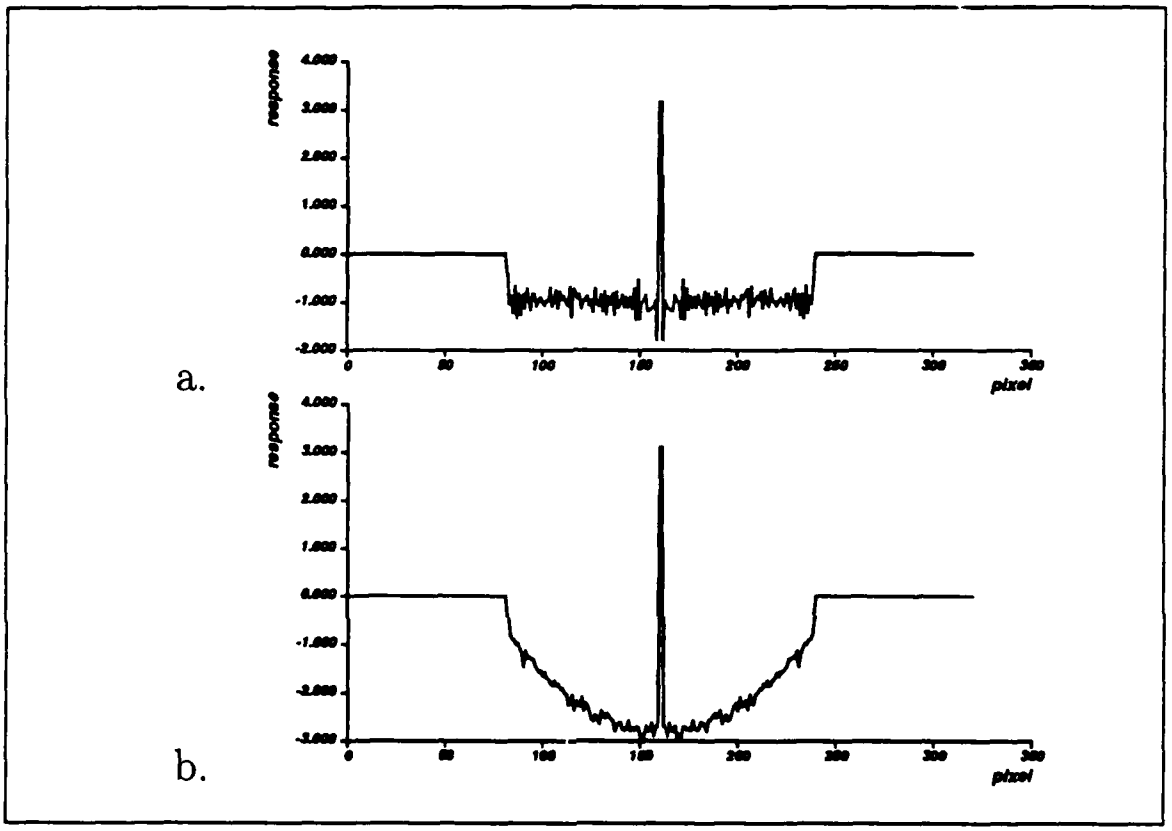

Figure 5.3: (a) Illustration of the linear operator that best maps, in the least squares sense, irradiance into reflectance, for input Mondrians under linear illumination gradients. The operator learned is a matrix that acts on one-dimensional vectors. The shape of the corresponding filter, shown here, is estimated by summating the central rows of the matrix, shifting them appropriately. The shallow surround extends but cannot be estimated reliably beyond the range shown here. (b) The filter estimated from a training set of vectors half of which had linear illumination gradients, the other half sinusoidally varying illumination gradients with random wavelength, phase and amplitude. The peak value of each filter has been scaled by 0.01 for convenience of display.

on each point in the image. Each row in the central part of L should therefore be the same as the row above but displaced by one element to the right. Inspection of the matrix confirms this expectation. To find the form of L in its center, we thus average the rows there, first shifting them appropriately. The result, shown in Figure 5.3(a), is a space-invariant filter with a narrow positive peak and a broad, shallow, negative surround. (We do not attempt to examine the structure of the surround beyond its central space-invariant range.)

(Because we start from the image irradiance equation in Equation 4.1, we assume that the illumination and reflectance vectors that are added together in the input training data are the logarithms of the "real" illumination and reflectance. Thus the "real" illumination gradients that we recover vary exponentially across space. When we use instead the logarithms of linear illumination gradients in the training data, we obtain a qualitatively similar receptive field.)

To observe how the shape of the estimated operator varies with the type of illumination gradient in the training set, we synthesize a second operator using a new set of examples. This set contains equal numbers of vectors with linear illumination gradients and vectors with slowly varying sinusoidal illumination components of random wavelength, phase and amplitude. Whereas the first operator, synthesized from examples with strictly linear illumination gradients, has a broad negative surround that remains virtually constant throughout its extent, the new operator's surround has a smaller extent and decays smoothly towards zero from its peak negative value in its center (see Figure 5.3(b)).

The form of the space-invariant filter is similar to that which results from the direct application of regularization methods exploiting the spatial constraints on reflectance and illumination described in Chapter 4. It also resembles the solution derived by posing the lightness problem as a Poisson equation. As noted in Chapter 4, the derived and estimated filters are similar to Land's most recent retinex operator [60], which divides the image irradiance at each pixel by a weighted average of the irradiance at all pixels in a large surround and takes the logarithm of that result to yield lightness. The

lightness triplets computed by the retinex operator agree well with human perception in a Mondrian world. The retinex operator and the matrix L both differ from Land's earlier retinex algorithms, which require a nonlinear thresholding step to eliminate smooth gradients of illumination.

The Fourier transform of the filter is approximately a bandpass filter that cuts out low frequencies due to slow gradients of illumination and preserves intermediate frequencies due to step changes in reflectance (see Figure 4.2). For linear gradients of illumination, which are effectively very low spatial frequencies, the surround is necessarily very broad in the space domain. When the spatial frequencies of illumination are increased, the surround shrinks, as expected. In contrast, the operator that recovers the illumination, e, takes the form of a low-pass filter. (Because of the initial log transformation on the image data, the estimated algorithm can be regarded as an example of homomorphic image processing.)

The shape of the linear lightness operator suggests another way in which simultaneous color (or brightness) contrast might be linked to color constancy. Here the lightness operator is "trained" to discount illumination gradients, the first step towards color constancy (spatial decomposition). Yet what results is an operator whose shape should naturally produce simultaneous contrast; a dark patch should appear darker when against a light background than against a dark one. The lightness operator trained on a mixture of illumination gradients has a negative surround which weights nearby pixels more heavily than distant ones, and thereby should produce local contrast effects. The lightness operator trained on linear gradients, on the other hand, will produce global contrast effects of the sort for which lightness algorithms are criticized (see Section 3.3.3). In fact, in the absence of normalizing factors, the output reflectance it computes for a patch of fixed input reflectance decreases linearly with increasing average irradiance of the input test vector in which the patch appears. That is, the lightness operator does enhance the contrast between a patch and its surround.

The linear operator of Figure 5.3(a) performs well on new test input vectors in

which the density and amplitude of the step changes of reflectance differ greatly from those on which it was trained. The operator performs well, for example, on an input vector representing one column of an image of a small patch of one reflectance against a uniform background of a different reflectance, the entire image under a linear illumination gradient. This result is consistent with psychophysical experiments that show that color constancy of a patch holds when its Mondrian background is replaced by an equivalent grey background [106].

The shape of the space-invariant portion of the linear operator, particularly of its large surround, is also suggestive of the non-classical fields that have been found in V4, a cortical area implicated in mechanisms underlying color constancy [27, 110, 116, 117] (see Chapter 8).

The technique of optimal linear estimation is closely related to optimal Bayesian estimation (see Appendix C). If we were to assume from the start that the optimal linear operator is space-invariant throughout its extent (which it is not), we could considerably simplify (and streamline) the computation by using standard correlation techniques ([1]; see Appendix C). Yet the resultant operator would probably not work as well as the one synthesized here.

5.3 Constructing a Lightness Operator by Backpropagation

Because the lightness problem as stated here is a linear problem, perhaps it is not too surprising that the solution provided by optimal linear estimation works so well. In the real world, the recovery of lightness from the image irradiance is not a linear problem, and one would not expect a linear operator to solve it successfully. We therefore explore nonlinear methods of constructing lightness operators from examples, in particular backpropagation and optimal polynomial estimation.

Backpropagation is gradient descent on a "neural" network with nonlinear units arranged in more than one layer [99]. We construct a two-layer backpropagation

percent edges	rms error	
	BP	lin
20	0.021	0.023
50	0.016	0.019
90	0.013	0.017

Table 5.1: Comparison of a BP net and estimated linear operator trained on the same set of 10000 examples under linear illumination gradients, with 85 percent edges. (In each example, reflectance edges occur at roughly 85 percent of the pixels.) Both the BP net and linear operator perform worse when the percent edges in the test vectors decreases. For each value of percent edges, the root-mean-square (rms) error (the pixel-by-pixel error between the actual and desired outputs) is calculated over 20000 test vectors.

network with 32 input units, each of which connects to each of 32 hidden units, each of which connects to each of 32 output units (see Figure 5.5). On each trial, each input unit receives the value of one pixel in a 32-pixel input vector. The hidden units and output units have sigmoidal nonlinearities. After many passes through a set of training vectors, the weights on the units stabilize to a configuration which, in the ideal case, minimizes the square error between the actual and desired outputs summed over the entire training set. (Note that performing gradient descent on a network with *linear* units is eqivalent to computing the regularized pseudoinverse (see Figure 5.4). Since the pseudoinverse is the unique best linear approximation in the L_2 norm, a gradient descent method that minimizes the square error between the actual output and desired output of a fully connected linear network is guaranteed to converge to the same global minimum.)

The backpropagation network requires an order of magnitude more time to converge to a stable configuration than does the linear estimator for the same set of 10000 32-pixel input vectors with linear gradients of illumination. The network's performance is slightly, yet consistently, better, measured as the root-mean-square error in output, averaged over sets of at least 20000 new input vectors (see Table 5.3). Interestingly, the backpropagation network and the linear estimator seem to err in

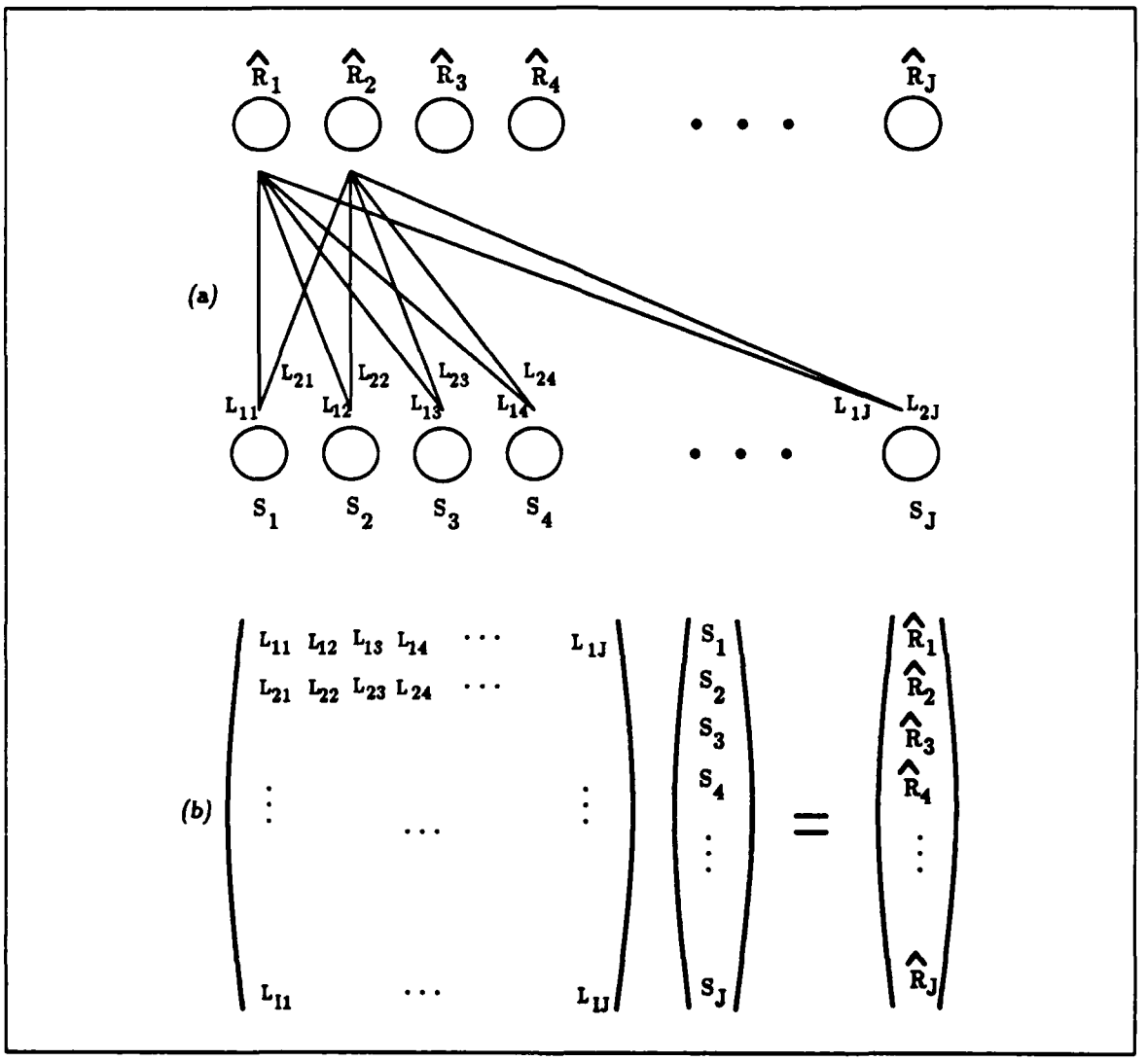

Figure 5.4: (a) Gradient descent performed on this network of linear units is equivalent to computing the regularized pseudoinverse. Each input vector in the training set, S_j $(j = 1...J)$, is fed to the network and its corresponding output vector $\hat{R}_j$ is computed using the current set of weights L_{ij}. The weights L_{ij} are changed in proportion to the size of the error between $\hat{R}_j$ and the correct output vector R_j. Over a finite number of iterations of each training vector in the set the weights converge to the matrix elements L_{ij} found by computing the regularized pseudoinverse shown in (b).

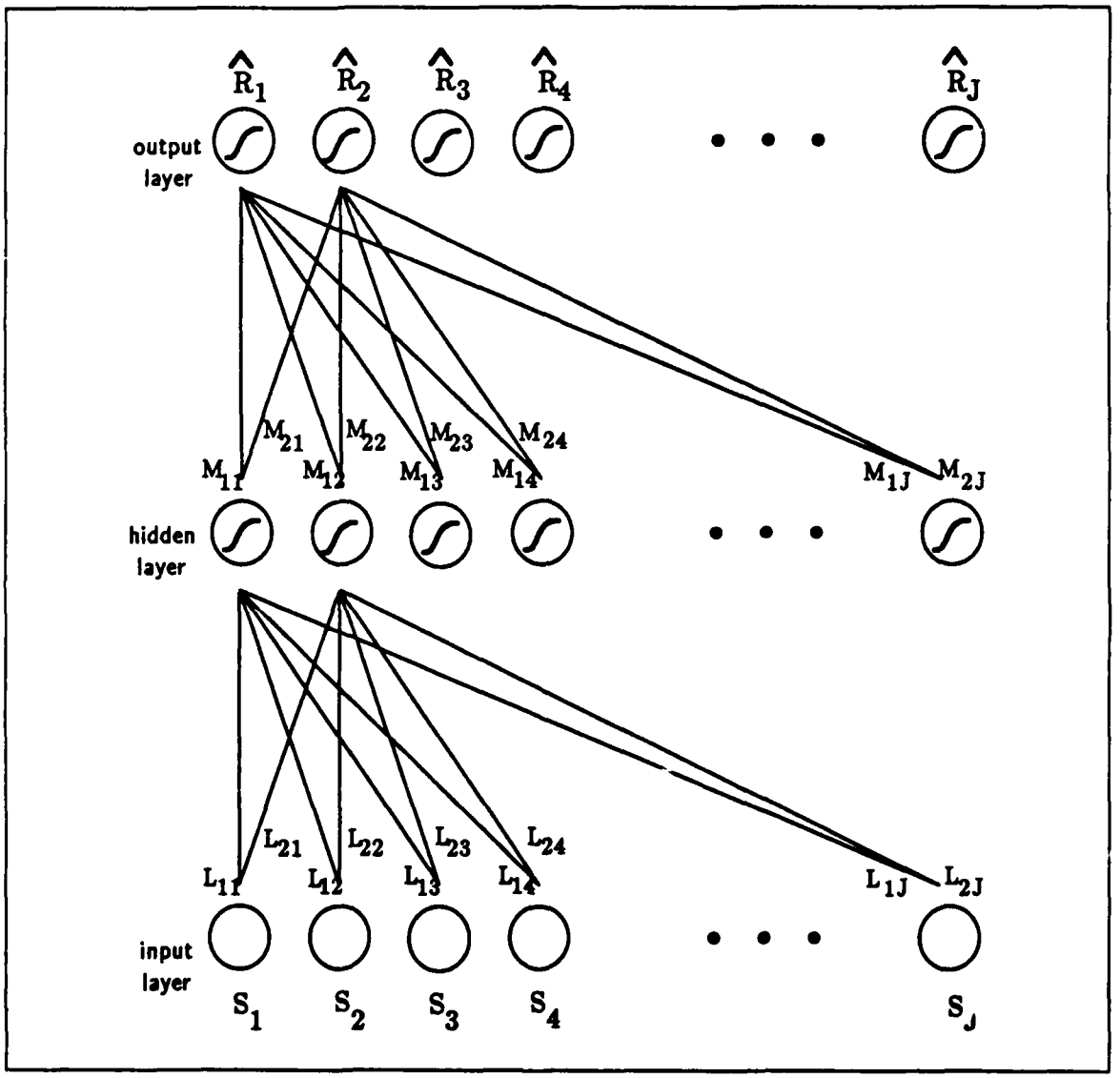

Figure 5.5: The backpropagation network. The outputs of the hidden and output layers are passed through a sigmoid function. As in gradient descent on the linear network in Figure 5.4, the error between $\hat{R}_j$ and the correct output vector R_j drives the changes made to the weights M_{ij} and L_{ij} on each iteration.

percent-edges	rms error	
	BP	lin
15	0.81	0.52
50	0.41	0.34
90	0.35	0.29

Table 5.2: Comparison of a BP net and estimated linear operator trained on the same set of 10000 examples under linear illumination gradients, with 85 percent edges, as described above. In this example, the BP net performs much worse than the estimated linear operator, because the configuration to which it has stabilized corresponds to a local rather than global minimum.

the same way on the same input vectors. Occasionally, the BP net does much worse than L, when, because it is not minimizing a quadratic functional, it falls into a local minimum during training (see Table 5.3).

We might expect the backpropagation network to perform significantly better than the linear estimator on tasks more complicated than the extraction of a linear illumination gradient. Accordingly, we train the BP net and a linear operator on the same set of 10000 input vectors representing Mondrians under *sinusoidally* varying illumination. The BP net again only slightly outperforms the linear operator (see Figure 5.6).

In the preceding examples, the input vectors are constructed by *adding* reflectance vectors to illumination vectors, on the assumption that a logarithmic transformation has been performed on the original intensity signal in which reflectance and illumination are multiplied. (Biological photoreceptors perform a similar transformation.) Trained on input vectors representing the *product* of illumination and reflectance, the linear operator performs slightly but consistently better than the backpropagation net (see Figure 5.7). This result is surprising since the task that the linear operator performs is no longer linear.

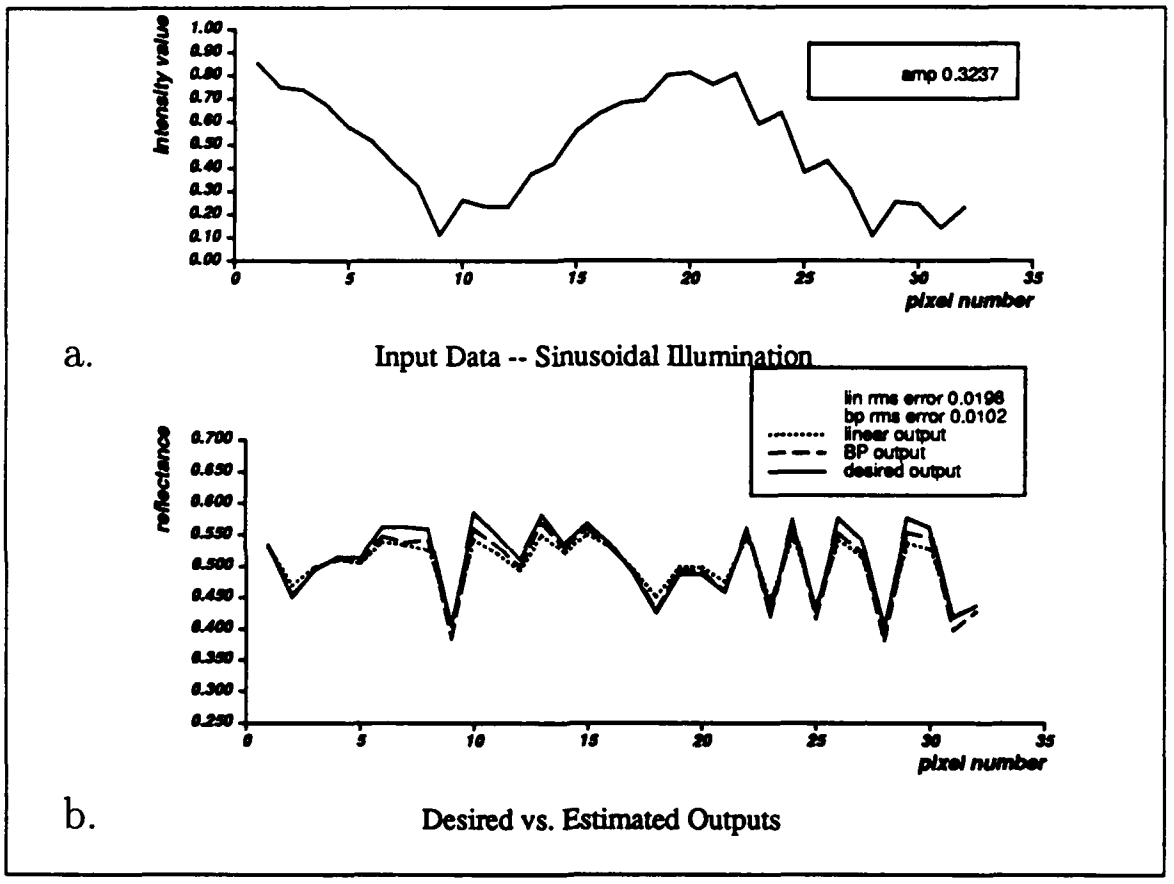

Figure 5.6: (a) Input data, a one-dimensional vector 32 pixels long, consisting of Mondrian reflectance pattern added to sinusoidally varying illumination gradient. (b) Outputs of linear operator (dotted line) and BP operator (dashed line) trained on the same set of 10000 examples under sinusoidal illumination gradients, of which (a) was not one, compared with correct output (solid line).

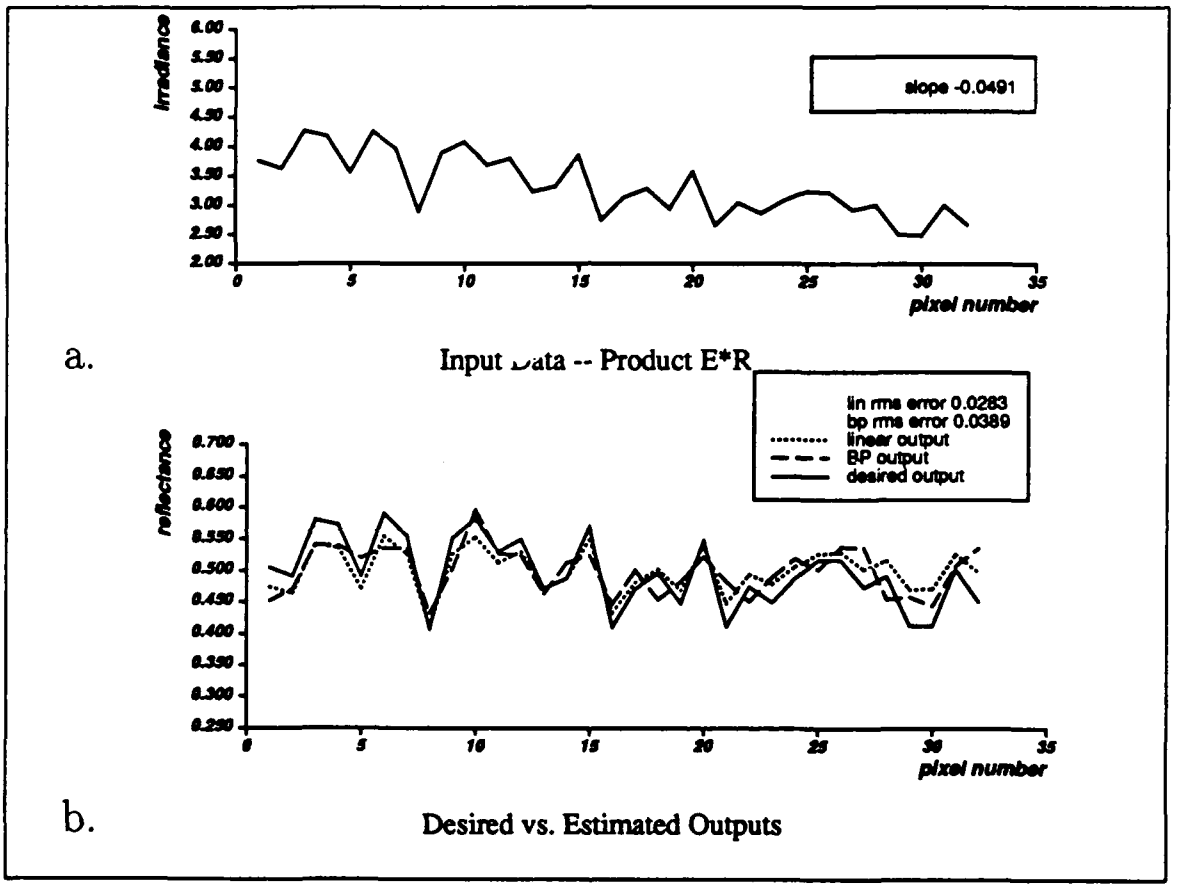

Figure 5.7: Input (a) and output (b) for linear (dotted line) and BP operators (dashed line) trained on the same set of 10000 examples representing the product of illumination and reflectance, compared with correct output (solid line). This BP net has stabilized to a global minimum.

5.4 Constructing a Lightness Operator by Optimal Polynomial Estimation

The natural extension of the optimal linear estimator is the optimal polynomial mapping. The Weierstrass-Stone theorem suggests that polynomial mappings, of which linear mappings are a special case, can approximate arbitrarily well all continuous real functions. We estimate the polynomial mapping of order two between input and output by estimating the optimal linear mapping for input vectors containing the original input components and all pairwise products between them. We compare this polynomial estimator with the BP net trained on the same set of original input vectors, with edge density 85 percent (each pixel in the reflectance vector has an 85 % chance of being at an edge). The polynomial estimator outperforms the BP net by a small margin that decreases as the edge density of the input decreases, as illustrated in Table 5.3. Figure 5.8 illustrates the performance of the BP net, the optimal polynomial estimator of order two, and the optimal linear estimator acting on a test input vector with a linear illumination gradient.

percent-edges	rms error	
	BP	poly
10	0.016	0.013
50	0.017	0.011
85	0.019	0.010

Table 5.3: Comparison of a BP net and optimal polynomial mapping of order two trained on the same set of 10000 examples under linear illumination gradients, with 85 percent edges, as described above. The rms errors are averaged over 2000 trials of new input vectors. This BP net has stabilized to a global minimum.

Recently, it has been demonstrated that a BP net can represent exactly any polynomial mapping, given polynomial output functions of at least order two for each unit and enough hidden units and layers ([85]). This statement, together with the Weierstrass-Stone theorem, implies that a BP net with sigmoid outputs and enough

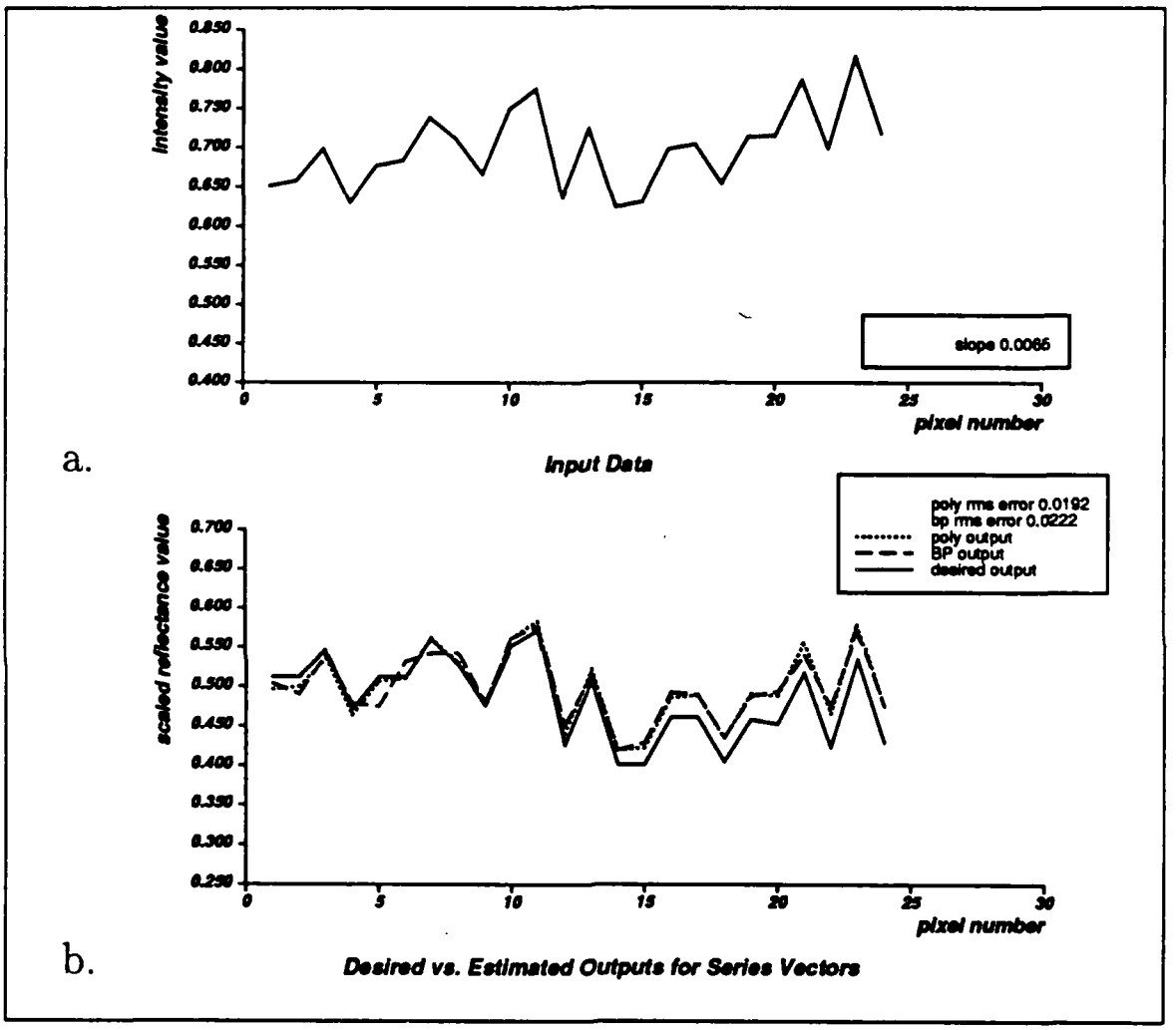

Figure 5.8: Input (a) and output (b) for BP operator and optimal polynomial mapping of order two trained on the same set of 10000 examples under linear illumination gradients. The polynomial operator was obtained by estimating the optimal linear operator for input vectors containing the original input components (only these were used for training the BP operator) and all pairwise products between them. This BP net has stabilized to a global minimum.

units and layers can approximate any real-valued continuous function. That the optimal polynomial mapping of order two performs better than does the two-layer BP net suggests that the net might need to be very much larger to approximate closely the lightness transformation. That the linear operator does almost as well as the other two suggests that it captures most of the essential features of that transformation.

5.5 Conclusions

The significance of these results lies in the facts that a simple statistical technique may be used to synthesize a lightness algorithm from examples and that a similar technique may be used for other problems in early vision. For the lightness problem, a two-layer BP net does not perform significantly better than the linear estimator under several different conditions, when : a) the input data is the *sum* of illumination and reflectance (or the *logarithm* of the product); b) the input data is the *product* of illumination and reflectance; c) the spatial dependency of the illumination is linear; and d) the spatial dependency of the illumination is sinusoidal. The BP net performs slightly worse than the optimal polynomial estimator of order two on input data with linear illumination gradients. The addition of nonlinear hidden units does not seem to improve performance, suggesting that compression of the input into a smaller set of important features would only degrade the performance of the linear operator.

All three operators synthesized here provide good solutions to the lightness problem. Yet as discussed in Chapter 3, lightness solutions cannot solve the full problem of color computation. Lightness operators by necessity will confuse shadows and specularities with reflectance edges if the former cause sufficiently sharp image irradiance changes. Figure 5.9 illustrates how the lightness operator synthesized by optimal linear estimation acts on a shadow. The Mondrian in Figure 5.9(a) is constructed in the same way as the Mondrian in Figure 5.2, with one linear illumination gradient applied uniformly to each column, except that each pixel value in the right half of the image is offset by a fixed amount. This simulates a shadow that satisfies the three

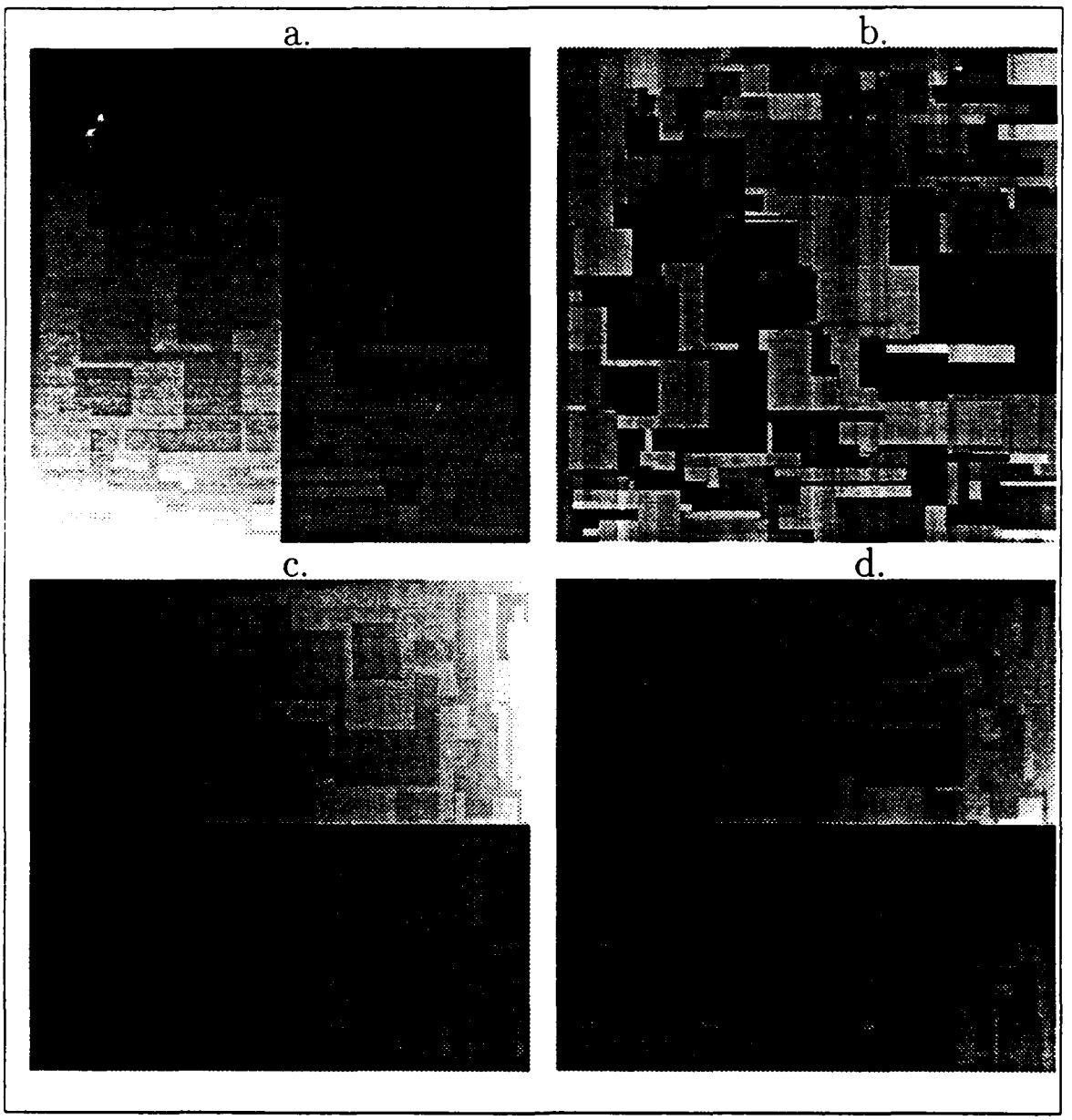

Figure 5.9: (a) Input Mondrian with vertical linear gradient of illumination extending across entire image. "Shadow" edge runs vertically down center of image. (b) Output of linear operator (shown in Figure 5.3) acting on (a). (c) Input Mondrian with horizontal linear gradient of illumination extending across entire image. "Shadow" edge runs horizontally across center of image. (d) Output of linear operator (shown in Figure 5.3) acting on (c). Notice that the operator interprets the shadow as a reflectance edge.

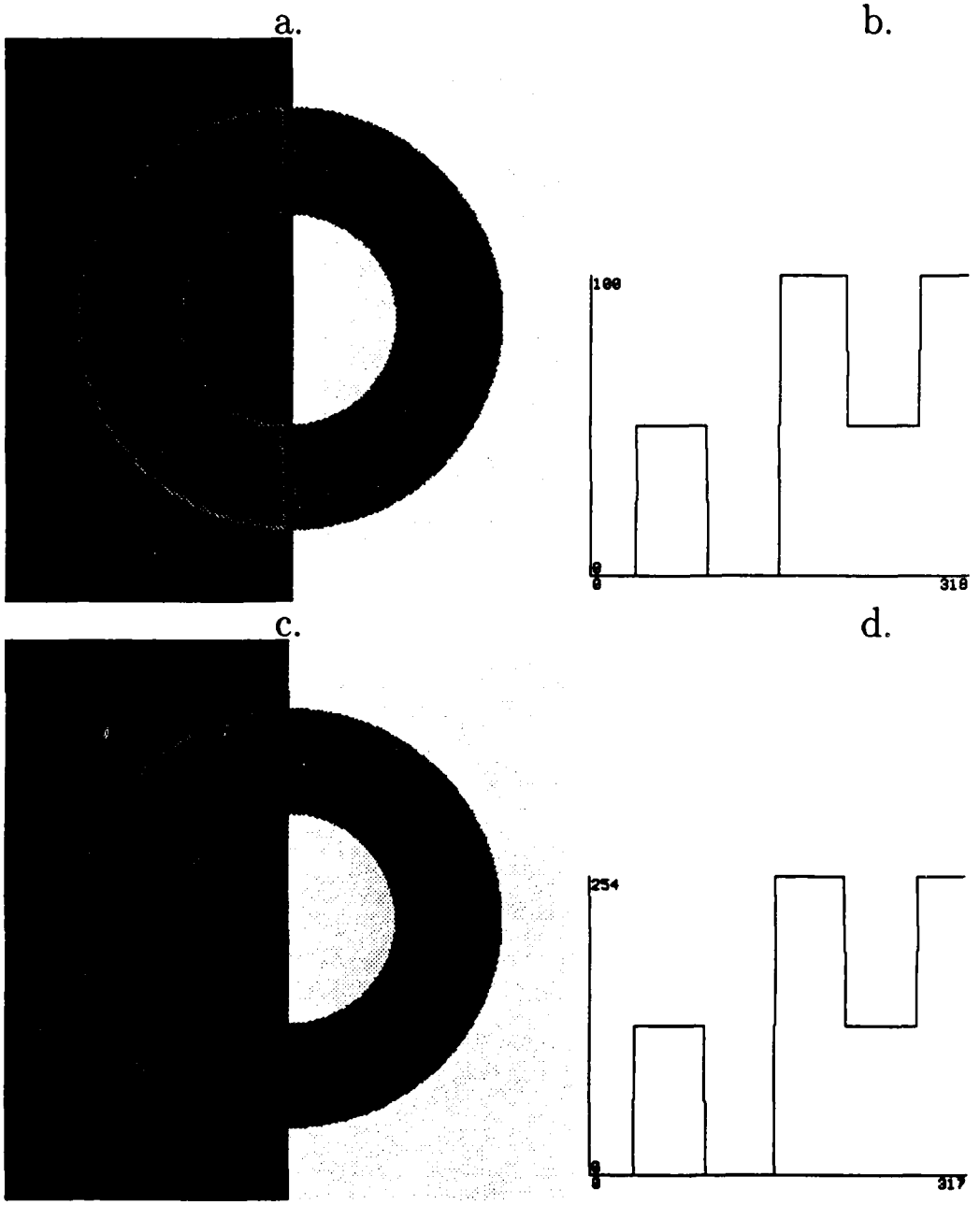

Figure 5.10: (a) Koffka Ring with midline drawn down center of uniform grey annulus against bipartite background. Left half of annulus appears lighter than right half. (b) Horizontal slice of (a) through center of ring. (c) Koffka Ring without midline drawn down center of annulus. Annulus now appears to be uniform grey throughout. (d) Horizontal slice of (c) through center of ring.

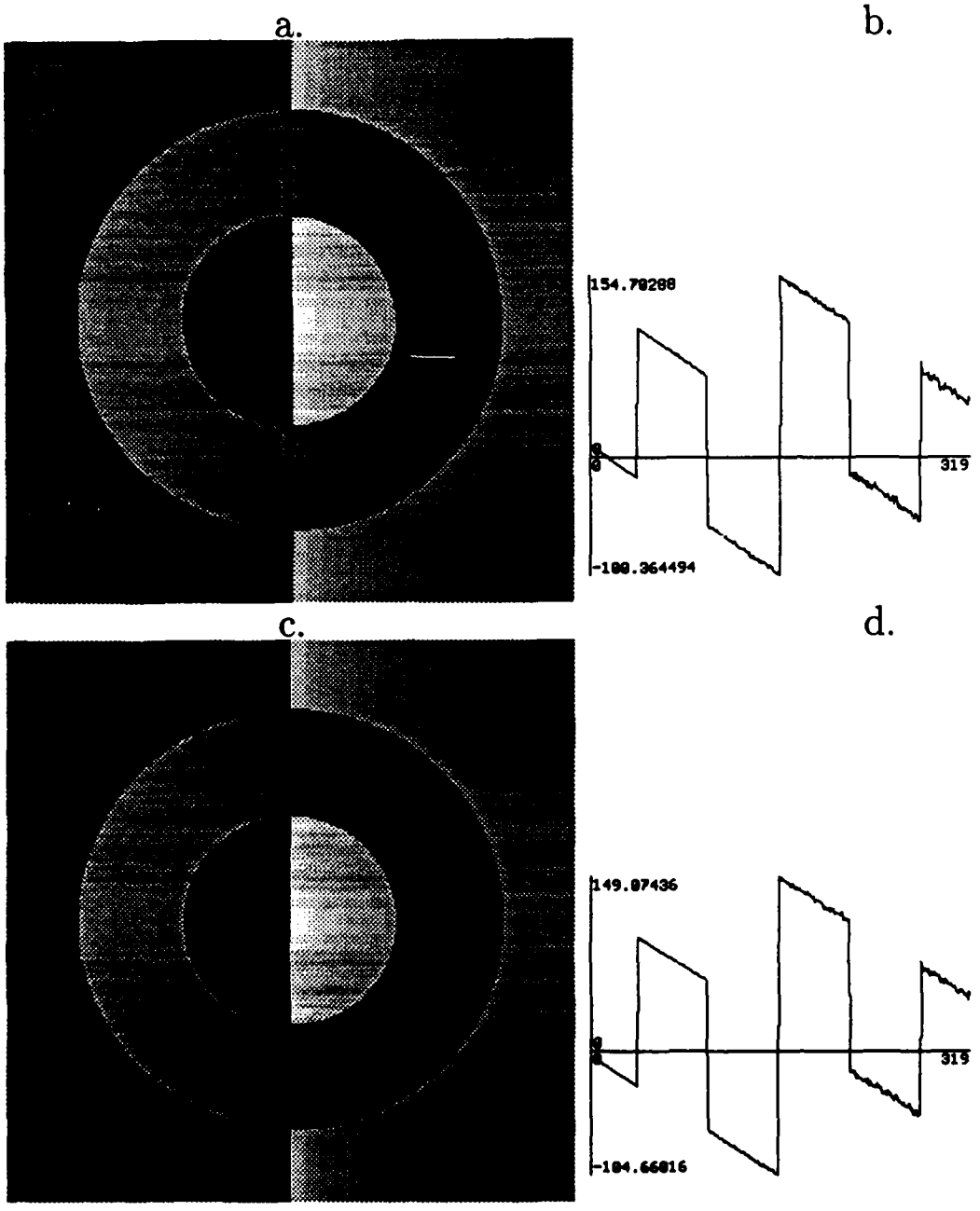

Figure 5.11: The linear operator gives the same results on Koffka Rings with (a) and without (c) midlines.

assumptions listed in Section 3.5: image irradiances in one wavelength channel scale in the same way everywhere across the shadow edge. (The assumptions in Chapter 3 also require that the ratios of image irradiances in distinct wavelength channels do not change across any patch that straddles the shadow edge, but for lightness algorithms this is not relevant.) Because the lightness operator is a one-dimensional operator that acts on each column independently, normalizing each separately, the shadow edge is effectively eliminated. But when the shadow edge runs counter to the columns, the lightness operator interprets it as a reflectance edge. Globally, though, the output of the lightness operator resembles the desired output more than it does the input. Patches that straddle the shadow edge are assigned inaccurate reflectances because the central value seen by the lightness filter (see Figure 5.3) changes greatly across the edge, while the surround changes little. The reflectances of patches far from the shadow edge, on the other hand, are quite accurately represented, because both center and surround values change little across them.

It is important to note, though, that in one sense the lightness operator does reproduce human perception. That is, we *sense* the image irradiance change across a shadow edge but *perceive* it as a shadow – we do not discount it completely. Acting on each of three wavelength channel images independently, the lightness operator would assign the same hue to different pieces of a single patch on either side of a shadow edge, but different luminances. (That is, the ratios of lightnesses in the distinct wavelength channels would remain the same across the shadow, but the sum of lightnesses would not.) We would also see that the patch has a constant hue and a changing brightness, but we would attribute the change in brightness to the effective illumination, not to the surface reflectance. One of the cues we might use in determining the cause of the shadow edge might well be the fact that hue does not change across it. This is the cue that segmentation algorithms use. Lightness algorithms may therefore provide an adequate model for the early stages of human perception.

Not only can lightness operators not distinguish between image irradiance changes due to different causes but also they can not use image irradiance changes to influence

color computation in a nonlinear way. Take, for example, the Koffka Ring [54]: a uniform gray annulus against a bipartite background (Figure 5.10(a)) appears to split into two halves of different lightnesses when the midline between the light and dark halves of the background is drawn across the annulus (Figure 5.10(b)). The estimated linear operator of Figure 5.3(a) acting on the Koffka Ring of Figure 5.10(a) reproduces our perception by assigning a lower output reflectance to the left half of the annulus (which appears darker to us) than to the right half. (Note that the operator achieves this by subtracting a non-existent illumination gradient from the input image irradiance. How it does so is apparent from the shape of the filter in Figure 5.3(a): the surround seen by the operator changes linearly across the Koffka annulus while the central value does not. It is also interesting to note that the operator effectively solves the Poisson equation (Equation 4.2) within a bounded region. Since a linear gradient is invisible to the ∇^2 operator, it may be added to any solution to the Poisson equation to satisfy the boundary conditions. Arend [3] demonstrates that the human visual system sometimes sees such "illusory lightness gradients.") Yet the operator gives this brightness contrast effect whether or not the midline is drawn across the annulus (Figure 5.11). Because the operator can perform only a linear transformation between the input and output images, it is not surprising that the addition of the midline in the input evokes so little change in the output. These results demonstrate that the linear operator alone cannot compute lightness in all worlds and suggest that an additional operator might be necessary to mark and guide it within bounded regions. Chapter 6 addresses the question of how a lightness-type operator might work in concert with luminance-edge finder.

Chapter 6

Segmentation Algorithms

To rely on color as a cue in recognizing objects, a visual system must have at least approximate color constancy. Otherwise it might ascribe different characteristics to the same object under different lights: a ripe yellow banana in the kitchen could become a raw green one in a forest glade. But the first step in using color for recognition, segmenting the scene into regions with distinct surface reflectance properties, does not require color constancy. In this crucial step color serves simply as a means of distinguishing one object from another in a given scene. Color differences, which mark material boundaries, are essential, while absolute color values are not. The goal of segmentation algorithms is to achieve this first step toward object recognition by marking discontinuities in the surface spectral reflectance function, or in other words, by finding color edges that mark material boundaries.

The segmentation algorithms introduced in this chapter, unlike lightness algorithms, start with triplets of input values that are transformations of the image irradiance values in the original sensor channels. Two of the transformed values ("hue" values) are analogous to the CIE chromaticity coordinates x and y in that each represents the irradiance in one channel divided by the sum of the irradiances in the three channels; the third, analogous to the tristimulus luminance value Y, is given by the sum of the three irradiances. Under the single source assumption the hue values change across space only when the surface reflectance changes, except when

specularities are present. Edges in the images defined by the hue values are localized and enhanced with the help of luminance edges. The color labels provided by the transformed input values are then smoothed and spread uniformly across demarcated regions. Because the luminance edges block the spreading of color values, this process mimics the filling-in phenomenon described in Chapte- 2 and reproduces the effects seen in the Koffka Ring discussed in Chapter 5.

6.1 Color Labels

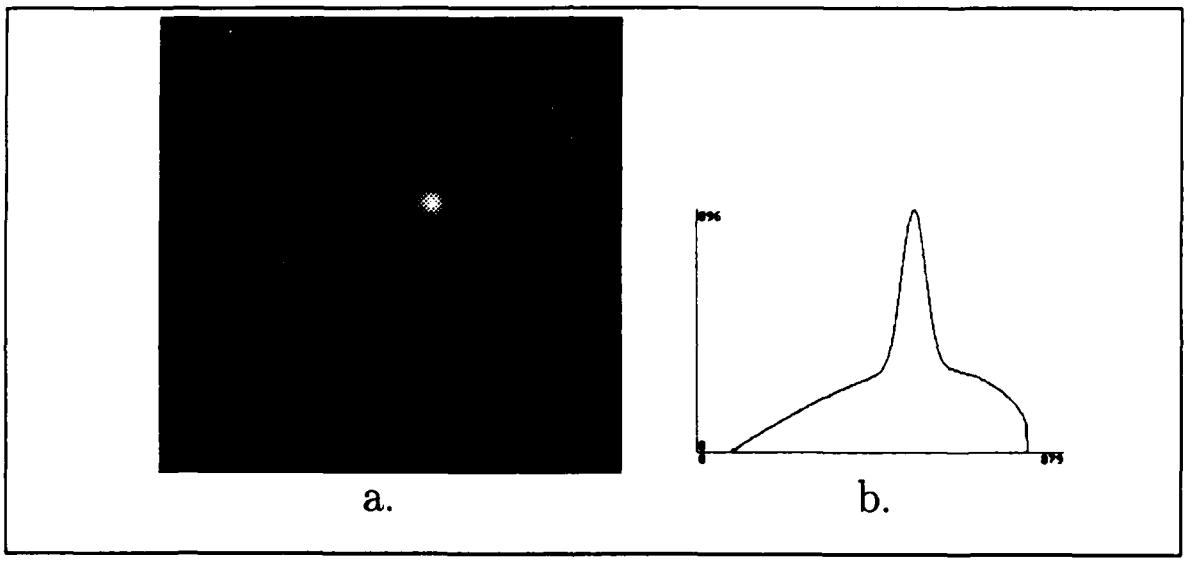

Figure 6.1: (a) A simulated sphere with uniform surface reflectance illuminated by a point source. This is the *luminance* image, in which grey levels correspond to the sum of red, green, and blue image intensity values. (b) Horizontal slice through center of specularity, starting at the arrow marked in (a).

Lightness algorithms identify all sharp changes in the image irradiance (*irradiance edges*) as reflectance edges, and hence as material boundaries. In large part, this is because the image irradiance signal on which they operate is given in three independent broad band spectral channels. If they were provided instead with the signal obtained by taking the difference (or ratio) of the signals in two of the orginal spectral channels, under certain conditions many of the confusing irradiance edges would

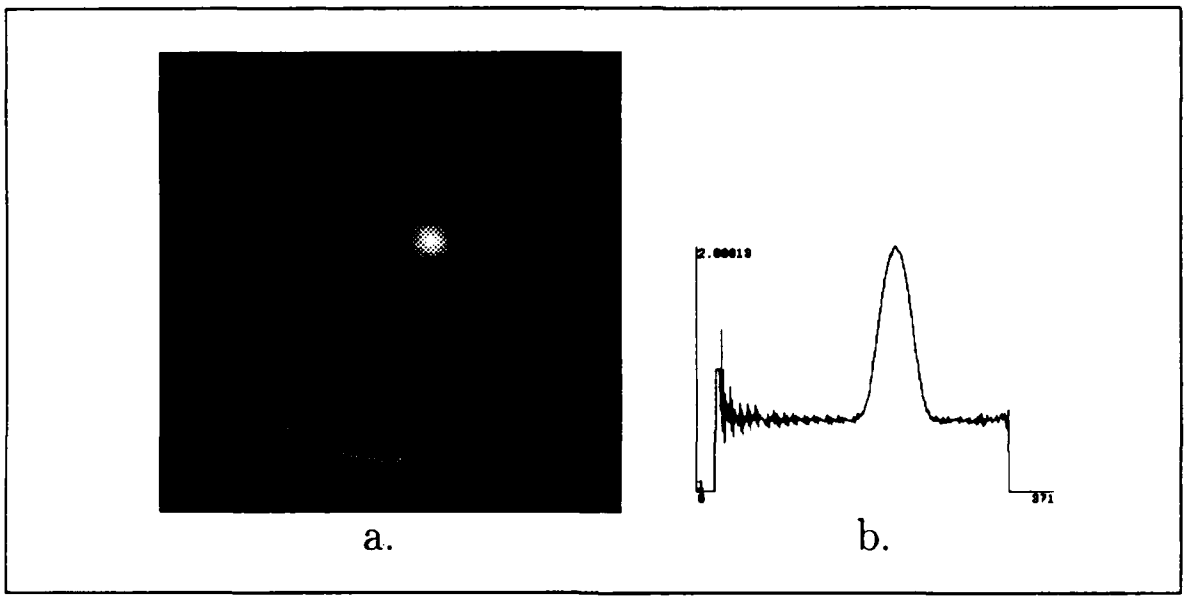

Figure 6.2: (a) Hue image of sphere in Figure 6.1: grey levels now correspond to hue values, which are the ratio of red and green image intensity values. (b) Horizontal slice through center of specularity at arrow. Note that although the slow changes in luminance due to shading disappear, the specularity is still visible.

already be eliminated in the input. Thus the choice of channel in which lightness is computed can significantly alter the color labels computed.

Ideally, one would like to perform a simple transformation on the original spectral channels, in which the responses are (S^i, S^j, S^k), to create new channels, in which the responses are, say, (H^a, H^b, H^c), so that all reflectance edges but no other type of irradiance edge are represented as edges in the H image and distinct H triplets label distinct surface reflectance properties. Then all a color algorithm need do is detect and enhance the edges in H. Rubin and Richards [98] employ the sign of the ratio of the difference of S^i and S^j to the sum $[(S^i - S^j)/(S^i + S^j)]$ to label materials within regions marked by a previous segmentation step (see Section 3.5). The numerator of the ratio is notably like the signal transmitted by physiological color-opponent cells.

Here we choose similar input image values by starting with Equation 3.13 and the single source assumption.

The single source assumption says that all objects in a scene are illuminated

by light with the same chromaticity. This assumption requires that the effective irradiance in Equation 3.13 separate into the product of two terms:

$$E^i(\mathbf{r}_s) \;=\; k^i E(\mathbf{r}_s) \tag{6.1}$$

which follows from

$$E(\lambda, \mathbf{r}_s) \;=\; E(\mathbf{r}_s)k(\lambda)$$

so that

$$S^i(\mathbf{r}) = k^i E(\mathbf{r}_s)\rho^i(\mathbf{r}_s). \tag{6.2}$$

In other words, the single source assumption states that the effective irradiance varies in the same way across space for each spectral channel. In general, the assumption is valid when there is a single light source; in particular, illumination changes due to shadows, shading and surface orientation changes usually affect all wavelengths in the same way, whereas those due to specularities do not. The assumption is clearly violated if there are two spatially and spectrally distinct light sources.[1]

Under the single source assumption, the spatially dependent term of the effective irradiance $E(\mathbf{r})$ will factor out in any ratio of the signals in two distinct channels:

$$h_{ij} \;=\; S^i/S^j \;=\; \frac{k^i E(\mathbf{r}_1)\rho^i(\mathbf{r})}{k^j E(\mathbf{r})\rho^j(\mathbf{r})} \;=\; \frac{k^i \rho^i(\mathbf{r})}{k^j \rho^j(\mathbf{r})}. \tag{6.3}$$

Thus discontinuities in h_{ij} will mark discontinuities in the surface spectral reflectance function, or material boundaries. (Here the superscript i is used to refer to the sensor responses S^i as the product of E^i and ρ^i, as derived in Equation 3.13, rather than the log of the product, as derived in Equation 3.18. The effective irradiance similarly disappears in the difference $h_{\mu\nu} = S^\mu - S^\nu$, where S^μ is the log of the sensor response. See Section 8.2.2 for a discussion of the physiological relevance of how the hue index is written.)

[1] Ramachandran[96] demonstrates that we appear to rely on a similar single source assumption (without reference to the spatial varation of the source color) in making judgments of 3-D shape from shading information.

Conversely, image regions across which h_{ij} is constant signify uniform surfaces of single materials. Synthetic images generated with standard computer graphics algorithms (using, for example, the Phong reflectance model) behave in this way: illumination changes due to shading, shadows, and surface curvature are factored out in the ratio h_{ij}, while those due to specular reflections are not. Figure 6.1 shows the luminance image ($I = I^r + I^g + I^b$, where I^r is the image intensity in the red channel, I^g the green, and I^b the blue) of a sphere with a uniform body reflectance function and a non-zero specular reflectance function simulating that of an inhomogeneous material. A horizontal slice through the center of the specularity reveals the smooth, slow changes in the irradiance function due to illumination gradients on the three-dimensional surface, to be compared with the sharp, fast changes due to specular reflections.

h_{ij}, on the other hand, remains constant throughout the regions of smooth shading on the sphere, as illustrated in the hue image of the sphere (Figure 6.2). The hue image depicts the ratio I^r/I^g at each pixel. As the horizontal slice in Figure 6.2(b) demonstrates, h_{ij} changes significantly across specularities. Thus, for non-Lambertian surfaces, changes in h_{ij} alone could not be used reliably to indicate material boundaries.

For Lambertian surfaces under single light sources, though, the ratio h_{ij} is a good candidate for an input image value in which edges mark material boundaries. The desired triplet of images defined by (H^a, H^b, H^c) may be formed by taking the three distinct pairwise combinations of the indices i, j and k. The segmentation problem may then be tackled by marking regions of constant h_{ij}, or, equivalently, by finding contours of discontinuities in h_{ij}.

In practice, real h_{ij} data are noisy and unreliable and therefore do not permit segmentation so easily. This is because h_{ij} is the quotient of numbers that are not only noisy themselves (see Figure 6.2) but also, at least for biological photosensor spectral sensitivities, very close to one another.

The first step toward constructing a robust segmentation algorithm is thus to

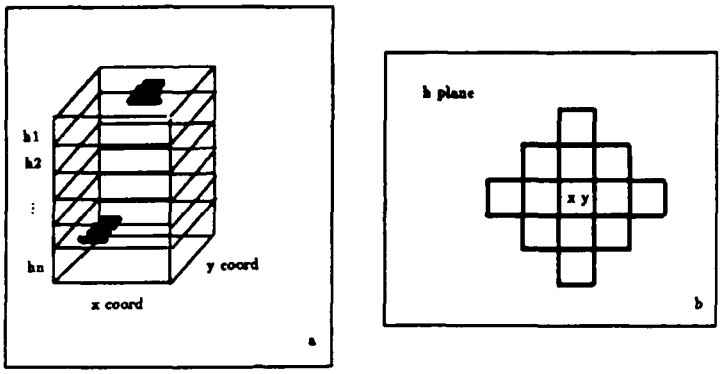

Figure 6.3: (a.) For the cooperative and winner-take-all algorithms, the stack of hue planes into which the hue values are initially loaded. (b.) The excitatory neighborhood of the x,y-th pixel in one hue plane.

choose more stable hue values that still exhibit the properties of h_{ij} defined above. By analogy with the CIE chromaticity coordinates x and y and the tristimulus value Y we choose the values u, v and w such that

$$u = \frac{I^r}{I^r + I^g + I^b}$$

$$v = \frac{I^g}{I^r + I^g + I^b}$$

$$w = I^r + I^g + I^b$$

where u and v are the new hue values and w is the luminance value. Whereas the ratio I^r/I^g hovers near 1 for most images, u and v range more widely between 0 and 1. The new values are still noisy though and do not provide clean segmentation without further processing.

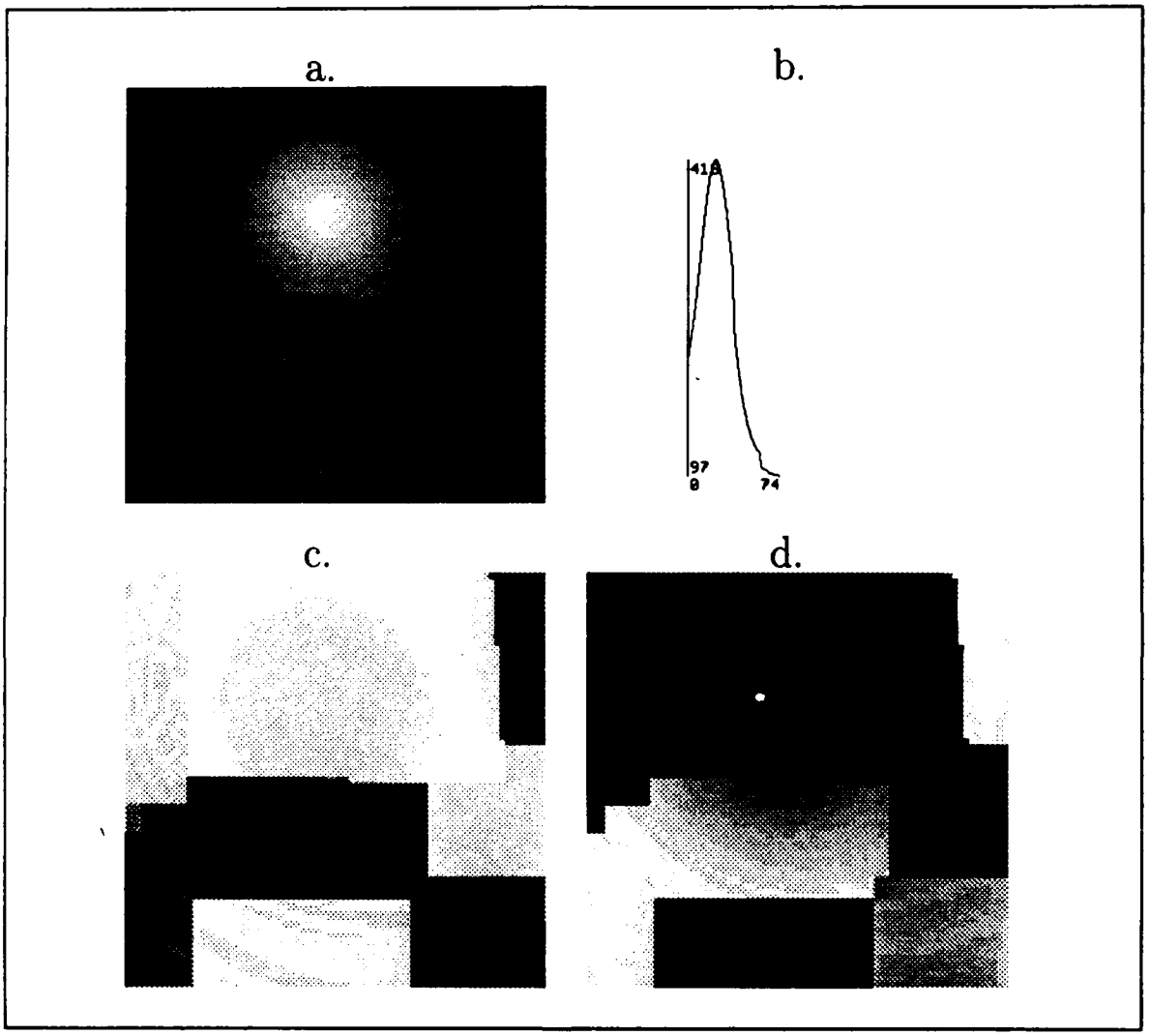

Figure 6.4: (a) 75x75 pixel region of the red channel image of a Mondrian-textured sphere with a specular surface illuminated by a point light source. (b) Vertical slice through center of specularity. (c)"u" $[I^r/(I^r+I^g+I^b)]$ image of (a). (d) "v" $[I^g/(I^r+I^g+I^b)]$ image of (a).

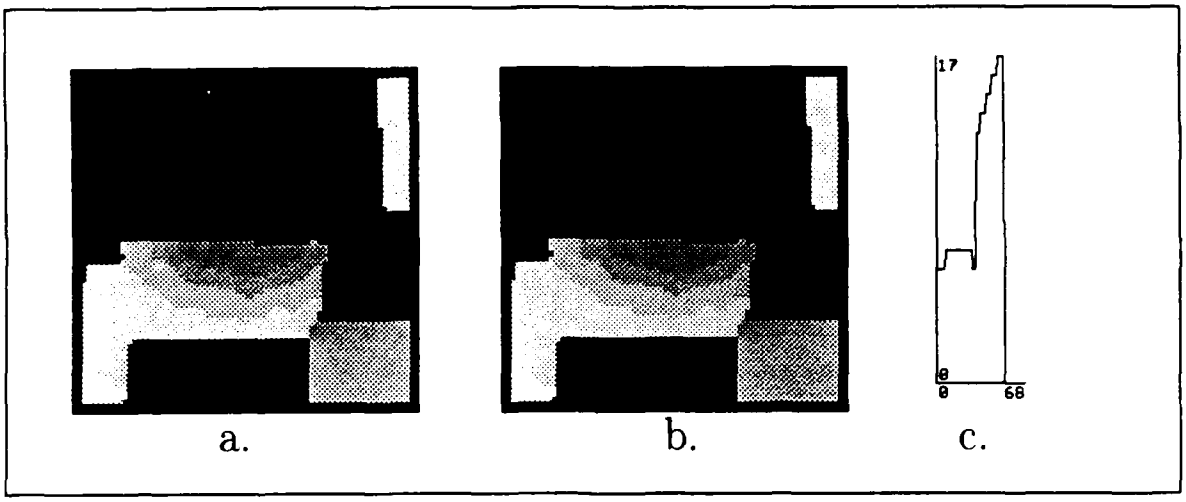

Figure 6.5: Images obtained after (a) five and (b) ten iterations of the winner-take-all scheme on the "v" image, using the Canny edges of the intensity image (see Figure 6.8(b)) as edge input. Note that even after ten iterations, the regions within intensity edges still are not smooth, because the initial discretization into bins cannot be overcome. (c) Vertical slice through (a), showing clearly the discrete levels.

6.2 The Role of Luminance Edges

The next step is therefore to find a method that enhances discontinuities in the hue values, smoothes the noise and fills in data where they are unreliable. Reasoning from the implications of many psychophysical phenomena, luminance edges are a natural choice for help in enhancing hue discontinuities. For example, isoluminant boundaries are fuzzy; high contrast luminance edges can capture and contain blobs of color even if they do not fall perfectly within the edges, as in impressionistic paintings or color cartoons; luminance edges block the filling-in of color in stabilized image experiments; and, as in the Koffka Ring, luminance edges can create apparent hue discontinuities.

Except in artificial isoluminant images, hue discontinuities rarely occur in the absence of discontinuities in luminance. Therefore the hue edges in an image should form a subset of the set of luminance edges, depending on the parameters used in detecting edges. In the algorithms below, we use luminance edges to activate colocalized

hue edges and to block the spreading of uniform hue values.[2]

6.3 The MRF Approach

One approach to the segmentation problem is to regularize the data using Markov Random Field (MRF) techniques. Similar techniques have been exploited by several authors (Poggio et. al., [94] Blake and Zisserman [8], Geman and Geman, [32] Marroquin [77] and Geiger and Girosi [31]) to solve other surface reconstruction problems in early vision in which the data supplied are sparse and noisy, and where the surface may be defined, for example, by depth, color or motion. MRFs specify the probability distribution of solutions to the surface reconstruction problem by providing an energy associated with each possible solution. The energy function is dependent only on local interactions within the reconstructed surface. That is, the probability that a given point on the surface has a given value – say, hue – is determined solely by the hues assigned to its neighbors on the surface.

Formally, the prior probability of a solution field f is given by

$$P(f) = (1/Z)e^{-E(f)/T}$$

where Z is a normalizing constant, $E(f)$ is the energy of the field f and T is the "temperature," a constant. The energy function $E(f)$ is the sum of contributions from each neighborhood in the surface, for which the local energy function has the same form. It typically includes a term that penalizes reconstructed values far from the original data and those far from neighboring values, when a continuous surface is desired. For surfaces in which discontinuities are allowed, the energy function may include a term representing the presence or absence of discontinuities which break the continuity constraint at specific locations. (Geman and Geman [32] first introduced

[2]Note that Land's original retinex algorithm [61], which thresholds and sums the differences in image irradiance between adjacent points along many paths, accounts for the contribution of edges to color, without introducing a separate luminance edge detector.

this term as a line process defined on the lattice dual to f; see Geiger and Girosi [31] for an alternative construction.)

Poggio et. al. [94] derive an energy function $E(f)$ in which the reconstructed hue values u and v are required to be piecewise smooth and close to the input hue values. Hue edges that colocalize with luminance edges are enhanced by a term that couples the probability of the two. This formulation leads to a stochastic algorithm, a form of which has been implemented on real images [94].

6.4 Cooperative Color Algorithms

The stochastic algorithm based on the MRF formulation selects the most likely surface from a random series of candidates, a procedure that is computationally expensive and biologically unlikely. A deterministic algorithm would in many cases be more efficient and biologically plausible. A deterministic approximation to the stochastic algorithm based on the MRF formulation is explored in the following section. In this section we examine an alternative to the MRF approach by drawing an analogy with the cooperative stereo network of Marr and Poggio [75].

We start with the observation that if we require hue to be piecewise constant, or less restrictively, piecewise smooth, and assume that the visual system we are simulating can distinguish only a finite set of discrete hues, then the problem of assigning hues to surface points becomes analogous to that of assigning depths. (As other authors have noted, see e.g. Blake and Zisserman [8], similar analogies can be drawn between many problems in early vision.) As in the problem of computing depth from binocular disparity, two constraints apply: uniqueness – each surface point may be assigned only one hue; and piecewise constancy or continuity – hue values are constant or at most smoothly varying within edge-bounded regions. As if in a Mondrian world, here we ask for piecewise constancy, but this constraint proves too restrictive, as shown below.

To implement these constraints, we construct a cooperative network, a type of

Hopfield net, that performs similarly to the cooperative stereo network of Marr and Poggio [75, 76].

We construct a three-dimensional binary array $C(x, y, u)$ corresponding to the array $u(x, y)$, where x and y are the image spatial coordinates and U labels discrete hue values in a finite set. The binary array $C(x, y, u)$ is loaded by setting all cells at (x, y) to 1 for which $u(x, y)$ is within a specified distance from U; otherwise, $C(x, y, u)$ is set to zero. At each instant in time, each cell $C(x, y, u)$ in the array receives "excitatory" contributions from cells in the same U layer belonging to its excitatory neighborhood (see Figure 6.3). It receives "inhibitory" votes from cells at the same (x, y) position but with different U. If the sum of these contributions exceeds a certain threshold the cell will be active at time $t = t_0 + 1$; otherwise the cell will be set to zero.

Thus,

$$C^{t+1}_{x,y,u} = \sigma(\Sigma_{x',y',u' \in E(x,y,u)} C^t_{x',y',u'} - \epsilon \Sigma_{x',y',u' \in I(x,y,u)} C^t_{x',y',u'} + \alpha C^0_{x,y,u}) \qquad (6.4)$$

where σ is a step function that implements the threshold, and $E(x, y, u)$ and $I(x, y, u)$ are the excitatory and inhibitory neighborhoods of (x, y, u), respectively. The network enforces local constancy and uniqueness by requiring that only one cell be active at each (x, y) across all U values.

To incorporate information from brightness edges, we construct a binary map with "black" and "white" regions corresponding to oppositely signed regions of the function $\nabla^2 G * w(x, y)$, where G is a Gaussian kernel and $w(x, y)$ is the image luminance. In Equation(6.1), the summation over E is then restricted to black cells if (x, y) is black, and to white cells if (x, y) is white.

Although the cooperative color network produces reasonably good results on synthetic images, there are a number of arguments against using it either as a model for biological color vision or as a machine vision algorithm. First, the number of iterations required to achieve a stable configuration of hues and hue boundaries is high.

Biological visual systems cannot rely on such a time-consuming process to segment scenes into regions of constant colors. Second, the initial partitioning of hue values into discrete bins is neither straightforward computationally nor plausible biologically. Noisy values from a single region that should be assigned a constant hue will be loaded initially into a different group of bins at each (x, y). After many iterations on the cooperative network, the "uniform" region may split up into islands of similar but distinct hue values (see Figure 6.5).

The first problem may be solved by implementing a one-pass winner-take-all scheme on the network. For each pixel for each hue plane, a local processor sums the binary votes from all pixels in its excitatory neighborhood. A second processor scans the hue planes and assigns to each (x, y) the hue of the hue plane with the most votes at that pixel. This algorithm is much quicker than the first but still suffers from side-effects of the initial discretization of hue values (see Figures 6.4 and 6.5). Other methods of loading the bins obviate this problem, but require pre-processing of the image that is both inefficient and biologically unlikely.

6.5 A Segmentation Algorithm

To avoid the problems inherent in the above approaches, we return to the MRF formulation, which provides a rigorous foundation for a simple algorithm. To avoid small step changes in hue across a uniform surface we relax the requirement for piecewise constancy and instead require only that hue be piecewise smooth. This obviates the need for specifying a finite range of allowed hue values. Although in principle the hue values may now vary smoothly within an infinite range across a scene, in practice they will not vary much, depending on how the smoothness criterion is implemented.

To implement the smoothness criterion we incorporate it into an energy function that the reconstructed surface should minimize. The full energy function for the point (x, y) is:

$$E^{xy} = (\overline{u}_{xy} - u^o_{xy})^2 + \alpha E^{N_{xy}} + \gamma \sum_{xy \in N_{xy}} E(l_{xy})$$

where u^o_{xy} is the initial hue data, $\overline{u}_{xy}$ is the reconstructed hue, N_{xy} is the local neighborhood of pixel (x, y), $E^{N_{xy}}$ is the term representing the smoothness criterion and $E(l_{xy})$ is the energy of the line process that marks hue discontinuities (including, if desired, interaction with luminance discontinuities).

For the deterministic approximation, we incorporate the term due to the line process into the term $E^{N_{xy}}$ and exclude the first term. Then $E^{N_{xy}}$ becomes equal to the following function:

$$(1 - d_{x+1,y})V(u_{x,y}, u_{x+1,y}) + (1 - d_{x,y+1})V(u_{x,y}, u_{x,y+1})$$

$$+(1 - d_{x-1,y})V(u_{x,y}, u_{x-1,y}) + (1 - d_{x,y-1})V(u_{x,y}, u_{x,y-1}),$$

where $V(u_1, u_2) = V(u_1 - u_2)$ is a quadratic potential around 0 and constant for $|u_1 - u_2|$ above a certain value (see Figure 6.6). (This potential is equivalent to the "weak membrane" energy discussed by Blake and Zisserman [8].)

The values d_{xy} are explicitly provided at each pixel (x, y) by edge detection on the hue and luminance images and are either 0 or 1 depending on the absence or presence of an edge at pixel (x, y) as described below. Thus, if a valid edge exists at a neighboring pixel (i.e., $d = 1$), the neighborhood interaction term goes to zero and the reconstructed value is exempted from the smoothness criterion. We derive an iterative algorithm by using gradient descent to solve for the stationary value of u. Successive changes in u are governed by the equation

$$du/dt = -\alpha \frac{\partial E^{N_{xy}}}{\partial u}.$$

For $\alpha = 1/10$, solving for the value of u at discrete times t yields the following expression:

$$u_{x,y}^{t+1} = \frac{1}{n(N^t)} \sum_{l,m \in N(u_{x,y}^t)} u_{l,m}^t = < u >_N$$

where $N(u_{x,y}^t)$ is the set of $n(N^t)$ pixels among the next neighbors of x, y that differ from $u_{x,y}^t$ by less than a specified amount and are not crossed by an edge in the edge map(s) (on the assumption that the pixel (x, y) itself does not belong to an edge). $n(N^t) = 5$ if none of the nearest neighbors are crossed by an edge.

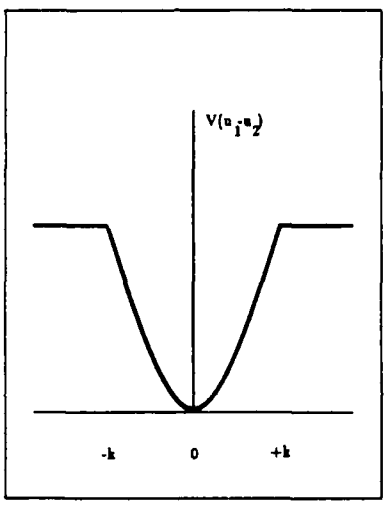

Figure 6.6: The energy potential that implements the piecewise smoothness constraint.

The result is an algorithm that simply replaces the value of each pixel in the hue image with the average of its local surround, iterating many times over the whole image. The algorithm takes as input the hue image (either the u-image or the v-image) and one or two edge images, either luminance edges alone, or luminance edges plus u or v edges, or u edges plus v edges. The edge images are obtained by performing Canny edge detection or by using a thresholded directional first derivative. On each iteration, the value at each pixel in the hue image is replaced by the average of its value and those in its contributing neighborhood (see Figure 6.7b). A neighboring pixel is allowed to contribute if (i) it is one of the four pixels sharing a full border

with the central pixel (see Figure 6.7a) (ii) it shares the same edge label with the central pixel in all input edge images (see Figure 6.7c) (iii) its value is non-zero and (iv) its value is within a set range of the central pixel value. The last requirement simply reinforces the edge label requirement when a hue image serves as an input edge image – the edge label requirement allows only those pixels that lie on the same side of an edge to be averaged, while the other insures that only those pixels with similar hues are averaged.

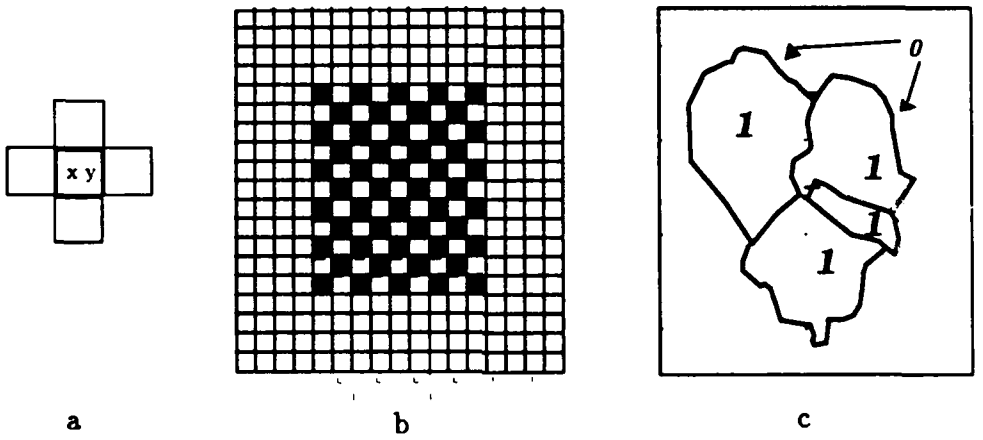

Figure 6.7: (a) For the averaging algorithm, the contributing neighborhood of the x,y-th pixel. (b) In the first iteration, each black pixel is set to the average of the values at it and the four neighboring white pixels (provided they share the same edge label – see (c) – and fall within the range for similar pixel values). In the second iteration, each white pixel is set to the average of the value at it and the four neighboring (newly averaged) black pixels. Successive iterations repeat this pattern. (c) An edge input image; all pixels lying within a closed edge are labeled "1", while all pixels lying on an edge are labeled "0." Only pixels with the same edge label are allowed to contribute to the same neighborhood.

Because the original data are used only as initial values for the iteration scheme, we expect that the algorithm will provide asymptotically piecewise constant regions bounded by discontinuities. If the term requiring closeness to the original data is restored to the energy function, then the iteration rule becomes:

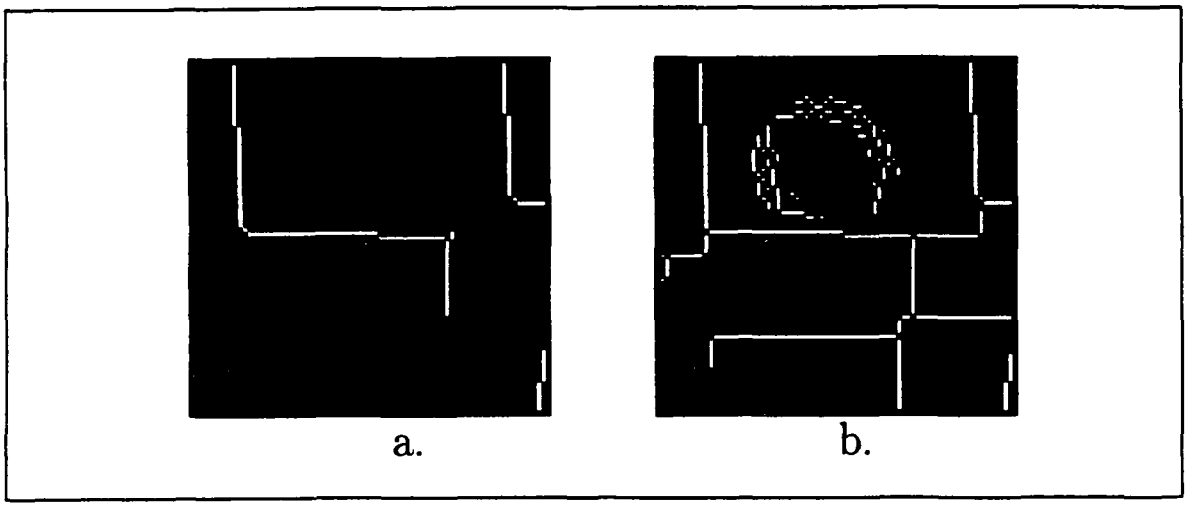

Figure 6.8: (a) Canny edges of the small intensity image using the default sigma. (b) Canny edges of the small intensity image using sigma = 0.5. Note that at this resolution the edge detector easily finds the borders of the specularity.

$$u_{x,y}^{t+1} = < u >_N - k(u_{x,y}^t - u_{x,y}^0)$$

where u^0 are the input hue data. The resulting "surface" is a generalized spline.

The local averaging smoothes noise in the hue values and spreads uniform hues across regions marked by the edge inputs. On images with shading but without strong specularities, the algorithm performs a clean segmentation into regions of different hues. Figure 6.9 illustrates the result of 1000 iterations of the algorithm on a synthetic image of a Mondrian-textured sphere with a specular surface generated using the Phong reflectance model. Figures 6.10 and 6.11 illustrate the performance of the algorithm on a natural image obtained with a CCD color camera.

This simple segmentation network also mimics our perception of the Koffka Ring. We replicate the "illusion" in the following way. The Koffka Ring is filtered through the lightness filter shown in Figure 5.3(a). (For black-and-white images this step replaces the operation of obtaining u and v: in both cases the goal is to eliminate spatial gradients in the effective illumination.) The filtered Koffka Ring is then fed to the averaging network together with its brightness edges. When in the input image

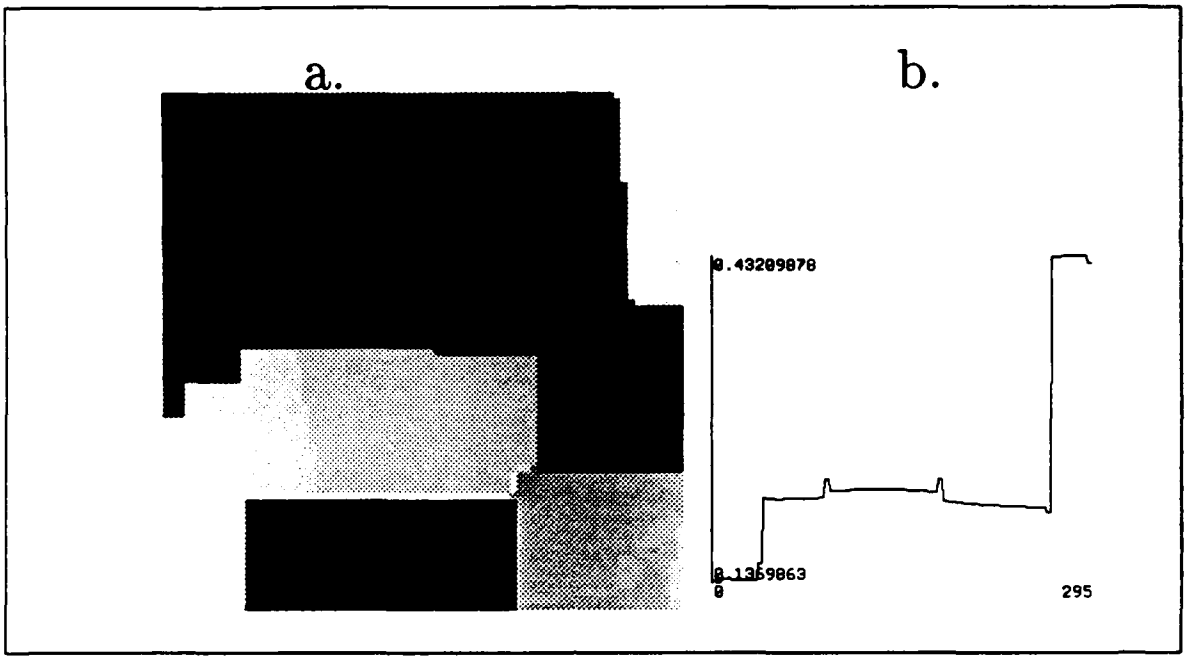

Figure 6.9: (a) The image obtained after 1000 iterations of the averaging network on the image in Figure 6.4(d), using as edge input the Canny edges of the luminance image shown in Figure 6.8(b). The edge input is aided by a threshold on differences in the v image which allows only similar v values to be averaged. (b) Horizontal slice through center of specularity in (a). Note that the specularity appears as a smooth patch with a slightly different v-value in the center of another patch.

the boundary between the two parts of the background continues across the annulus, in the output image (after 2000 iterations of the averaging network) the annulus splits into two semi-annuli of different lightnesses in the output image, dark grey against the white half, light grey against the black half (see Figure 6.12). When the boundary does not continue across the annulus, the annulus color relaxes to a uniform grey, but only after approximately 20 times as many iterations.

The averaging scheme finds constant hue regions under the single source assumption in a Lambertian world. Often the real world complies with these constraints, as in the image in Figure 6.10. Although a strong highlight may originate an edge that could then "break" the averaging operation, the specular reflections in the teddy-bear image are weak enough that they average out and disappear from the smoothed hue

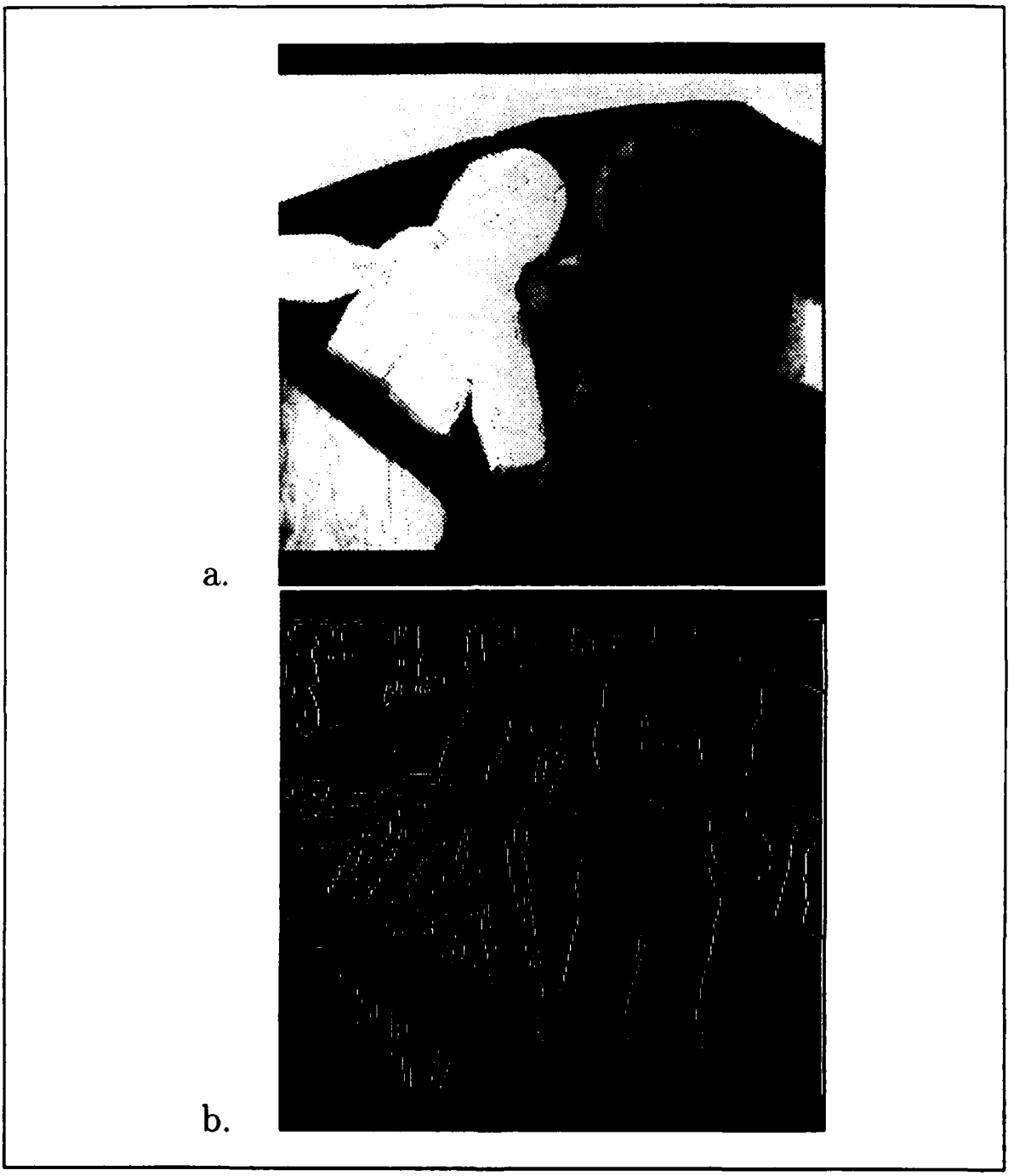

Figure 6.10: (a) Digitized red-channel image of natural scene (teddy bear with jacket) obtained with CCD camera. (b) Canny edges obtained using the default sigma of intensity image of same scene.

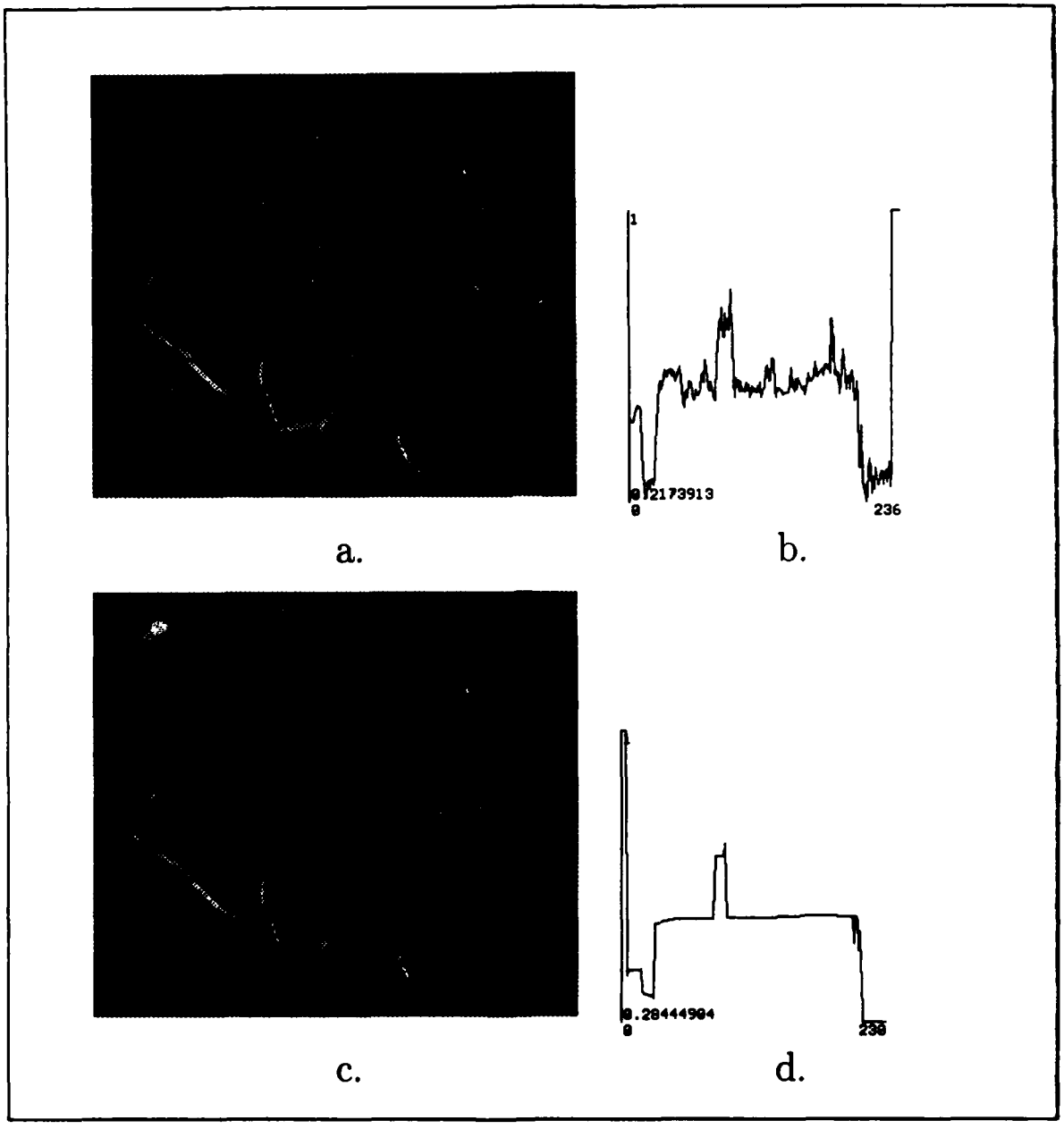

Figure 6.11: (a) Eight-bit "U" image of image in Figure 6.10. (b) Vertical slice through (a) at arrow. (c) Image obtained after 500 iterations of averaging scheme on (a), using intensity image and (a) as edge inputs. The dips at the two extremes correspond to the blanket; the small peak in the center corresponds to the bow tie. (d) Vertical slice through (c) at arrow. Note that the noise is smoothed by averaging while discontinuities in hue are preserved.

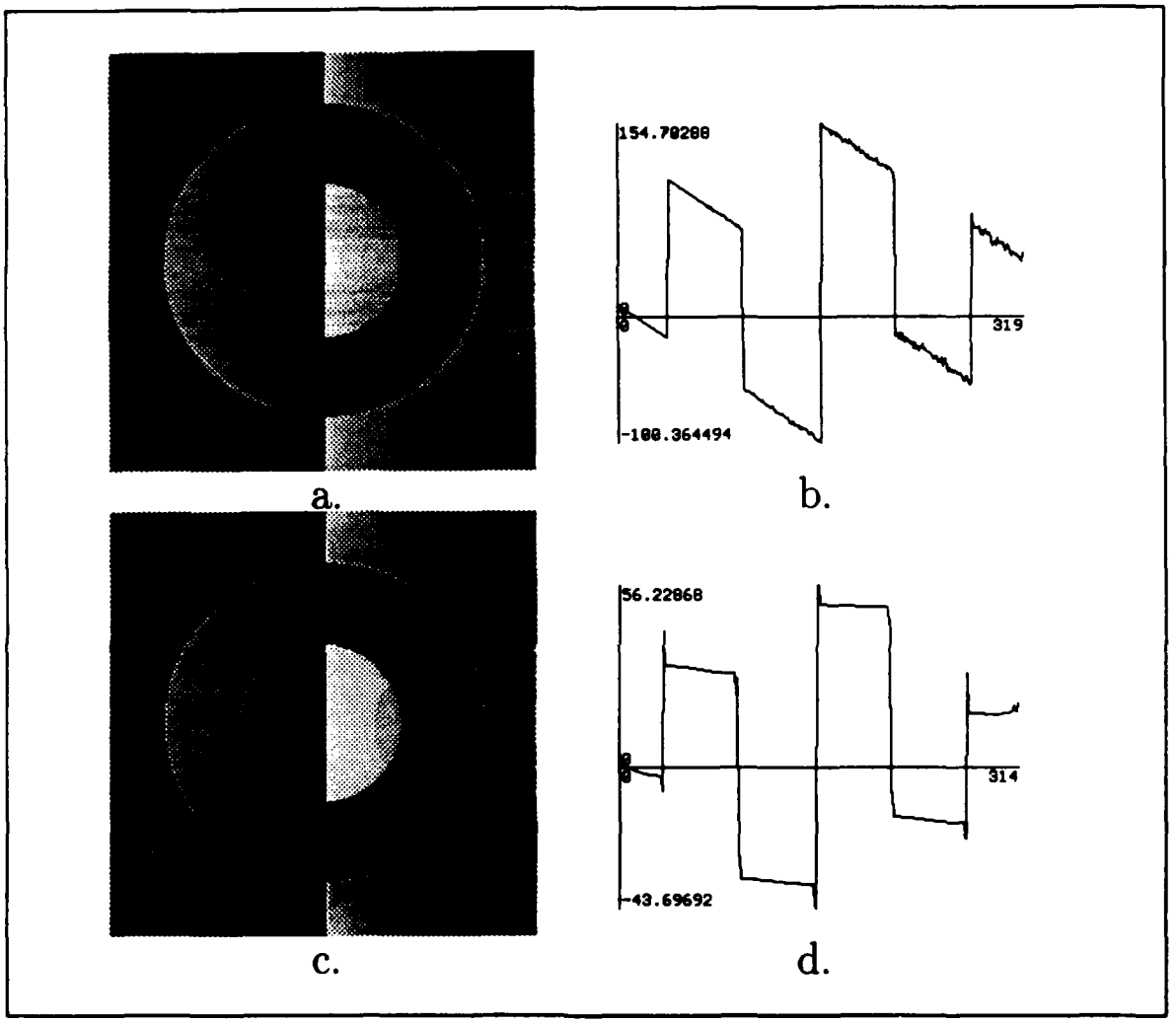

Figure 6.12: (a) Output of linear estimator obtained as described in Chapter 5 acting on Koffka Ring with midline. (b) Horizontal slice through center of (a). (c) Output of averaging algorithm after 2000 iterations, acting on (a). (d) Horizontal slice through center of (b).

image (see Figure 6.11). The iterative averaging scheme completely eliminates the remaining gradients in hue.

Yet a robust segmentation algorithm for the real world will clearly require some method for identifying irradiance edges induced by specular reflections. This method may well entail specialized routines and higher-level knowledge. For example, the interpretation of virtually identical local image intensity changes may vary according to the global context. In one image a bright white line against a dark background may be a painted stripe on a piece of cloth; in another, it might be an extended specularity on plastic. Furthermore, positive identification of the physical cause of an edge may require interaction between several visual modules. Specular reflections, for example, vary widely in magnitude with the angle of view and thus image motion due to eye movements may provide valuable cues for their detection. Despite its limitations, though, this simple segmentation algorithm still captures some of the integral features of human vision.

Chapter 7

Cues to the Color of the Illuminant

The methods for recovering constant colors discussed in chapter 3 also recover, directly or indirectly, the color of the illuminant. (Conversely, a method that correctly recovers the illuminant color can achieve perfect color constancy, at least for nonmetameric objects.) Like the constant object colors they recover, the illuminant color is an estimate dependent on the assumptions that each method employs to make equation (3.18) (the *image irradiance equation*) tractable. Hence one way to test a method's accuracy and flexibility is to examine the illuminant colors it predicts under different conditions. Similarly, one way to test whether a method closely mimics the one that humans use is to compare the illuminant colors it estimates with those that humans perceive under the same conditions. What one really tests by comparing the outputs of two methods – an explicit machine vision algorithm versus a hidden biological mechanism – is whether they rely on the same assumptions about the physical world, not whether they use identical operations in stepping from light signal to solution. If we rely on the grey-world assumption, for example, then our perception of the illuminant color will vary with the bias in reflectance distributions of scenes, as does the estimate obtained by lightness algorithms that set the average to grey. On the other hand, if we are able to use information from specular reflections as the chromaticity convergence algorithm does, then when sufficient specularities are

present we too will accurately estimate the illuminant color even when the surface reflectance distribution is skewed.[1]

These predictions make two critical assumptions: (1) humans *are* able to judge the illuminant color and (2) we use the same computation both to estimate the illuminant color and to derive constant object colors. Can humans perceive the color of the light source without "looking" directly at it? If we do not, and yet we do perceive roughly constant colors of objects, then intermediate stages of the color computation must be lost to consciousness. Yet if we are capable of judging the illuminant color, still it does not follow that we do so by accessing the same computation for computing constant colors.

In fact, experience and experiments tell us that we are capable of judging illuminant color. At sunset, we sense a reddish glow to things and call it the color of the light; under indoor incandescent lights, we see the source as yellow while simultaneously recognizing apples as red and eggplants as purple. Yet in many circumstances, we would not be correct in assuming the illuminant color to be the same as that of the primary light source we see. In a blue-walled room, with a blue rug underfoot, and blue furniture about, we correctly sense the illuminant color as bluish, despite the greenness of the sunlight - reflected from leaves and grass - pouring in through the window. Although the primary source is the sun, interreflections between the walls, the furniture and other objects insure that the incident light striking any object in the room is far more biased toward blue. As Evans said, we live under strange lights largely determined by our interior decorations: "When the light reaching a particular object is to be evaluated, the light reaching it from all the reflecting surfaces in its neighborhood has to be taken into account (p.54)" ... "...If sunlight enters the window and falls on a bright red carpet, all other objects in the room are lighted by *red* light of the color of the carpet (p.81)"[30].

Experiments performed by Arend and Reeves [4] indicate that illuminant differ-

[1]The work described in this chapter was done in collaboration with Hsien-Che Lee and Heinrich H. Bülthoff.

ences between otherwise identical scenes (in their experiments, computer-simulated two-dimensional Mondrians) are readily visible even without reference surfaces that provide conclusive cues. Furthermore, their results indicate that observers can use cues to the illuminant color from reference surfaces to improve color constancy, suggesting that the two computations, if not the same, are strongly linked.

These comments suggest that an experiment in which human observers judge the illuminant color on different scenes might reveal something about the constraints on color constancy. This chapter reports on such an experiment. To insure readily interpretable results, we presented observers with simple scenes that were nonetheless complex enough to test whether their perceptions agreed with predictions from the computational methods described above. Each scene (presented on a computer-driven color CRT monitor) consisted of a single smooth-surfaced Mondrian-textured sphere illuminated by a single point source. This choice of scene eliminated confusion caused by interreflections, while still providing a range of surface colors. The sphere surface was either specular or non-specular, simulated by different reflectance functions. For some specular spheres, the Mondrian reflectance distribution was biased away from grey, while the source color was differently biased, allowing the grey world assumption and information from specular reflections to come into direct conflict.

7.1 Methods

Seven observers (5 males, 2 females; age range 11 -34) participated in the experiments reported here. Observers were screened for color vision deficits using the Ishihara protanopic and deuteranopic confusion plates and the Farnsworth F2 Tritan plate. (One observer (ENJ) demonstrated a mild protanopia but was nonetheless included in the experiment; all others tested normal – see results.)

The observer viewed a series of computer-generated images displayed on a calibrated color CRT monitor in an otherwise completely dark room. He or she was positioned in a chin-rest at a fixed viewing distance (104 cm) from the monitor. The

observer's line of sight was fixed at a perpendicular angle to the screen. In each image the Mondrian-textured sphere (see Figure 7.1) subtended approximately 4.5 degrees in the center of the observer's visual field. The simulated light source appeared to be suspended between the screen and observer above and to the right of the sphere. Beneath the sphere, separated by .75 degree of black screen, was a rectangular palette of 256 colors, representing a section of the CIE chromaticity diagram that includes the CIE daylight curve (see Figure 7.2).

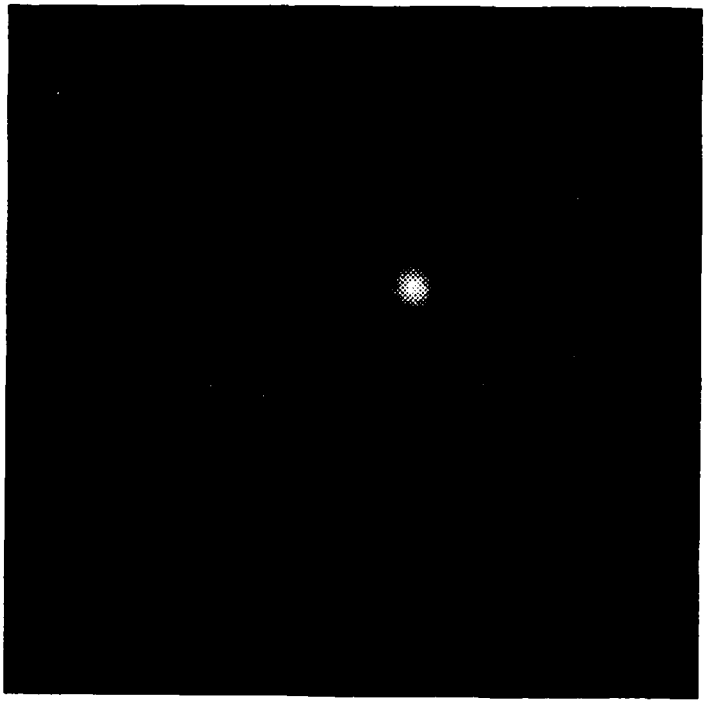

Figure 7.1: Red channel image of a Mondrian-textured sphere, with specularity.

Observers were asked to choose the color of the perceived illuminant on the sphere as the color that white paper in the void would be under the same illuminant. The white paper was simulated by a small patch at the edge of the display, approximately 2 degrees distant from the closest border of sphere. The observer selected its color by clicking on choices in the palette below the sphere. In each trial, the sphere

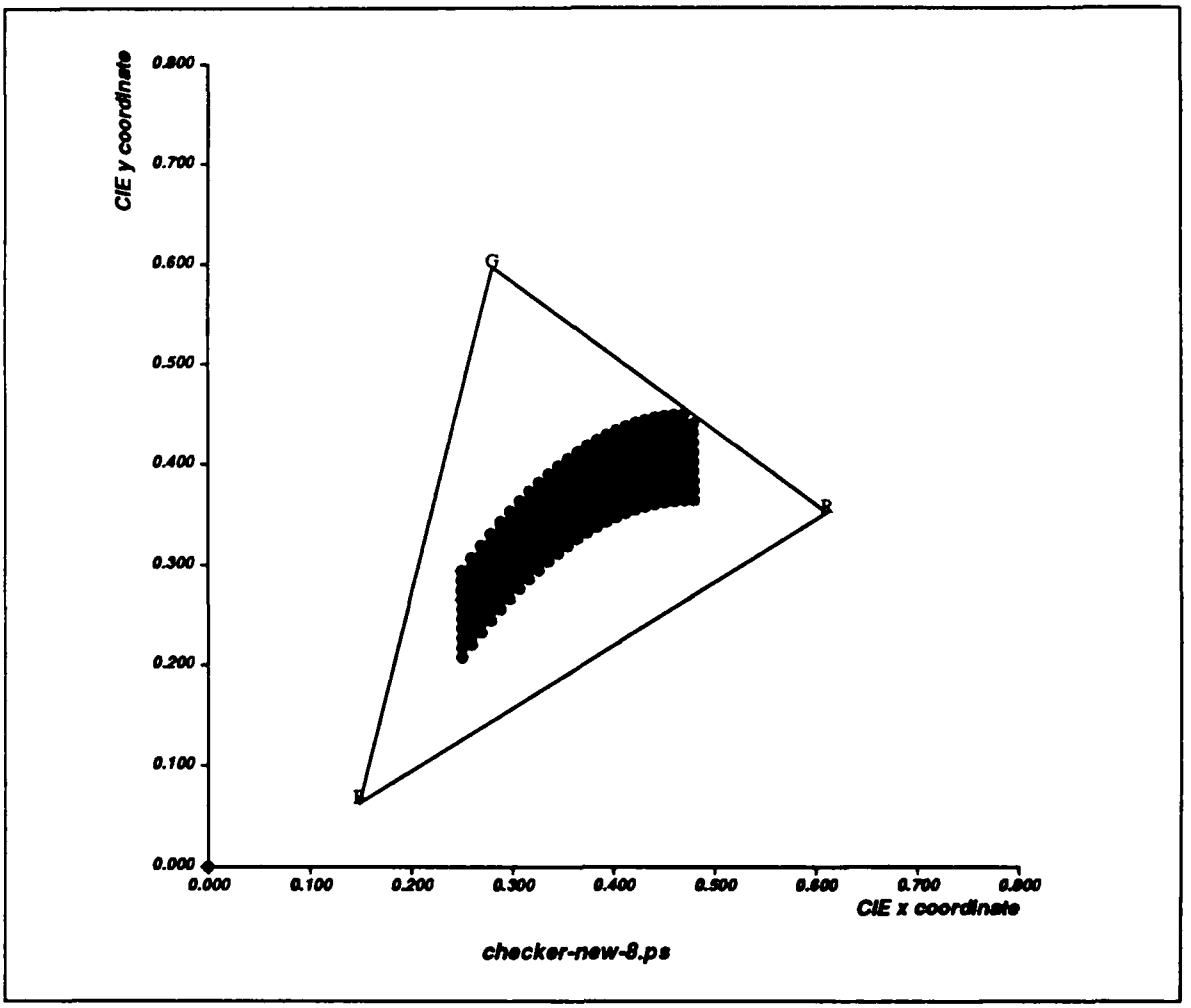

Figure 7.2: Palette colors on CIE chromaticity diagram. The CIE daylight curve runs longitudinally through the center of the band. R, G and B mark the loci of red, green and blue monitor phosphors.

was randomly selected from a large set generated as described below; its illuminant color, chosen as described below, was one of five colors on the palette.[2] Observers were instructed to move their gaze across the entire screen when each new image appeared, to maintain a constant state of adaptation. Successive trials were separated in time by approximately 15 seconds to reduce the effects of successive color contrast. Although there was no time limit on the choice, observers were asked to give as instantaneous a reaction as possible. They received no feedback during the experiment. Afterwards, observers were asked to rate their confidence in their choices and to speculate on the methods they employed to judge illuminant color.

The spheres were generated using the NIR model (see section 3.2) to combine interface and body reflection components and the Phong model [90] for the geometry of the interface reflection. 2-D Mondrian reflectance images were projected onto 3-D spheres of fixed radius. The spheres were then illuminated by a point light source at a fixed position and distance and by ambient light of uniform color and intensity. The three gun values for pixel (x, y) of the final image, $I_r(x, y)$, $I_g(x, y)$ and $I_b(x, y)$[3], were given by the following equations:

$$I_r(x,y) = k_d E_r[\rho_r(f(\mathbf{n},\mathbf{s},\mathbf{v}) + k_a) + \rho_s g(\mathbf{h},\mathbf{v},\mathbf{n},\mathbf{s})]$$

$$I_g(x,y) = k_d E_g[\rho_g(f(\mathbf{n},\mathbf{s},\mathbf{v}) + k_a) + \rho_s g(\mathbf{h},\mathbf{v},\mathbf{n},\mathbf{s})]$$

$$I_b(x,y) = k_d E_b[\rho_b(f(\mathbf{n},\mathbf{s},\mathbf{v}) + k_a) + \rho_s g(\mathbf{h},\mathbf{v},\mathbf{n},\mathbf{s})].$$

Pixel (x, y) is the projection of the point p_n on the sphere characterized by surface normal $\mathbf{n}$. $\mathbf{s}$ is the vector from the surface point p_n to the source and $\mathbf{v}$ is the vector from p_n to the viewer. $\mathbf{h}$ is the vector in the direction of the perfect specular reflection at p_n. In the first term in the square brackets, $\rho_r f(\mathbf{n},\mathbf{s},\mathbf{v})$ represents the

[2]To minimize the difficulty of the task, the palette colors were chosen from an isoluminant plane of CIE chromaticity space, allowing the observer only two degrees of freedom of choice in illuminant color: hue and saturation. Observers were instructed to ignore, as much as possible, any variations in brightness amongst the images.

[3]Note that in this computer-simulated scene, image irradiance and scene radiance are the same, both specified by the gun values at each pixel.

body reflection component, where ρ_i is the body reflectance factor of the sphere surface in the ith channel. (The body reflectance factors ρ_i are constant across each patch of the Mondrian image.) $f(\mathbf{n},\mathbf{s},\mathbf{v})$ is proportional to $(\mathbf{n}\cdot\mathbf{s})/(s^2\cdot v^2)$. The specular reflection component is determined by the second term, in which ρ_s is the specular reflectance factor and $g(\mathbf{h},\mathbf{v},\mathbf{n},\mathbf{s})$ is proportional to $(\mathbf{h}\cdot\mathbf{v})^m$, with $m = 40$. (This is the Phong approximation to the true form of the specular reflection term.) The form of g is chosen to simulate an inhomogeneous material with a largely smooth surface – e.g. a painted surface or plastic – in which the specular reflection tends to take on the illuminant color. A fixed amount of ambient light was added at each pixel, its color determined by the body reflection component $(E_i\cdot\rho_i)$ and its magnitude set by the scaling factor k_a to be a small percentage (< 10) of the minimum irradiance in the final image.

Nine 400x400 pixel Mondrian images were generated, each with roughly 200 patches. In each subset of three Mondrians, the reflectances were randomly chosen from a different distribution. In the set of blue-biased Mondrians the ρ_b values for each patch were randomly chosen from a uniform distribution between minimum and maximum values of 35 and 95, while the ρ_g and ρ_r values varied between 5 and 65. In neutral Mondrians all three reflectance factors varied randomly between 5 and 65, while in yellow-biased Mondrians, ρ_g and ρ_r varied between 35 and 95 and ρ_b varied between 5 and 65.

In images with specularities, ρ_s was set to 0.8; in images with body reflection only, ρ_s was set to 0.0. Source color values E_i varied between 0 and 1. The constant k_d in the image irradiance equation was set so that final image irradiance values varied from 0 to 1000. To convert images to an 8-bit format, the maximum value (max) for each image was then set to 254.5, and all other pixel values scaled by the ratio $254.5/max$.

Images were generated on a SUN 3/60, transferred via EtherNet to a Symbolics 3600 Lisp machine, and displayed on a Mitsubishi color monitor hosted by the Lisp machine. The Symbolics 24-bit color map (which accessed 10-bit DACs on each gun)

was calibrated so that the 8-bit pixel values produced a linear range of luminosities for each gun, as measured by a Minolta Luminance Meter.

The spectral energy distributions of the three gun phosphors of the color monitor were measured using a Tracor multi-channel spectrometer, corrected for the sensitivity of the Tracor detectors in each wavelength interval. The spectra were interpolated to 1 nanometer wavelength intervals between 360 and 750 nanometers and normalized to the energy at the peak wavelength. The interpolated, normalized spectra were integrated with the CIE 1931 $\overline{x}$, $\overline{y}$ and $\overline{z}$ color-matching functions to give CIE tristimulus values X, Y and Z. In the calculation of the normalizing factor k_{10}, the source spectral energy distribution was set to unity for all wavelengths, so that

$$k_{10} = 100 \sum_i \overline{y}(\lambda_i) \simeq 100 \int \overline{y}(\lambda) d\lambda.$$

From the tristimulus values, the CIE chromaticity values x, y, and z were computed and from them, the matrix for transforming pixel values (r, g, b) into CIE coordinates (x, y, Y).

The set of colors at a fixed luminance (or fixed tristimulus value Y) given by a linear combination of the energies of the three phosphors forms a color triangle within the CIE x, y chromaticity plane. The vertices of the triangle are given by the x, y chromaticity coordinates of the three phosphors (see Figure 7.4). In the subsequent analysis, all image parameters and observer choices are first converted from (r, g, b) values to CIE (x, y) coordinates.

7.1.1 Control Experiment

The control set of images, designed to acquaint the observer with the task and to obtain an estimate of the observer's response variability, consisted of 6 images in random order of spheres with untextured, uniform surface reflectances, one pair – with and without specularities – for each of three illuminants. The observer was told that each sphere, whether glossy or matte, was white, illuminated by a colored light source (see Figure 7.3). He or she was asked to select the color that a flat white paper

would be under the same light source illuminating the sphere. The task, considerably simplified by the given information that the spheres were white, was reduced almost to color matching, but still difficult because the shading across the curved surface of the sphere had to be discounted to make a direct match. It enabled the observer to practice visualizing the comparison patch as a white paper and selecting its color with the mouse. The control set was presented twice, at the start and middle of a 2-3 hour session that included the following two experiments.

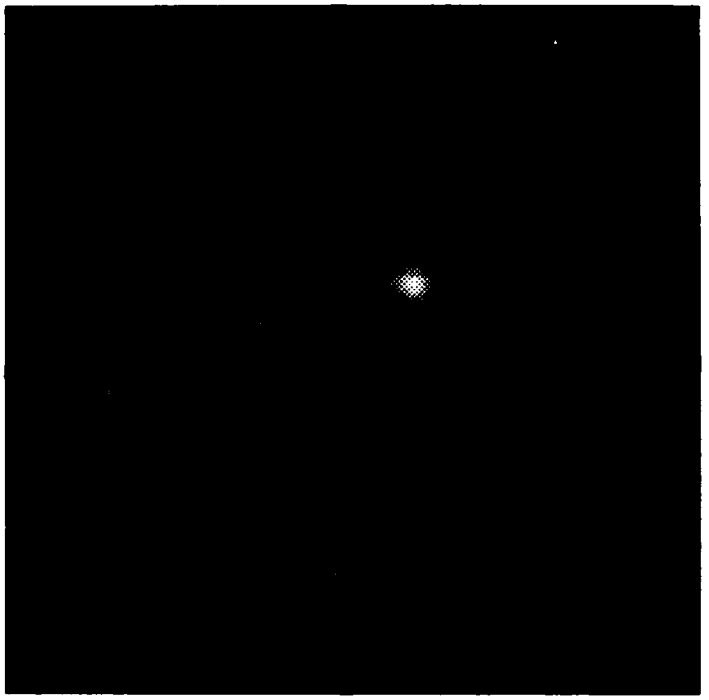

Figure 7.3: Red channel image of a sphere with uniform reflectance, with specularity.

7.1.2 Experiment I

Experiment I was designed to test whether observers rely on the grey world assumption in judging illuminant color, using images without specularities ($\rho_s = 0$). Group I consisted of 9 images generated by randomly combining each of 3 neutral Mondri-

ans with each of 3 strongly biased light sources [source Ia CIE coordinates, (0.4513 0.4283), source Ib (0.4130 0.3699), source Ic (0.2505 0.2270)]. Group II consisted of 18 images generated from 6 biased Mondrians and the same 3 light sources. Each image was presented twice for a total of 54 trials. In images generated with neutral Mondrians, the grey-world estimate of the source color correlated strongly with the actual source color, whereas in most of the images with biased Mondrians, it did not.

7.1.3 Experiment II

Experiment II was designed to test whether specular reflections influence observers' perceptions of the illuminant. Images in group I were generated from six Mondrians (two neutral, four biased) and two weakly biased light sources [source IIa CIE coordinates (.3939, .3805), source IIb (.3366, .3320)], *without* specularities ($\rho_s = 0.0$). Images in group II were identical to group I except for the addition of specularities ($\rho_s = 0.8$). Each of the 12 images in Group I was presented one or two times, while each of the 12 images in Group II was presented two or three times, yielding 48 trials total. An additional two images were added to group II, and three to Group I. In these, the color of one central patch was changed "by hand" in order to increase the distance between the average-to-grey and normalize-to-white estimates.

7.1.4 Image Parameters

Each image was characterized by 6 or 7 parameters, each a potential guide to the observer's choice of illuminant color. The first parameter, marked S in Figure 7.4, is the actual source color. G is the estimate of the source color obtained by averaging to grey, computed by scaling E_i by the ratio of the average Mondrian reflectance in the ith channel (ρ_i^{ave}) to the maximum reflectance in all channels (ρ_{max}):

$$G_i = \frac{\rho_i^{ave}}{\rho^{max}} \cdot E_i.$$

The average reflectance is taken over all pixels in the Mondrian image (i.e., the reflectance of each patch is weighted by its area). Note that this estimate is not

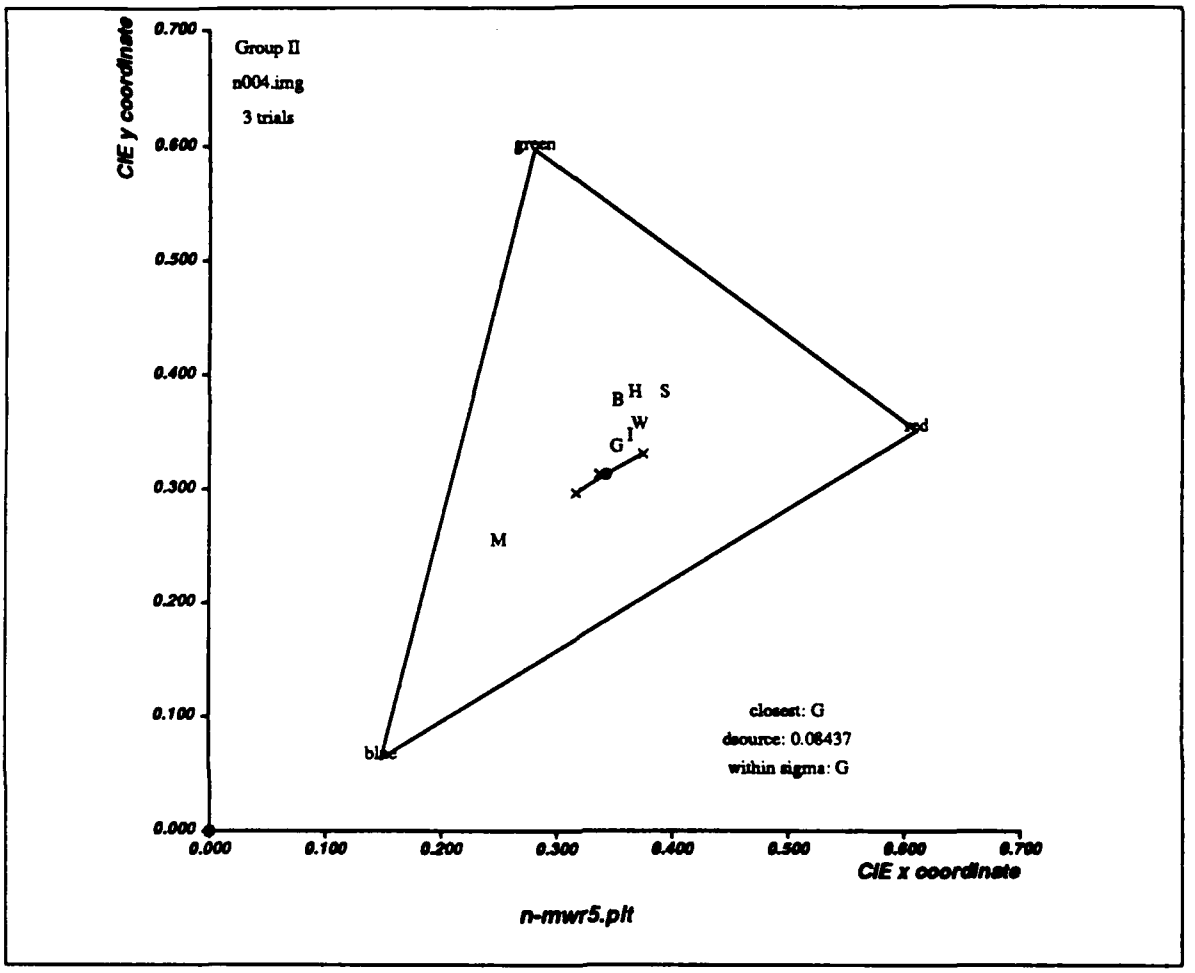

Figure 7.4: Illustration of variability in observer choice. Observer MWR's choice of source color (points marked by small x) for a yellow-biased Mondrian illuminated by a pinkish source (source IIa) with specularities. S marks the actual source color, W the source color obtained by normalizing to white, G the source color obtained by averaging to grey, I the average image irradiance color, H the image irradiance color at the peak of the specularity, M the average Mondrian reflectance, and B the color at the brightest image point. Text in the upper-left hand corner indicates the group to which the image belongs and the number of trials (in this case three). Text in the lower right-hand corner indicates the image parameter closest to the observer's choice; the distance between the observer's choice and the source color, averaged over all trials of this image; and the image parameters that fall within distance sigma of the observer's choice, where sigma is the square root of variance of the subject's choices, averaged over all trials of all images *within a group*.

precisely the same as that predicted by Land's retinex theory. The retinex model, as described in section 4.1.1, would scale E_i by

$$\frac{\overline{log\rho_i}}{log\rho^{max}},$$

where $\overline{log\rho_i}$ denotes the average of the logarithm of ρ_i taken over the entire image.

I is the average image irradiance, computed by averaging the irradiance values of the final image pixel by pixel. Thus I is effectively the estimate predicted by Evans' formulation of the grey-world assumption[30]: to balance color photographs, Evans scaled the pixel values in the original image so that the final image would average to grey. This method is based on the assumption that each scene has a random collection of colors (the grey-world assumption) and that if the illumination is white, the scene will average to grey.

W is the estimate obtained by normalizing to white, computed by scaling E_i by the ratio of the maximum Mondrian reflectance factor in the ith channel (ρ_i^{max}) to the maximum reflectance in all channels (ρ^{max}). That is,

$$W_i = \frac{\rho_i^{max}}{\rho^{max}} \cdot E_i.$$

So the normalize-to-white estimate of the illuminant in one channel varies proportionately with the highest luminance in that channel.

B is the color of the patch with the highest luminosity. H is the color at the peak of the specularity in the final image. (Because the light source position and sphere radius are the same for each image, the point of maximum source radiance is the same on each sphere whether or not specular reflections are present. This locus is not always the same as that of the maximum luminance of the final image (B), which depends also on the reflectances of the patches nearest to the light source.) M is the average color of the input Mondrian, computed by averaging pixel by pixel. W should approximate S only when the reflectances of the brightest patches in each color channel are the same (when the "whites" in each channel are the same). G

approximates the source color when the average reflectance in each color channel is the same.

Figure 7.5 illustrates that the illuminant color estimate given by the chromaticity convergence algorithm is exactly the source color. Note, though, that the convergence lines are short and therefore provide a relatively weak signal. This is due to the fact that the specularity, which falls off as $cos^{40}\theta$, is very concentrated, so that only at the borders of the specularity does the image irradiance vary slowly enough to create a spread of chromaticity values along a line. The central region of the specularity registers as a separate patch, the color of which is a linear combination of the underlying Mondrian reflectance and the source color.[4]

7.2 Results

Figure 7.6 illustrates observer DSC's performance on the control experiment. In (a) are plotted the results from six trials (two presentations for each of three images) of uniform-reflectance spheres without specularities; in (b) the same for images with specularities. In (a) and (b) the choice variability and accuracy are the same. DSC is a typical observer; some estimated the source color more accurately, others less. Table 7.1 depicts similar results for all 7 observers. For each observer there is no significant difference between the average choice variances for control images with and without specularities (for all there is approximately 50% or greater chance of being wrong in rejecting the hypothesis that the two variances are equal). Although for two observers (JK, MWR) the source-choice distance in images without specularities is approximately one-half that of images with specularities, for neither is the difference significant (for MWR, α is less than 0.2). [5] Note that, although ENJ's mild color

[4]The stray dots in the diagram are largely due to quantization noise introduced by the conversion of real values to 8-bit values in the final image. In the chromaticity diagram of the whole image, the convergence signal is more difficult to see in the swarm of noise.

[5]Note that in computing the t-statistic, the source-choice distance variance, obtained by pooling the source-choice distances from all trials within one group, is used rather than the choice variance. This is on the assumption that only one parameter, the actual source color, characterizes each image. Furthermore the observer's task is simply to match that color. Therefore on repeated trials of all

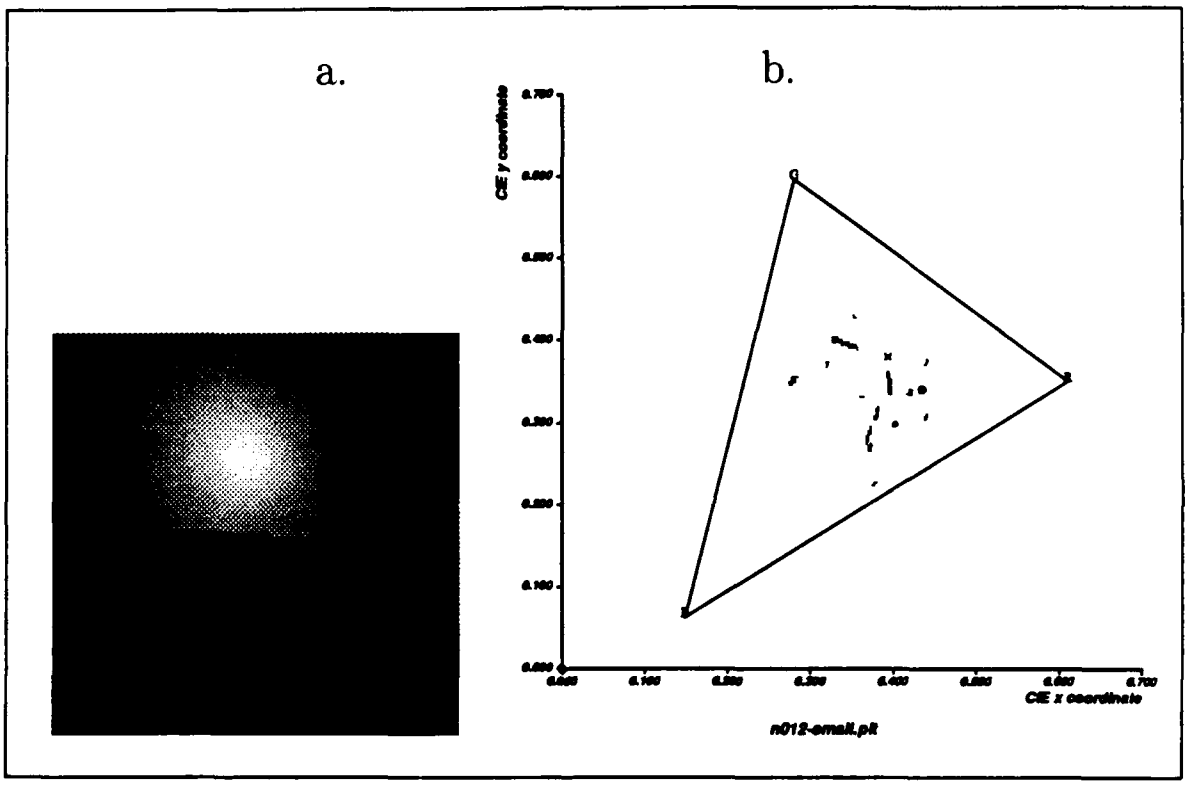

Figure 7.5: (a) A 75x75 pixel region of the red image of the Mondrian-textured sphere, including the specularity. (b) Chromaticity convergence diagram of the region, in which the cross marks the source color. Three short lines clearly intersect at the cross, providing the source signal.

blindness might be expected to affect the variability of his choices, it appears not to when compared with other observers (see Table 7.1 and Figure 7.8).

7.2.1 Experiment I

Table 7.2 summarizes the results from Experiment I (excluding the six trials for which any image parameter fell outside the range of palette choices, i.e. those of a blue-biased Mondrian under the bluish illuminant, source Ic). The average distance between the source (S) and choice (C) is higher for biased Mondrians (Group II) than for unbiased (Group I), whereas the average distance between the grey-world estimate (G) and observer choice is lower in Group II than in Group I. The average variance of the choice remains roughly the same in the two groups. In the data pooled from all observers, the average distance between S and C in Group I is significantly lower than that in Group II ($\alpha < 0.005$). Furthermore, whereas the distance between C and S is significantly higher than the distance between C and G in Group II ($\alpha < 0.005$), the two distances are not different from one another in Group I ($\alpha >> 0.20$).[6] These results suggest that for unbiased Mondrians (which uphold the grey-world assumption), the observer's choice is close to the source only because the source is close to the grey-world estimate. Consequently, for biased Mondrians (which violate the grey-world assumption), when the source is further away from the grey-world estimate, (see Table 7.3) the observer's choice is closer to the grey-world estimate than to the source.

Similarly, for biased Mondrians observer choice is closer to the normalize-to-white estimate (W) than to the source. In fact, for both groups, observer choices are roughly equally close to both G and W. This result is not unexpected since, because

images within a group, the source-choice distance should fall in a distribution with a constant mean and variance. This is not true for images in experiments I and II, as discussed below. Making this assumption increases the degrees of freedom (d.o.f.) in the t-statistic test here and therefore increases the confidence in the statement that the mean distances are the same in the two groups.

[6]To compute the t-statistics for these tests, the variances of the *distances* (between S and C, and between C and G) rather than the variance of the choice is used. These variances are only slightly different from one another, as expected.

observer	*image type (N)*	*ave. distance source-choice*	*average* σ_{choice}	*average* σ^2_{choice}	F	α	*t-stat.*
DSC	I. no spec (3)	.020160	.009732	.000095	1.309	$\simeq 1$	-0.0985
	II. spec (3)	.020670	.008507	.000072			
ENJ	I. no spec (3)	.016644	.018457	.000341	.9893	$\simeq 1$	-0.274
	II. spec (3)	.019684	.018556	.000344			
K	I. no spec (3)	.019667	.011329	.000128	.2424	< 0.5	-1.148
	II. spec (3)	.035084	.023009	.000529			
MEK	I. no spec (3)	.016524	.017978	.000323	3.426	< 0.5	-0.508
	II. spec (3)	.022885	.009713	.000094			
MSD	I. no spec (3)	.020304	.025686	.000660	3.198	$\simeq 0.5$	0.742
	II. spec (3)	.012770	.014364	.000206			
MWR	I. no spec (3)	.011006	.007805	.000061	.3597	$\simeq 0.5$	-1.970
	II. spec (3)	.022192	.013014	.000169			
PWU	I. no spec (3)	.004207	.007402	.000055	.5254	< 0.5	-0.570
	II. spec (3)	.006710	.010212	.000104			
pooled observers	I. no spec (21)	.015502	.015410	.000237	1.093	$\simeq 1$	-1.221
	II. spec (21)	.019999	.014738	.000217			

Table 7.1: Results for control experiment, comparing observer performance on images with and without specularities. The average source-choice distance is computed by averaging over all trials of all images within a group (6 total). The average variance of the choice, which gives a measure of observer response variability, has N-1 degrees of freedom (d.o.f.), where N is the sum of d.o.f. (= 1) of each sample variance contributing to the average. Each sample variance is computed for 2 trials of one image. F is the ratio of σ^2 (average choice variance for images without specularities) to σ^2_H (average choice variance for images with specularities). α is the level at which the hypothesis $\sigma^2 = \sigma^2_H$ may be rejected – e.g. for observer MWR there is approximately 50% chance of being wrong in rejecting the hypothesis. The t-statistic provides a test of the hypothesis that the average source-choice distance is the same for the two groups of images. It is computed using the pooled variance from the two groups, with d.o.f. = (total number of images - 2). Only for observer MWR may the hypothesis $\mu_I = \mu_{II}$ be rejected with α less than 0.2, which is not significant. All other t-statistics correspond to α approximately 0.4.

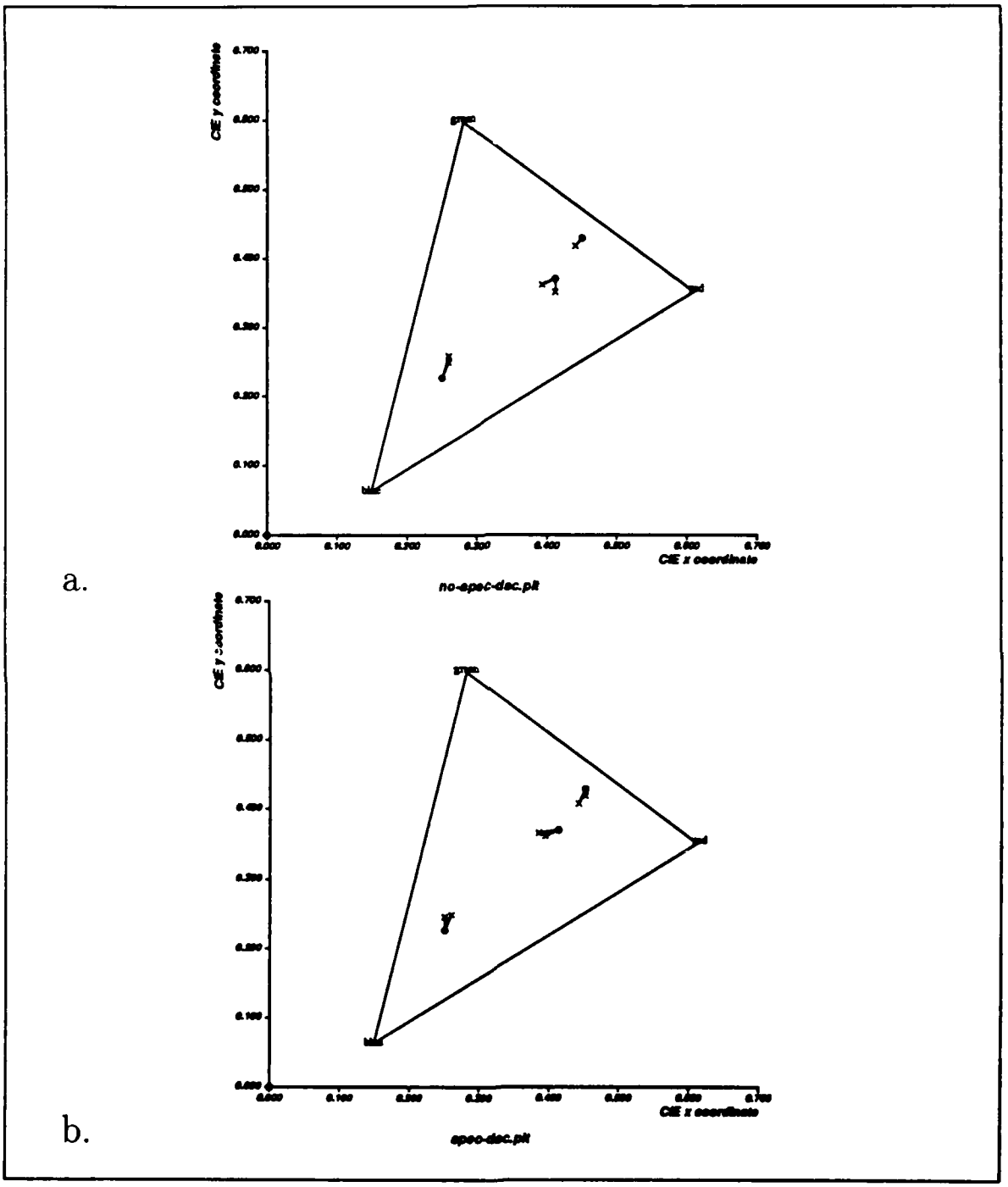

Figure 7.6: Observer DSC's performance on the control experiment: (a) images without specularities; (b) images with specularities. Dots mark actual source color; each cross connected to a source color marks observer's estimate of it on one trial (two trials per source).

observer	*image group (N)*	*average distance between choice and* S	G	W	σ_{choice}
DSC	I. U (20)	.0363	.0355	.0373	.0190
	II.B (30)	.0405	.0322	.0261	.0257
ENJ	I. U (20)	.0367	.035	.0375	.0176
	II.B (30)	.0446	.0236	.0292	.0165
MEK	I. U (20)	.0421	.0409	.0421	.0243
	II.B (29)	.0479	.0297	.0314	.0206
MSD	I. U (19)	.0383	.037	.0378	.0251
	II.B (30)	.0452	.0299	.0278	.0309
PWU	I. U (19)	.0237	.0242	.0242	.0326
	II.B (30)	.0475	.0307	.0314	.0291
pooled observers	I. U (98)	.0355	.0346	.0359	.0241
	II. B (149)	.0451	.0292	.0292	.0252

Table 7.2: Results for Experiment I, comparing observer performance on images in Group I (unbiased Mondrians) and GroupII (biased Mondrians). N is total number of trials contributing to average distance between choice and: source (S); grey-world estimate of illuminant (G); and normalize-to-white estimate (W). σ_{choice} computed as in Table 7.4. Note that if the variance is calculated in a uniform color space such as *Lab*, it still does not vary significantly with the source color.

image group	*d(S, G)*	*d(S, W)*
I. Unbiased	0.0043	0.0030
II.Biased	0.0490	0.0274

Table 7.3: Distances between source and grey-world estimate (G), and between source and normalize-to-white estimate (W) for each group (I – unbiased Mondrians; II – biased Mondrians), averaged over all images within the group.

of the strong bias in sources, G and W tend to cluster together. (S, G, and W are inseparable on all images in Group I and on almost half in Group II (see Figure 7.10).) On the six trials for which the distance between G and W was appreciable (images of yellow-biased Mondrians under the bluish illuminant, source Ic) two subjects (ENJ, MEK) consistently made choices close to G, one (DSC) consistently chose W, and two subjects were inconsistent. Figure 7.7 illustrates DSC's choices. Note, though, that observers would be more *accurate* if they chose W over G, as the distance between S and W is lower than the distance between S and W for both groups (see Table 7.3).

Although all observers followed the trend described above, there were slight differences and peculiarities amongst individual responses. Figure 7.8 illustrates the range of observer variability, measured on 4 trials of one image in Group I. Figure 7.9 illustrates for one observer (PWU) his occasional inexplicable choice of the actual source color when it was distant from all other image parameters. In this image of a yellow-biased Mondrian under the bluish illuminant (source Ic), the grey-world estimate of the source is close to neutral, yet PWU is able to determine the source color as blue, evidently using other cues. Figure 7.10 illustrates one observer's (MEK's) tendency to choose sources less saturated than the grey-world estimate predicts.

7.2.2 Experiment II

Whereas in the control experiment, the presence of a specularity influenced neither the accuracy nor variance of the choice, in Experiment II specularities appeared

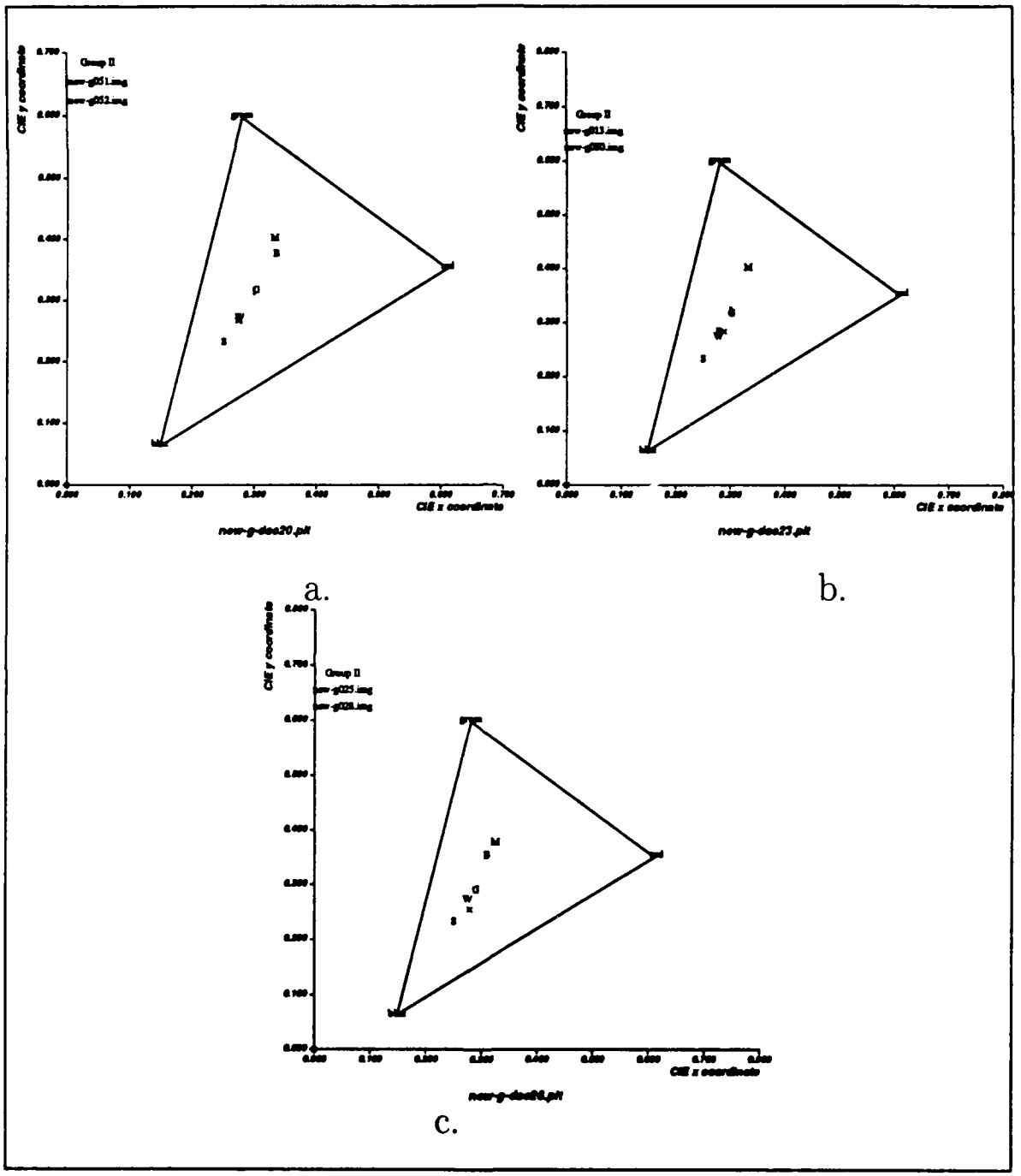

Figure 7.7: Observer DSC's choice of source color for three images (two trials each) in which the grey-world estimate G and the normalize-to-white estimate W are well separated. Note his tendency to choose colors close to W.

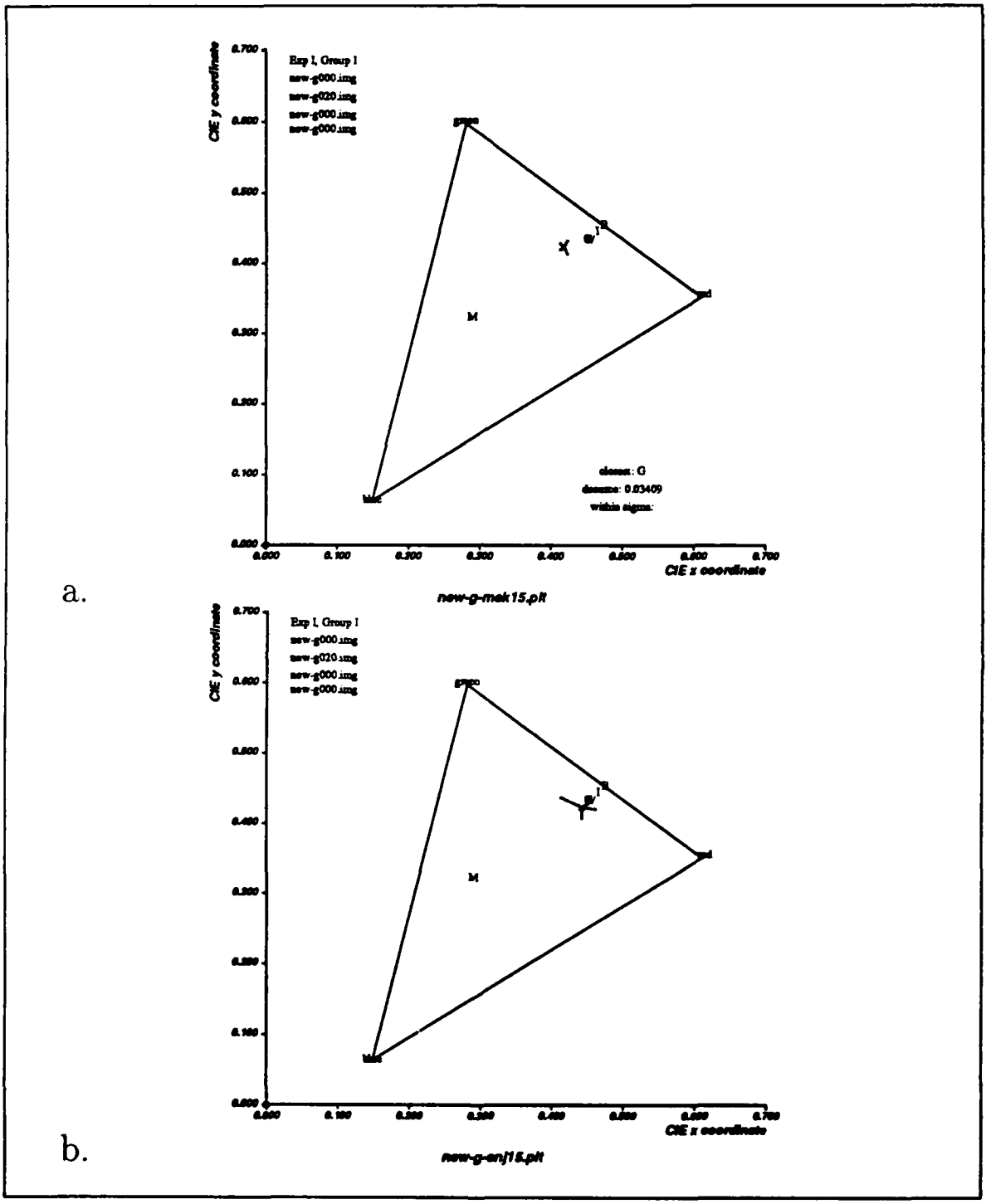

Figure 7.8: Choice variability on 4 trials of one image in Group I, Experiment I. MEK (a) and ENJ (b) represent the two extremes of variability, although ENJ, despite his mild protanopia, is not alone at the high end.

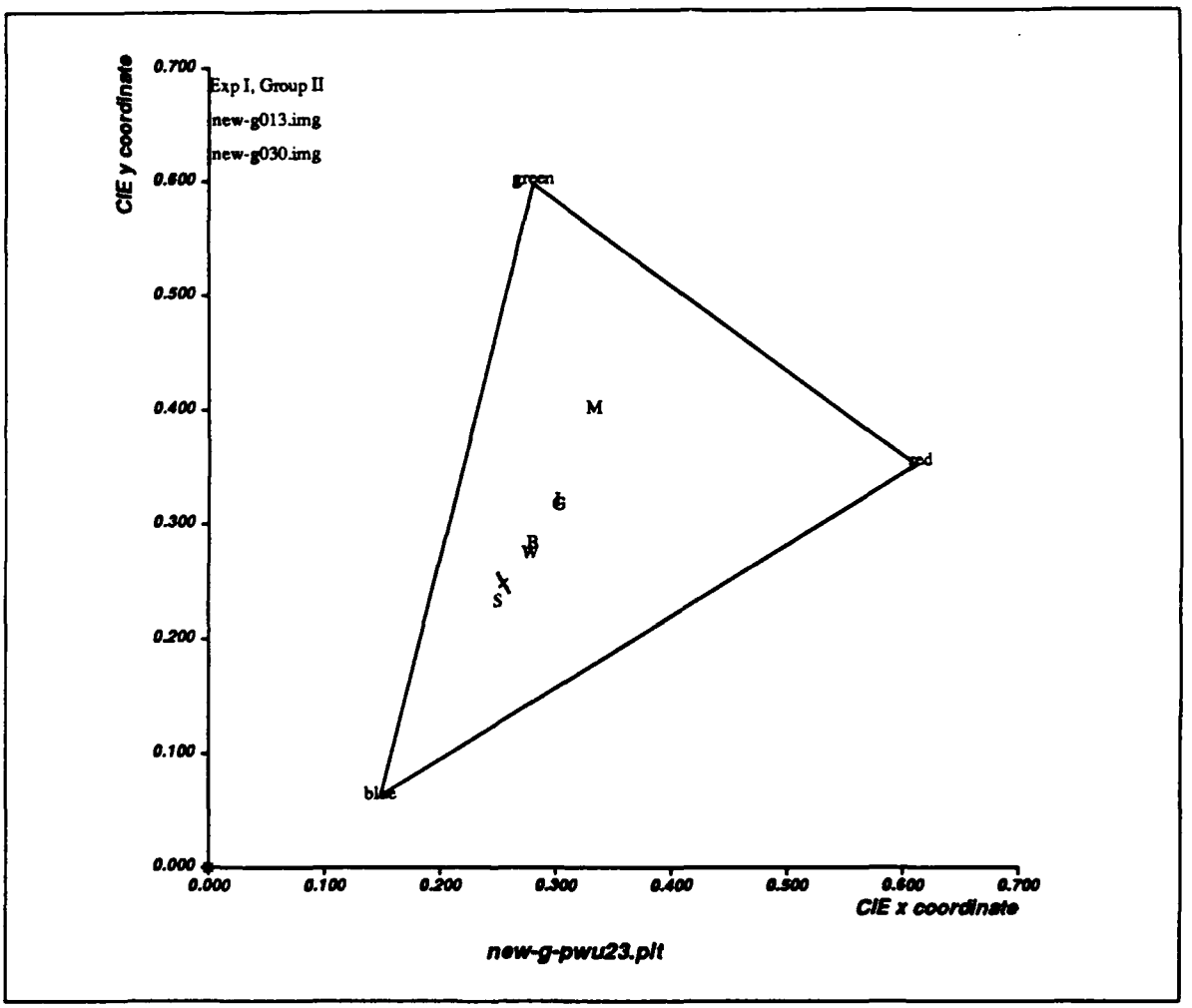

Figure 7.9: Illustration of observer PWU's occasional accuracy of source choice in absence of obvious cues.

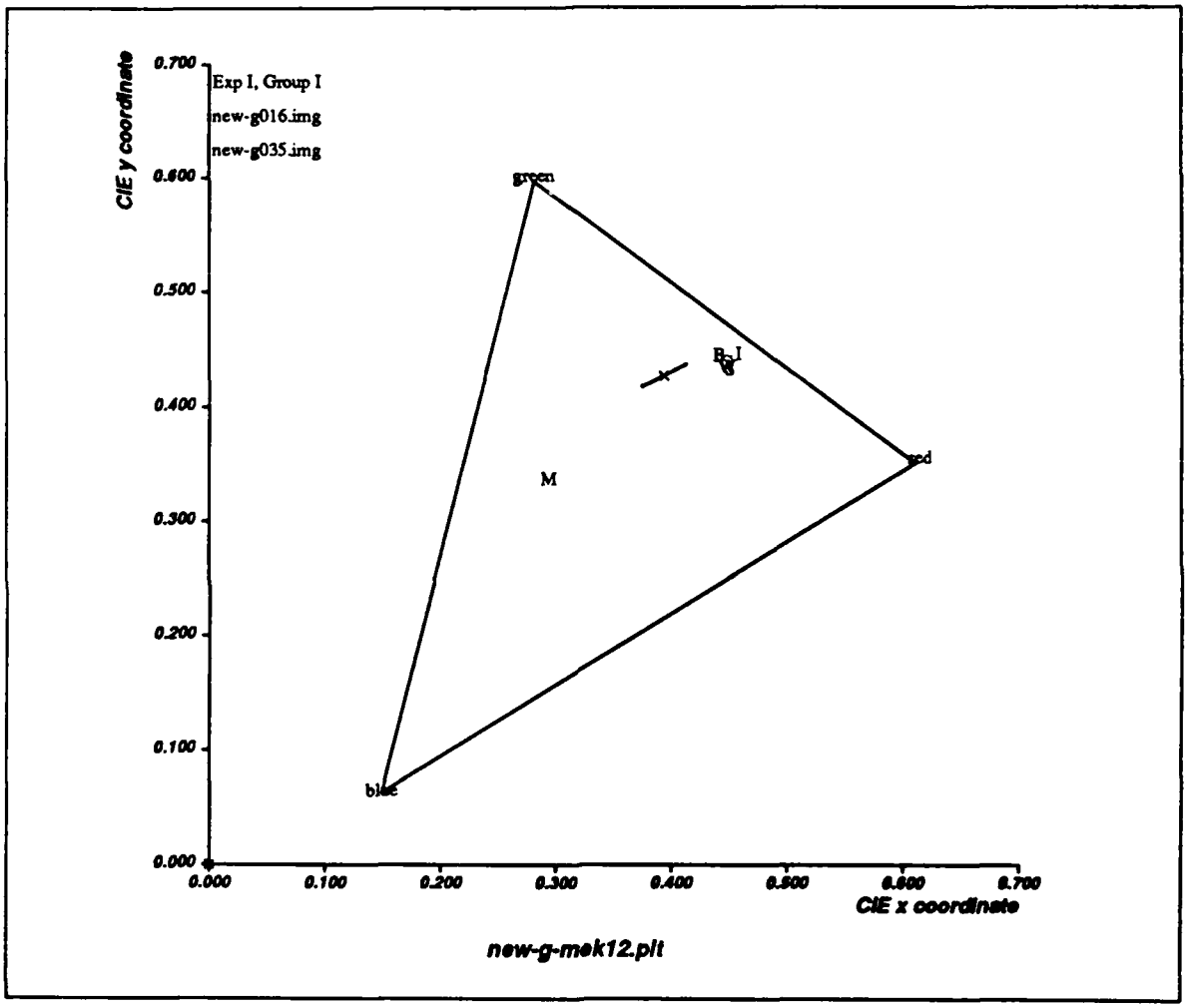

Figure 7.10: Illustration of observer MEK's tendency to choose colors less saturated than either grey-world estimate or actual source color.

observer	*image type (N)*	*ave. distance source-choice*	*average* σ_{choice}	*average* σ^2_{choice}	F (σ^2/σ^2_H)	α
DSC	I. no spec (7)	.0399	.020934	.000438	.2655	< 0.2
	II. spec (17)	.0490	.040623	.001650		
JK	I. no spec (7)	.0541	.025911	.000671	.4029	$< .3$
	II. spec (18)	.0588	.040823	.001666		
MSD	I. no spec (7)	.0478	.038955	.001517	2.242	< 0.2
	II. spec (18)	.0497	.020615	.000677		
MWR	I. no spec (7)	.0525	.045936	.002110	4.217	< 0.05
	II. spec (18)	.0510	.022384	.000501		
PWU	I. no spec (7)	.0778	.020667	.000427	.3864	< 0.4
	II. spec (18)	.0778	.033248	.001105		
pooled observers	I. no spec (35)	.0544	.032138	.001033	.930	$\simeq 1$
	II. spec (89)	.0573	.033332	.001111		

Table 7.4: Results for Experiment II, comparing observer performance on images without specularities (Group I) and with specularities (Group II). The average source-choice distance is computed by averaging over all images within a group. The average variance of the choice, which gives a measure of observer response variability, has N-1 degrees of freedom (d.o.f.), where N is the sum of d.o.f. of each sample variance contributing to the average. Each sample variance is computed for 2-4 presentations of one image. F is the ratio of σ^2 (average choice variance for images without specularities) to σ^2_H (average choice variance for images with specularities). α is the level at which the hypothesis $\sigma^2 = \sigma^2_H$ may be rejected – e.g. for observer DSC there is less than 20% chance of being wrong in rejecting the hypothesis.

to influence choice variance, increasing it for three observers, decreasing it for two. For data pooled from all observers, these opposing influences cancel, resulting in no difference between groups. For all observers, choice variance was higher than in the control experiment for images both with and without specularities, in part reflecting the increased difficulty of the task. For all but one observer (PWU), the standard deviation was approximately 50% or greater of the mean source-choice distance. Figure 7.4 illustrates this variability for the one observer (MWR) for whom the difference in variance between the two groups is significant (α less than 0.05).[7]

As in Experiment I, for pooled observer data here the average distance between S and C was significantly higher for biased Mondrians than for unbiased Mondrians, with or without specularities, whereas the distance between S and G was significantly lower, as illustrated in Table 7.5. Table 7.5 also illustrates that, comparing groups of like-biased Mondrians, specularities consistently neither improve nor worsen the choice accuracy (as measured by the average distance between S and C). Figure 7.11 illustrates the tendency for observers to choose estimates close to G even when specularities provide conflicting information.

Table 7.6 illustrates these results in another way. The correlation coefficient between the choice and the grey-world estimate (averaged over all images within a group, using data pooled from all observers) is higher for biased Mondrians (Groups Ib, IIb) than for unbiased Mondrians (Group Ia, IIa), and, for biased Mondrians, higher than the correlation coefficient between choice and source. Because these groups included images in which G and W were more widely separated than in Experiment I, one could

[7]Note that the choice variance here is computed in the most conservative way possible. The difference between groups would be more significant if there were a parameter for which we could assume that the parameter-choice distance sampled over all images within a group forms a normal distribution with a single mean and variance. The variance of that parameter-choice distance could then be used as the choice variance. This would mean presupposing the parameter which guides the observer's judgment of illuminant color, by assuming that the observer attempts to maintain a constant (zero or non-zero) distance from the parameter. Instead, we make the less restrictive and more plausible assumption that the observer's choice forms a normal distribution when sampled over repeated trials of one image, and that the distributions for different images within a group have different means but the same variance.

observer	*image group (N)*	*average distance between choice and* S	G	W	σ_{choice}
DSC	Ia. no spec-U (10)	.0312	.0317	.0323	.0151
	Ib. no spec-B (12)	.0472	.0310	.0218	.0244
	IIa. spec-U (12)	.0376	.0406	.0367	.0311
	IIb. spec-B (19)	.0562	.0406	.0396	.0452
JK	Ia. no spec-U (10)	.0473	.0472	.0435	.0219
	Ib. no spec-B (12)	.0597	.0353	.0342	.0285
	IIa. spec-U (12)	.0693	.0726	.0672	.0578
	IIb. spec-B (20)	.0524	.0413	.0409	.0288
MSD	Ia. no spec-U (10)	.0351	.0375	.0392	.0384
	Ib. no spec-B (12)	.0585	.0329	.0310	.0394
	IIa. spec-U (12)	.0329	.0340	.0347	.0313
	IIb. spec-B (20)	.0598	.0300	.0336	.0229
MWR	Ia. no spec-U (10)	.0324	.0341	.0352	.0524
	Ib. no spec-B (12)	.0692	.0448	.0492	.0404
	IIa. spec-U (12)	.0268	.0297	.0296	.0250
	IIb. spec-B (20)	.0655	.0236	.0359	.0209
PWU	Ia. no spec-U (10)	.0570	.0612	.0595	.0247
	Ib. no spec-B (12)	.0951	.0341	.0556	.0171
	IIa. spec-U (12)	.0683	.0693	.0657	.0448
	IIb. spec-B (20)	.0836	.0345	.0518	.0256
pooled observers	Ia. no spec-U (50)	.0406	.0423	.0419	.0333
	Ib. no spec-B (70)	.0659	.0356	.0384	.0313
	IIa. spec-U (50)	.0470	.0492	.0468	.0398
	IIb. spec-B (99)	.0636	.0339	.0404	.0295

Table 7.5: Results for Experiment II, comparing observer performance on images in Group Ia (unbiased Mondrians, without specularities), GroupIb (biased Mondrians, without specularities), Group IIa (unbiased Mondrians, with specularities) and Group IIb (biased Mondrians, with specularities). N is total number of trials contributing to average distance between choice and: source (S); grey-world estimate of illuminant (G); and normalize-to-white estimate (W). σ_{choice} computed as in Table 7.4.

	correlation coefficient between					
	choice and			*choice-source dist. and*		
image group	*S*	*G*	*W*	*d(S, G)*	*d(S, W)*	*d(S, B)*
Ia. no spec-U ()	.5759	.5763	.5197	.1436	-.1813	-.1119
Ib. no spec-B ()	.5162	.8459	.8451	.4111	.4916	.1587
II.spec-U ()	.6915	.6920	.5875	.1737	-.1755	-.1492
II. spec-B ()	.5012	.8796	.8645	.5572	.4769	.3371

Table 7.6: Columns 1-3: correlation coefficients between choice and: source (*S*; grey-world estimate (*G*); and normalize-to-white estimate (*W*). Columns 4-6: correlation coefficients between source-choice distance and: distance between source and grey-world estimate (*d(S, G)*); distance between source and normalize-to-white estimate (*d(S, G)*); and distance between source and most luminous patch (*d(S, B)*). All distances and parameters averaged over all images within a group.

have expected to see a difference in the correlation coefficients for the two parameters if either were preferred as a guide to the illuminant color. In fact, no appreciable difference exists. If the observer's choice did tend to correlate more strongly with G, then the choice distance from S $[d(C,S)]$ should be low when the distance of S from G $[d(S,G)]$ is low, and high when the latter is high. In other words, the correlation between $d(C,S)$ and $d(S,G)$ should be high. Although it is high for biased Mondrians, it is not significantly higher than the correlation between $d(C,S)$ and $d(S,W)$. Thus, the results are again inconclusive in distinguishing between averaging-to-grey and normalizing-to-white as candidate methods for determining the illuminant color.

Almost all observers who participated in the experiment formed a strong impression of the illuminant color on first viewing of each sphere. (One observer (MWR) commented that whereas he could be "quite good" at "guessing the color of the object, [it was] a curiously difficult thing to guess the color of the illuminant." He performed the task, he felt, by "answering the question: what hue is this sphere biased toward," i.e. by presupposing the grey-world assumption.) Most observers complained that they could not reliably narrow their choice to a single rectangle in the palette. Thus

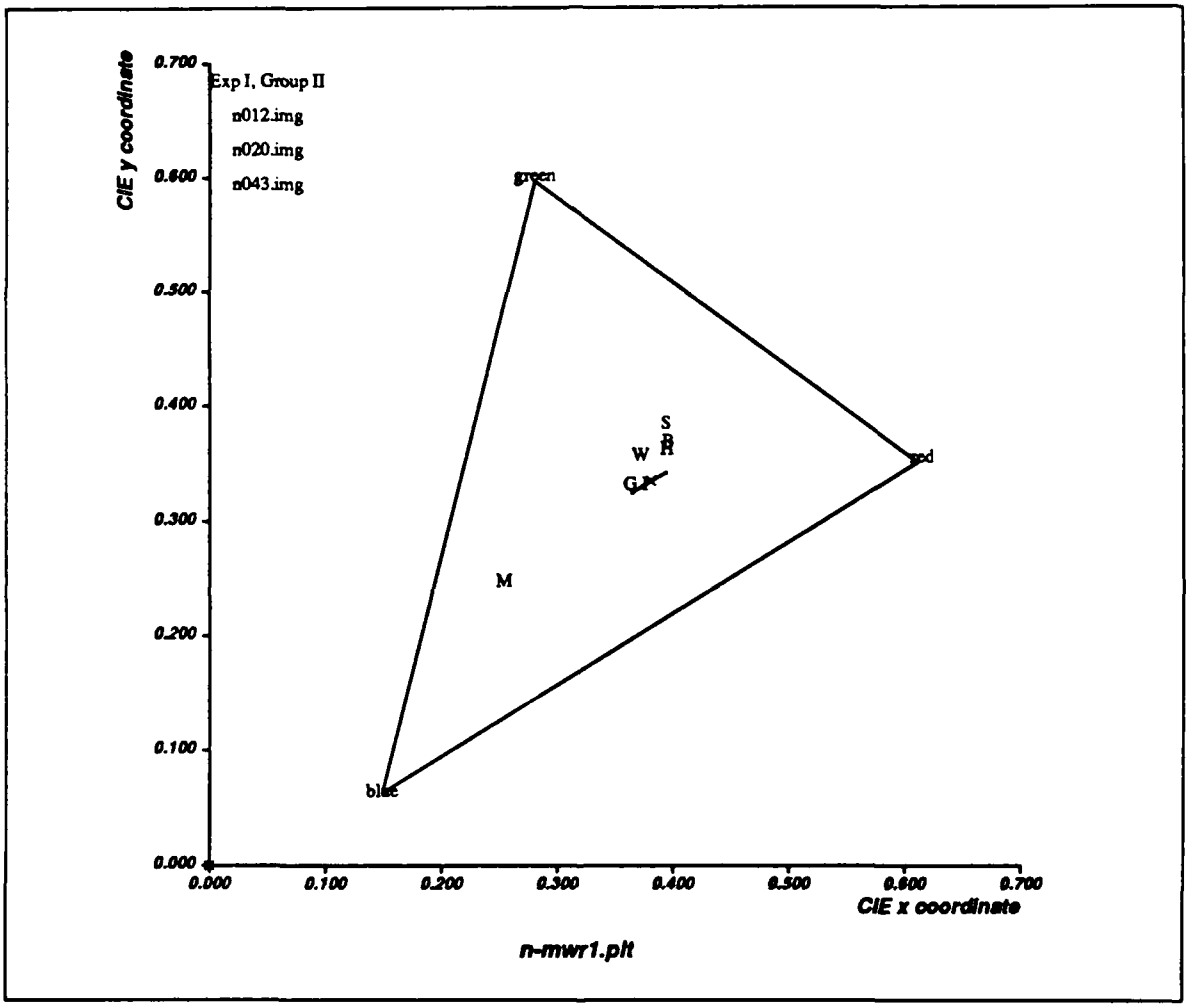

Figure 7.11: Illustration of observers' tendency to choose estimates close to the grey-world estimate even when specularities are present. (This example shows MWR 's choices for a blue-biased Mondrian under the pinkish illuminant, source Ia.)

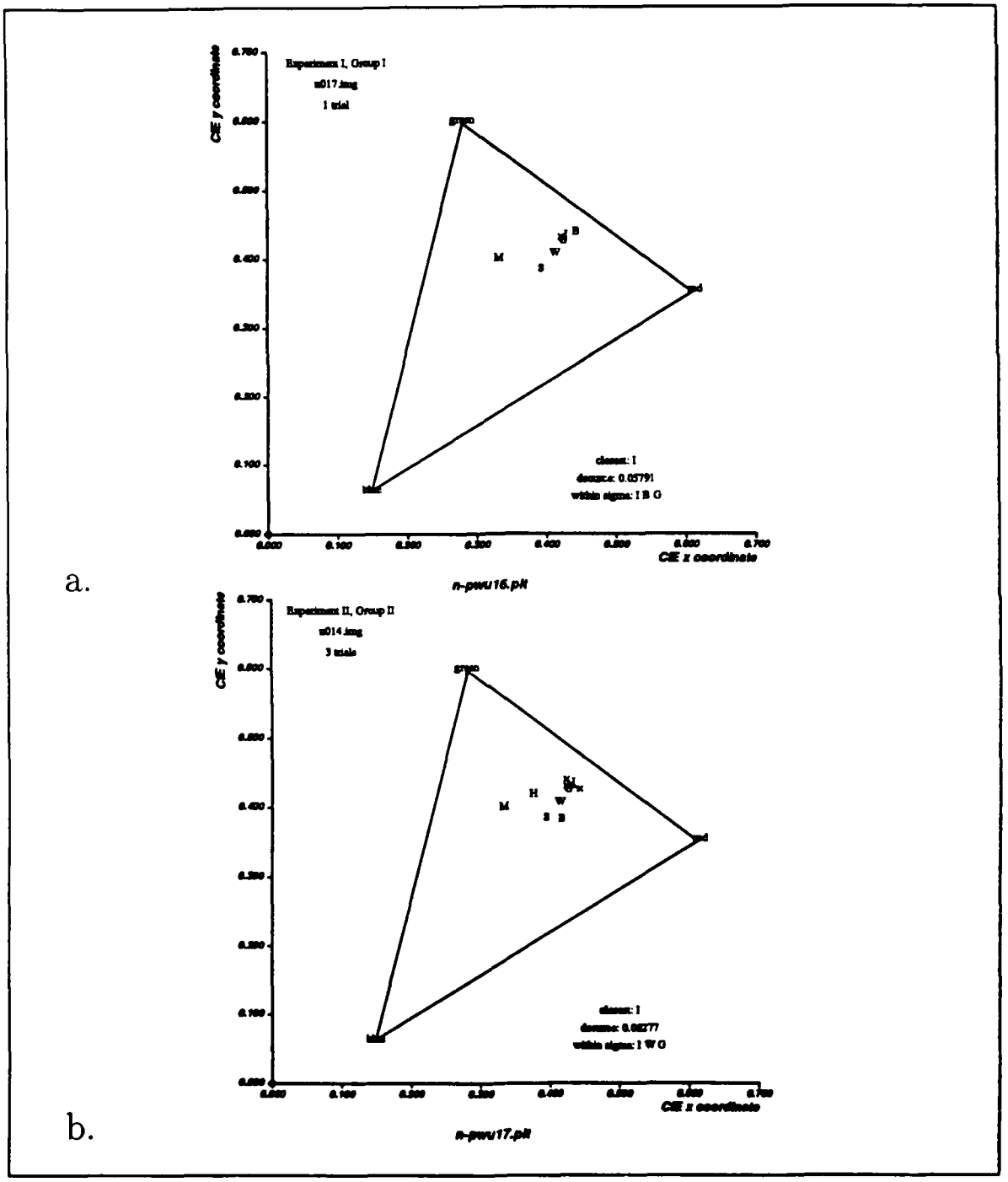

Figure 7.12: Observer PWU's choice of source color (points marked by small x) for a neutral Mondrian illuminated by a pinkish source (IIa): (a) without specularities, one trial; (b), with specularities, three trials. Note that PWU's choice varies little between the two images; the presence of a highlight in (b) appears not to influence his choice.

although the colors in the palette were easily discriminable as reflectances, they were not as illuminants. This coarseness of judgment accounts for at least part of the variability in observer choice. Some observers also noted that the first impression might fade after several minutes of viewing, but in a way that affected only the saturation, not the hue, of their final choice.

All but one observer who participated in the experiments, when asked to consider which aspect of the image might have served as a guide to the illuminant color, cited the average color of the sphere, e.g. "the overall cast of color," "the average", "the shading, tone", the "dominant color." One observer (MSD) felt that he was guided more by individual patches rather than a global average: he either "looked for the most brilliant color," to rule out its complementary color as the source, or "looked for the cast of the greyest patch." Despite the lengthy time interval between successive trials, successive color contrast effects were still strong enough to induce one observer (who did not participate in the full experiment) to comment that he could judge the illuminant color only by using – arbitrarily – the previous sphere as a reference "white."

Two observers claimed not to have noticed the highlights that appeared in Experiment II. One (MEK) commented that highlights made the task easier and one (JK) that it made the task harder. Others noted that highlights had no influence on their choices.

Spheres in which the Mondrian reflectance distribution was biased oppositely from the actual source color – e.g. a yellowish sphere under a bluish light – appeared odd to observers aware of the image generation technique (none of whom were included in the experiment). Naive observers, even when pressed to comment on any unusual features of the images, did not mention this effect. Thus the conflict between cues from specularities and the grey-world assumption did not rise to the level of conscious awareness.

7.3 Discussion

The above results indicate that, at least for these artificial images, the human visual system relies on the grey-world assumption (or on the closely-related method of normalizing to white) to estimate the illuminant color, even when cues from specular reflections provide conflicting information. Whereas for all observers the presence of specularities influenced choice variance, for none did it increase the accuracy of the choice.

A number of caveats apply to these conclusions. First, although we can judge the illuminant color easily and immediately, the judgment is precise only on a scale much coarser than that on which we can discriminate reflectances. Second, the axiom "correlation does not imply causation" reminds us that we cannot rule out other methods of computing the illuminant color simply because observer responses correlate with the grey-world estimate. Third, the phrase *at least for these artificial images* is critical. Despite the robustness of the results for computer-simulated scenes, they might be radically different for natural scenes. Computer-generated images have the advantage of being easily and quantitatively manipulated, but the disadvantage of not being real. Although the physical stimulus produced by a real sphere under a real light differs from that produced by the computer image that simulates it, the retinal responses to the two stimuli are the same. Yet despite the equivalence of the pixel-by-pixel retinal signal, the observer can usually use other cues to distinguish a real scene from its computer simulation. For example, in the real world it is rare to see surfaces with perfectly uniform chromaticity values. Furthermore, Gilchrist has shown that in ambiguous situations, we use high level information to distinguish reflectance from illumination edges, information that is missing in most computer simulations. (In the images used here, high level information concerning the position and number of light sources is supplied at the outset, and the edges are not ambiguous.) The range of light intensities produced by a CRT monitor is also severely limited in comparison with the range in the natural world, reducing the amount of information that can

be conveyed in the size of intensity edges. Finally, at some level the observer must face the paradox raised by knowing that the monitor is a self-luminous object and believing that the sphere within it is lit by source outside the monitor. The question is whether that paradox alters the computation the observer makes to judge the illuminant color. Unfortunately the answer is not known, despite much discussion on the merits of computer simulations in visual psychophysics.

Some visual illusions more striking in reality than in computer simulations, presumably because the greater abundance of cues in the real version creates more opportunity for conflicting interpretations. The Mach Card illusion, for example, cannot be reproduced at all by computer simulation (pers. comm., H. H. Bülthoff). Given that, it is possible that observers would be more consistent and confident in their responses if asked to judge the illuminant color on real scenes, but still not certain that they would continue to ignore information from specular reflections. Despite the fact that, as Figure 7.5 illustrates, there appears to be sufficient information from specular reflections to determine the source color, the question remains: is it enough? The simple fact that the specularities in the images used here are simulated might rule them out as cues to the illuminant color. (No matter how realistic they appear, they still do not behave as real specularities. For example, the requirement that the observer's angle of gaze be fixed obviated the problem that simulated specularities do not move with angle of gaze, but there might be other features for which we did not control.) More importantly, in a natural scene, variations in pigment density, surface orientation, and strength of specularities might provide a far more overwhelming convergence signal. The visual system might have good reason to discard information deliberately when it arises from a single specularity, while using such information when it is confirmed at several locations across the scene. That is, the visual system might operate on the single source assumption, preferring to make a global estimate that applies fairly well across a scene rather than local estimates that might vary from point to point. When the assumption is violated by information from a single location, the system simply reinstates it by disregarding that information. In this

sense, the results reported here may be taken as support for the argument that the computation of the illuminant color is global rather than local.

Perhaps the strongest conclusions that can be drawn are (1) human observers are capable of judging confidently the illuminant color without given a reference "white", at least for computer simulations of simple scenes without object interreflections; (2) they appear to do so by averaging to grey or normalizing to white; and (3) specularities do not play a role in this judgment.

What do these results imply for color constancy? May we conclude that humans use the same method to judge the illuminant color as to compute constant colors? Is color constancy consonant with (a) the degree of variability in observers' perception of the illuminant color and (b) the errors predicted by violations of the grey-world assumption? Answers to these questions await a more quantitative exploration of color constancy.

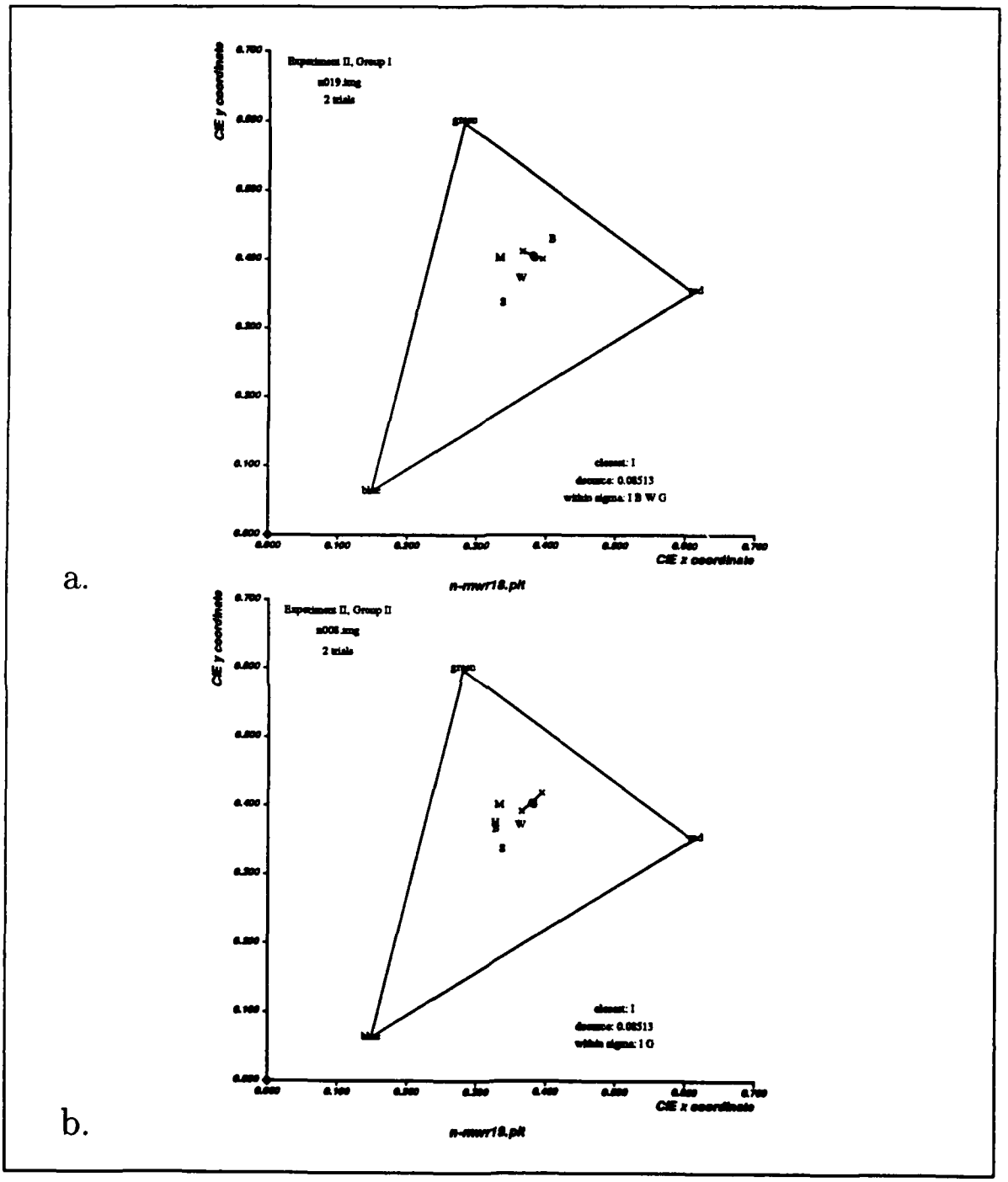

Figure 7.13: Observer MWR's choice of source color (points marked by small x) for a neutral Mondrian illuminated by a pinkish source (IIa): (a) without specularities, two trials; (b), with specularities, two trials. Note that MWR's choice varies little between the two images; the presence of a highlight in (b) appears not to influence his choice.

Chapter 8

Physiological Mechanisms Underlying Color Computation

Recent results in the neurophysiology of color vision make it possible to explore mechanisms of color computation at the level of individual cells and pathways. This thesis raises several questions: Are there "lightness neurons" – neurons with non-classical receptive fields that compute lightness in one chromatic channel – and if so, where? What are the spectral characteristics of the chromatic channels in which color is computed and where and how are the channels represented? Are there distinct "edge-finding" operators and "filling-in" operators, as suggested by both computational and psychophysical results, and if so, where are they and how do they interact? This chapter addresses aspects of these and other questions.

8.1 Physiological Data

8.1.1 Chrominance Cells

In the primate visual system, there are several distinct classes of cells selectively reponsive to the wavelength and/or color of stimulating light. These "chrominance" cells are likely to be involved at some level in the computation of color. Although in reality each class probably consists of a spectrum of types, it may be characterized by an ideal prototype. The real properties of cells in a class are difficult to itemize

because different researchers use different techniques to classify cells and because even cells classed together under one technique do not always satisfy the same criteria. The following list describes the properties of the prototypes in each class and discusses why some are more ideal than real. (All data is from Old World monkeys unless otherwise noted.)

- The ideal **single-opponent** (Type I) cell has spatially and chromatically opponent concentric fields, the center and surround each fed by a single cone type. The center of an L+/M− single-opponent cell gives an ON response to a spot of long-wavelength light (L+; i.e., increases its firing rate in response to an increment of long-wavelength light); its surround gives an OFF response to a larger spot of middle-wavelength light (M−; i.e., increases its firing rate to a decrement of middle-wavelength light). According to Livingstone and Hubel [66] about 80 % of the cells in the parvocellular layers of the lateral geniculate nucleus are Type I; 63 % of layer $4C\beta$ cells in V1 are Type I.

- **Center-only** (Type II) cells have a single circularly-symmetric chromatically opponent field. An L+M− center-only cell gives an ON response to a spot of long-wavelength light and an OFF response to a same-sized spot of middle-wavelength light.

- **Broad-band** (Type III) cells have spatially but not chromatically opponent concentric fields. An on-center broad-band (B-B) cell gives an ON response to a spot of light of any spectral composition in its center, and gives an OFF response to light of any spectral composition in its surround.

- **Double-opponent cells** have spatially and chromatically opponent concentric receptive fields, the center and surround each fed by two cone types in opponency. The center of an L+M−/L−M+ double-opponent (D-O) cell gives an ON response to a spot of long-wavelength light and an OFF response to a middle-wavelength light spot; its surround gives an ON response to an annu-

lus of middle-wavelength light and an OFF response to a long-wavelength light annulus.

D-O cells were first reported in goldfish retina [21], and have since been found in macaque monkey V1 [48, 29, 82], where they constitute more than 50 % of cells in metabolically active 'blobs' found in all layers but IVA and IVC [66], and in metabolically active thin stripes in V2 [47]. D-O cells have not yet been found in the retina or geniculate of primates.

Although D-O cells are abundant within blobs, they are not found outside of blobs in V1, and they contribute only a small percentage of the total cell population in visual cortex (e.g., about 10 % of cells in the foveal representation of layers II-III, about 4 % in the non-foveal representation [66]).

- A **complex double-opponent** cell gives a double opponent response to a small spot of light positioned anywhere within a much larger field (up to 16 times the optimum size of the spot.) Complex D-O cells constitute less than 5 % of cells in V2 and are clustered in regions that appear as thin dark stripes when stained with metabolic enzymes. Complex color-opponent and broad-band cells are also found in the thin stripes [67].

- Ts'o and Gilbert [105] find **modified type II** cells rather than double-opponent cells within blobs in V1. A modified type II cell has a type II cell center and a broad-band suppressive surround. T'so and Gilbert are unable to elicit single-opponent responses from annular stimuli presented to the receptive field surround of a modified type II cell, and therefore conclude that it is not a double-opponent cell. On the other hand, Livingstone and Hubel [66] do not explicitly elicit surround responses from cells they classify as double-opponent, but demonstrate a suppression of the center on and off responses when the central spot is expanded into the receptive field surround. Daw [21] did elicit single-opponent surround responses in double-opponent cells in the goldfish retina. On the basis of the Ts'o and Gilbert classification of modified type II cells, it

might be concluded that ideal double-opponent cells do not exist in primate visual cortex.

- **Oriented color-opponent** cells have oriented, wavelength-selective simple or complex receptive fields. A "red"-selective complex cell responds best to a bar of long-wavelength light oriented at its preferred angle, positioned against a white background in any of several excitatory regions within its relatively large receptive field [66].

  Livingstone and Hubel [66] report that outside of blobs in layers II and III of V1, 39 % of cells in the foveal representation and 17 % in the non-foveal representation are complex oriented wavelength-selective. Michael [83] reports that 13 % of wavelength-selective cells in V1 are simple oriented double-opponent cells, found almost exclusively in layer IV. Such a cell responds best to a bar of a preferred color oriented at a preferred angle on a background of the opponent color. It is not clear whether these cells are in fact the same as the oriented color-coded cells described in layers II and III and, if not, why they have not been described in other reports of color-coded cells.

8.1.2 The Chrominance Pathway

The psychophysical demonstrations of the role of edges in lightness and color perception suggest an interaction between two distinct mechanisms: one that finds luminance edges and uses them to bound regions and another that fills in color within regions. The evidence for the existence of anatomically and physiologically distinct pathways subserving form and color perception supports the idea of interacting mechanisms, but does not yet explain where or how the interaction takes place. DeYoe and Van Essen [28] summarize the evidence for form information being carried by the parvocellular geniculate–V1 blob–V2 thin stripe–V4 subregion pathway and for color information being carried by the parvocellular geniculate–V1 interblob–V2 interstripe–V4 subregion pathway. Livingstone and Hubel [67] argue that the magnocellular

geniculate also feeds into the blobs, perhaps providing luminance information as a third axis in a color space in which the other two axes are red-green and yellow-blue opponent axes. This is unlikely, since the relatively large receptive fields of magnocellular cells cannot provide the high spatial resolution characteristic of the luminance domain.

Data from De Valois et.al. [23] and from Logothetis et.al. [68], among others, indicate that a substantial portion of parvocellular cells are not strongly wavelength-selective. These may directly provide the luminance information necessary for form vision. That information may also be carried indirectly by color-opponent cells, as Young [114] has shown by applying prinicipal component analysis to the data obtained by De Valois et.al. [23], and as we discuss in section 8.2.1 below.

8.1.3 Extrastriate Color Cells

Computational analyses of the lightness problem suggest that lightness may be recovered from the image irradiance signal by filtering it through a receptive field with a small excitatory center and a very large inhibitory surround – the receptive field of a hypothetical "lightness neuron" (see section 8.2). This type of field has been termed "non-classical". Recent results suggest the presence of similar neurons in primate V4. Desimone, et. al [27] describe spectrally-selective neurons in V4 with a small excitatory center (1-2 degrees) and a large silent suppressive surround (10-12 degrees). The optimal stimulus for such cells is often a central spot of one wavelength surrounded by a large stimulus of a different wavelength.

Zeki [116] [117] describes neurons in V4 that appear to have the specific characteristics of lightness (or color) neurons. He defines two types: 'wavelength-selective color-only' cells which respond to a monochromatic light of a preferred wavelength, and to patches of the preferred color only within a Mondrian under a range of illuminants; and 'color-only' cells which do not respond to monochromatic light but do respond to patches of a preferred color within a Mondrian under a range of illuminants. Both cells are reported to have large receptive fields, but the extent of the

surround necessary to evoke (for color-only cells) or influence (for wavelength-selective color-only cells) the response has not been determined.

Schein (pers. comm.) has recently re-analyzed data that he gathered with Desimone in [27], concluding that there is strong evidence for lightness neurons in V4. A typical cell in V4 has one or two narrow peaks in the spectral sensitivity curve of its center (roughly 2 degrees), gives strong ON responses and weak OFF responses with the same spectral sensitivity, and has a strong suppressive surround with roughly the same spectral sensitivity of the center.

Further support for the role of V4 in color computation comes from Wild et. al. [110], who found that V4-ablated monkeys were unable to discriminate between two Mondrians under illuminants different from those under which they had been trained on the discrimination task. Normal (V4-intact) monkeys performed well on the task under a range of illuminants. On other behavioral tasks, including wavelength-, brightness- and pattern- discrimination tasks, the two groups of monkey performed at the same level. These results have the dual implication that V4 is crucial to color constancy and that color constancy is not crucial to all visual recognition tasks.

8.2 Computational Interpretations

8.2.1 The Puzzle of Single-Opponent Cells

Existing theories of color vision seem unable not only to predict but even to explain the puzzlingly high number of single opponent neurons in the visual pathway. From a computational viewpoint, such cells appear nonsensical at first sight because they mix chromatic and spatial information in a seemingly impenetrable way (as Marr observed in 1982 [74]). For analyzing luminance information, it would seem to make more sense to compare activities in cones of like type across space (e.g., L+ center/ L− surround), and for analyzing chrominance information, to compare activities in cones of opposite type at the same location (e.g., L+M− center).

In fact, the visual system may instead be forced by the physical and evolutionary

limitations of its design to create single-opponent cells (although, as discussed below, these are not the ideal single-opponent cells described above). Yet it still succeeds in meeting its computational goals. Single-opponent cells serve double duty admirably well.

One goal of the system is to collect and exploit color information: this requires photoreceptors of different spectral sensitivities in the retina. Another goal is to achieve high spatial resolution in the luminance domain, i.e. to maximize visual acuity. This requires closely packed photoreceptors in the fovea, and an ocular optics closely matched to their spacing. The resolving power of the human eye is limited by diffraction at the pupil and refraction in the anterior chamber, inducing an effective cutoff spatial frequency of around 60 cycles/degree (cpd). The maximum distance between receptors in a regular array that will allow reconstruction of a 60 cpd grating without aliasing is, according to the sampling theorem, approximately 30 seconds of arc. The human foveal cones are in fact arranged perfectly to meet these constraints. They lie in a triangular array (in which each cone is surrounded by six nearest neighbors) with a row spacing of 2.4-2.6 μm, which corresponds to 30-32 seconds of arc [111]. [1]

Several lines of evidence suggest that the dense lattice of receptors in the fovea in turn feeds a dense lattice of center-surround ganglion cells many of which have a single cone contributing to the center. Two-point visual acuity (approximately 1 minute of arc [109]) requires the existence of ganglion cells with a single cone contributing to the center [73], for example. The fact that gratings with spatial frequency at the Nyquist limit can be seen at low contrast when they are projected directly on the fovea using laser interference fringes, thereby bypassing the ocular optics, also implies that the spatial resolution of the ganglion cells must match that of the receptors [111]. Physiological data on the primate [22, 26, 35, 36], and anatomical data on the cat [11] also suggests that some ganglion cell centers are each fed by a single cone. Recently,

[1] As Williams [111] notes, the row spacing determines the Nyquist limit because "the lowest spatial frequency gratings that can produce aliasing ambiguity in a triangular array are those oriented in one of the three orientations lying parallel to rows of sampling elements."

Kaplan (pers. comm.) has found evidence that many single-opponent cells even at the level of the lateral geniculate nucleus have centers each fed by a single cone.

The surround of center-surround cells provides a filter through which the retinal image is smoothed. In order to minimize noise in the image, the output of this filter must be sampled as finely as possible. That is, if there are randomly-placed gaps in the surround, its output will be more variable than if there are no gaps. This implies that the surround of center-surround cells should be fed by all the nearest neighbors of the cone feeding the center. But this in turn implies that the ideal single-opponent cells described above cannot exist, at least not abundantly enough to supply complete coverage of the visual field. That is, if L and M cones are closely packed in a triangular array as the physiological, anatomical and psychophysical data suggest, then they cannot create enough L+/M− and M+/L− cells with center and surround each fed by pure opponent cone types, unless holes are allowed in the surround.

The numbers of such pure opponent cells are limited by the regularity of the cone mosaic and by the relative numbers of the two types of cones. Although the proportion of S cones in the primate retina has been determined by several methods, all of which converge to the same number, the relative numbers of L and M cones in the fovea are less well established. Smith and Pokorny [103] estimated an L:M ratio of 1.6:1 for the Judd Standard Observer by fitting the cone spectral sensitivities to the luminosity function V_λ. More recent estimates for the L:M ratio of the human fovea, based on the detection of spatially tiny flashes of light, are 1.46:1 – 2.36:1 [16] and 1.6:1 – 4.0:1 [107]. Marc and Sperling [72] conclude from anatomical data that in the baboon retina the ratio is reversed, with 40 x 10^3 M cones/mm^2 and 20x 10^3 L cones/mm^2 at eccentricity 0.5 degrees.

If there were exactly 2 L cones for each M cone and the two cone types were arranged in regular arrays, then the only allowed single-opponent cells with purely opponent center and surround would be the M+/L− or M−/L+ types (see Figure 8.1(a)). A cell with a single L cone feeding its center by necessity would have L and M cones intermixed in its surround.

Recently, Packer et. al. [88] have obtained results which imply that the L and M cones do not separately form regular submosaics, but instead are arranged in irregular arrays that interleave to form a regular mosaic of both cone types. By selectively stimulating the L and M cones in the fovea, Packer et. al. have demonstrated that the cutoff frequency above which aliasing is obtained is the same for the L and M cone submosaics as for the L+M mosaic, although the image noise increases when the submosaics alone are stimulated. If the L and M cone submosaics are random as these results imply, then neither L center/M surround nor M center/L surround single-opponent types may be pure. Both will have a mixture of L and M cones in their surrounds, with L cones generally more heavily weighted (see Figure 8.1(b)). Lennie et. al. [64] simulated single-opponent cells constructed from retinal mosaics in which the distributions of the L and M cones and their contribution weights to center and surround were systematically varied. They found that cells whose centers are fed by single cones but whose surrounds draw their inputs from a random mixture of cone types best fit the data obtained by Derrington et. al. [25] on the range of chromatic opponency in primate lateral geniculate nucleus.

We are now confronted with a paradox. Long- (L) and middle-wavelength-selective (M) cones are intermixed in the fovea: the system requires them to be, in order to analyze spectral information. But as a result, single-opponent cells mix spectral and spatial information, since the center must be spectrally different from the surround. In order to achieve visual acuity, avoid aliasing and at the same time gather color information, the system must pay by creating an imperfect single-opponent cell! How does the system compensate?

One attractive hypothesis is that this is the reason for the strong overlap in the sensitivity spectrum of long- and middle-wavelength-selective cones (see Figure 1.1), and possibly for the near absence in the fovea of blue cones (which have a quite different spectral sensitivity). Thus, for visual acuity in the luminance domain, the spectral differences between the center and surround of single-opponent cells may effectively be negligible. (The difference in sensitivity of the two cone types to light

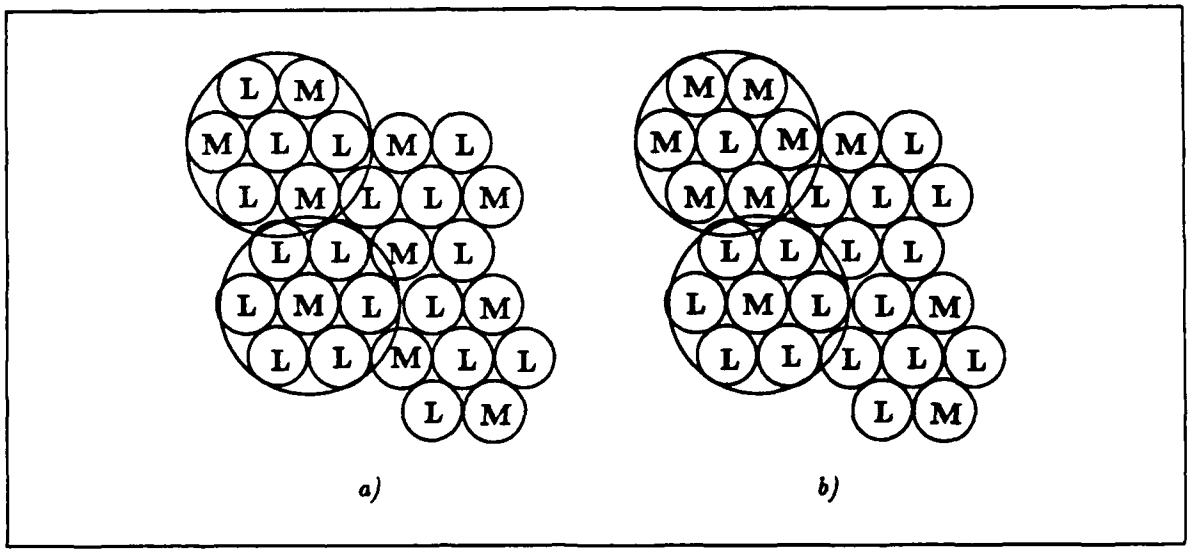

Figure 8.1: (a) L and M cones in a 2:1 ratio, arranged in regular submosaics so that each M cone is surrounded by L cones only. All L cones have a mixture of cone types in their surrounds. The circled regions illustrate typical center-surround cells thus allowed. (b) L and M cones in a 2:1 ratio, but arranged in more realistic random submosaics. Now both L and M cones may form the centers of pure cone-opponent cells (circled regions), but not everywhere in the fovea.

of the same wavelength could be compensated for by a change in gain of the center or surround response.)

Recent results from Nathans et.al. [87] put these comments into an evolutionary perspective. The close homology between genes coding for the L and M cone types suggest that the two arose as mutations of a single gene coding for a single cone type. This suggests that our ancestors had only one cone type in the fovea (with perhaps a second type more sparsely distributed)[2], closely packed for high acuity. When the advantage of color vision became obvious, evolution selected those mutations which retained the structure of the retina but simply, and only slightly, changed the pigments of some of its receptors.

Ingling and Martinez [53] have shown that for ideal single-opponent cells, it is not necessary to invoke similar spectral sensitivities of the cone types in order to extract a luminance signal from them. They illustrate that at high spatial frequencies, single-opponent cells effectively become luminance change detectors, by summing cone inputs. At low spatial frequencies, they are chrominance change detectors (see Figure 8.2).

Barlow has suggested (pers. comm.) that single-opponent cells might simply be one stage in the synthesis of double-opponent cells. (In the sense that they are only one of a few cell types constructed at the retinal ganglion level, they must be building blocks for cells at further stages in the visual pathway.) His scheme is sketched in Figure 8.3(a): single-opponent cells of like opponency are summed from one region of the visual field to create the center of a double-opponent cell; the responses of similar single-opponent cells are inverted before converging to create the surround. This is equivalent to disentangling the spatial and chromatic information that is intertwined in their output, but the chromatic information can be retrieved only at

[2]Gouras and Zrenner [37] suggest that the short-wavelength (S) cone might have evolved from the rod, to which it is electrophysiologically similar, in order to provide color contrast against "an existing midspectral cone mechanism," which itself split into two mechanisms later in evolution to create three axes for color vision. This implies that early in evolution rods and cones were intermixed in the fovea and that early color vision, mediated by the single midspectral cone and the S cone, had even coarser spatial resolution than trichromatic color vision.

the expense of spatial resolution (see Figure 8.3(b)). It suggests that the center of a double opponent cell should be at least the same size as the entire receptive field of a corresponding single opponent cell. In principle, of course, double opponent cells with similar low resolution receptive fields could be built directly from the receptors without the intermediate step of small single opponent cells.

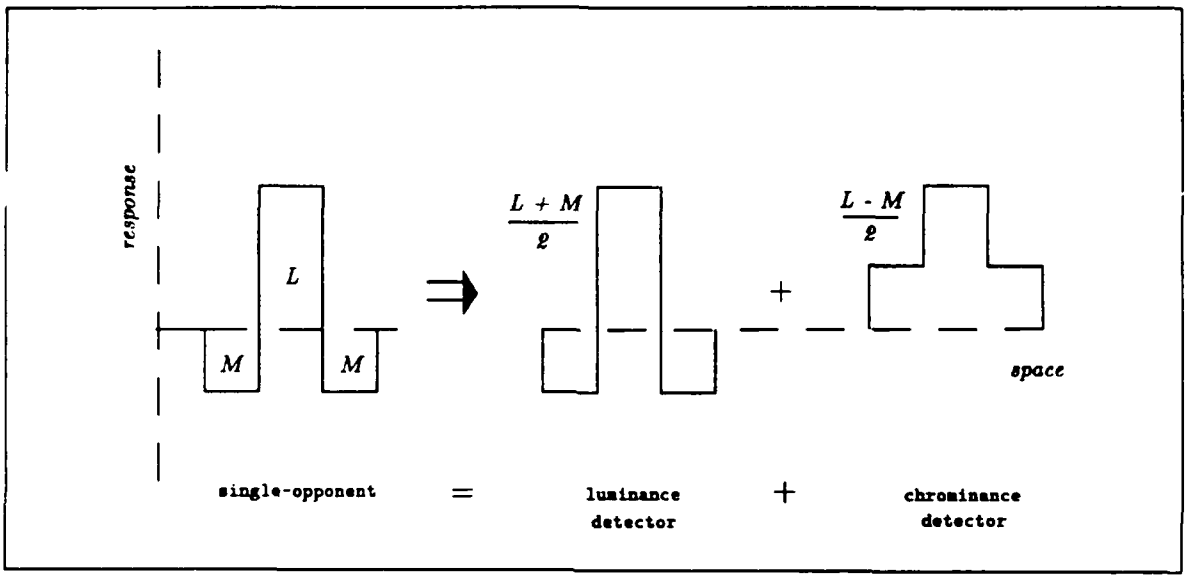

Figure 8.2: The response of an ideal single-opponent cell in one spatial dimension can be decomposed into the responses of a luminance change detector at high spatial frequencies and a chrominance change detector at low spatial frequencies. Redrawn from [53].

On the basis of these arguments, one may speculate that high visual acuity is incompatible with widely different cone spectral sensitivities. Thus, animals that put a premium on color vision at the expense of acuity might have receptor sensitivities more broadly separated than those who value acuity for survival (e.g. primates and birds). The roach, a cyprinid fish, has four photoreceptor types with well-spaced spectral sensitivities [9]. One might conclude that, to find its food, the fish does not require high acuity spatial vision.

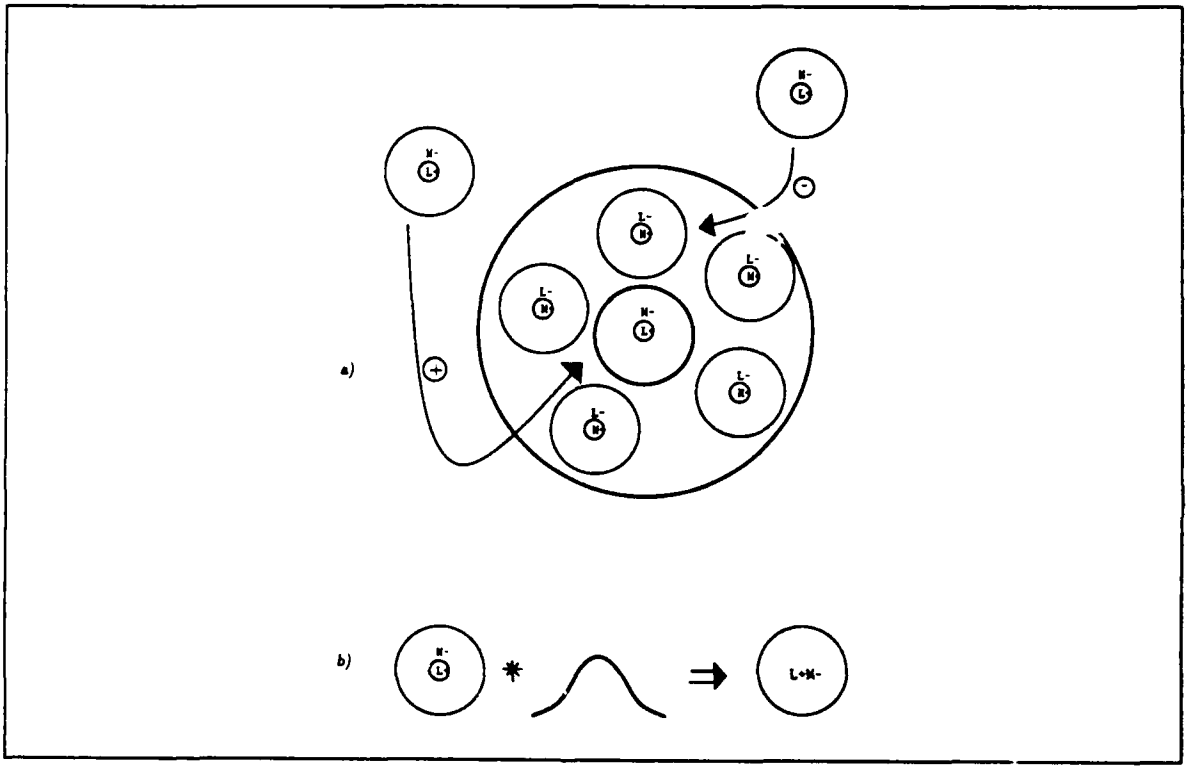

Figure 8.3: (a) Single-opponent cells converge to create a double opponent cell. (b) A center surround receptive field – the sum of two gaussian distributions - is convolved with a large gaussian: the outcome is effectively an opponent center only receptive field.

8.2.2 Double-Opponent Cells as Segmentation Operators

At first sight, the duty of double-opponent (D-O) cells seems to be to signal changes in a hue index like that defined in Equation 6.3 (h_{ij}). But the question of how the cone opponency is implemented in the D-O cell center and surround is critical in determining what precisely the D-O cell does signal.

The receptive fields of D-O cells have typically been described as center-surround with opposing cone-opponent inputs to the center and surround, e.g. $L+M-$ in the center and $M+L-$ in the surround. If D-O cells indeed exist in this ideal form questions still remain on the nature of the opponency within and between the center and surround. Are cone inputs combined linearly? If so, then if the cone responses themselves are the logarithms of image irradiance, the D-O receptive field may be described as

$$(\log I^L - \log I^M)_{center} // (\log I^M - \log I^L)_{surround}$$

or as

$$\log(\frac{I^L}{I^M})_{center} // \log(\frac{I^M}{I^L})_{surround}$$

where $I^L = k^L E(\mathbf{r})\rho^L(\mathbf{r})$ and L and M label the long- and middle-wavelength-sensitive photoreceptors respectively. If, further, the center and surround responses are combined linearly then the operation the D-O cell performs may be written as

$$G * \nabla^2 \log(\frac{I^L}{I^M}),$$

i.e. as the gaussian (G) of a spatial derivative operator on the log of the hue index h_{ij}.

As the preceding string of "ifs" suggests, there are other interpretations of the D-O cell receptive field. If the interaction between the center and surround is division rather than subtraction then the D-O cell may register

$$\frac{(\log \frac{I^L}{I^M})_{center}}{(\log \frac{I^M}{I^L})_{surround}},$$

a quantity that fits less easily into a segmentation scheme. In fact, recent reports from Kaplan (pers. comm.) suggest that the interaction between center and surround in single cone-opponent (Type I) cells is of the shunting inhibition type. This suggests that Type I cells, usually described as $L+_{center}//M-_{surround}$ are in fact registering $\frac{L_{center}}{M_{surround}}$, which, if $L = \log I^L$ as above, becomes

$$\frac{\log I^L_{center}}{\log I^M_{surround}}.$$

If on the other hand, the cone responses may be considered linear in small operating ranges, Type I cell responses become

$$\frac{I^L_{center}}{I^M_{surround}}.$$

Since Type I cells might be primary building blocks for D-O cells (one Type I cell per D-O center), this would suggest that the D-O center, instead of being $\log(\frac{I^L}{I^M})$, is either $\frac{\log I^L}{\log I^M}$ or $\frac{I^L}{I^M}$.

In the absence of more restrictive data, the possibilities are numerous!

8.2.3 Lightness Neurons

From the computational point of view, it is unlikely that the physiological visual system uses a retinex-type lightness algorithm to compute color in a three-dimensional world in which shape, shading and shadows confound the intensity signal. Yet it is possible that the visual system performs a similar computation using similar operations. In what follows, we speculate on the sorts of basic physiological operations that lightness algorithms would require for their implementation.

The two basic operations in lightness algorithms – *spatial differentiation* and *integration* – imply two basic types of operators that the human visual system should possess if it does implement a lightness algorithm. In the implementation of the

lightness algorithm the two steps acquire crucially different characters: differentiation is a *local* process that mediates spatial decomposition of the intensity signal by separating illumination gradients from reflectance changes, and integration is a *global* process that mediates spectral normalization by averaging the reflectance over a large portion of the visual field. The local operator must therefore take the difference in light intensity between nearby parts of the image, whereas the global operator must sum light intensity from virtually all parts of the image.

Spatial differentiation

The local operator may in turn be of two types: the directional gradient operator which responds best when the direction of maximum intensity change coincides with its preferred direction (Land, Crick, Blake; see Chapter 4) and a nondirectional Laplacian operator with a circularly symmetric receptive field (Horn). The formal arguments in Chapter 4 show that each local operator is effectively weighted by its distance from the point at which lightness is evaluated.

Equation 4.20 suggests that the gradient and Laplacian operator may coexist and work in conjunction, the sum of their responses over different parts of the visual image yielding lightness. The contribution from the Laplacian operator at point (ζ, η) to the lightness at point (x, y) is the same in any direction at a given radius r, because the Laplacian is isotropic. Thus, the Laplacian operator simply sends out the same signal along its connections in all directions. The gradient operator, on the other hand, is directional and therefore its contribution from point (ζ, η) must be weighted appropriately according to its orientation with respect to the point (x, y). In the physiological implementation of the second term of Equation 4.20, for example, the connections between the operator neurons and the lightness neurons must be highly specific, so that the operator neuron sends out either a single signal in a single direction, or several signals, different in each direction.

The contribution from the Laplacian operator is weighted by a factor of $\ln r$, so it would seem at first glance that points further away from (x, y) contribute more to

the lightness there than closer points. In fact, $\ln r$ increases slowly enough so that at the high end of the range of known lengths of cell-to-cell connections (where the longest is $\simeq 10$ times that of the smallest), $\ln\ r$ is practically constant.

As Figure 8.4 illustrates, the Laplacian of Equation 4.19 (convolved with a Gaussian, see later), operating at an edge, gives a strong positive result on one side and an equally strong negative result on the other. The sum of the two *weighted* signals is on the order of $A(\frac{\Delta r}{r_0})$, where A is the amplitude of the signal on either side, r_0 is the distance to the nearest side, and Δr is the additional distance to the far side. Thus, for large r, the signals sent by Laplacians on either side of an edge tend to cancel each other at (x, y). That is, if the sum is taken locally and then sent to (x, y), it will tend to attenuate over the long distance, and, similarly, if the two weighted signals are sent independently and then summed, fluctuations over the distance will tend to equalize them. Nearby edges, for which Δr is a significant fraction of r, will therefore contribute most to the lightness.

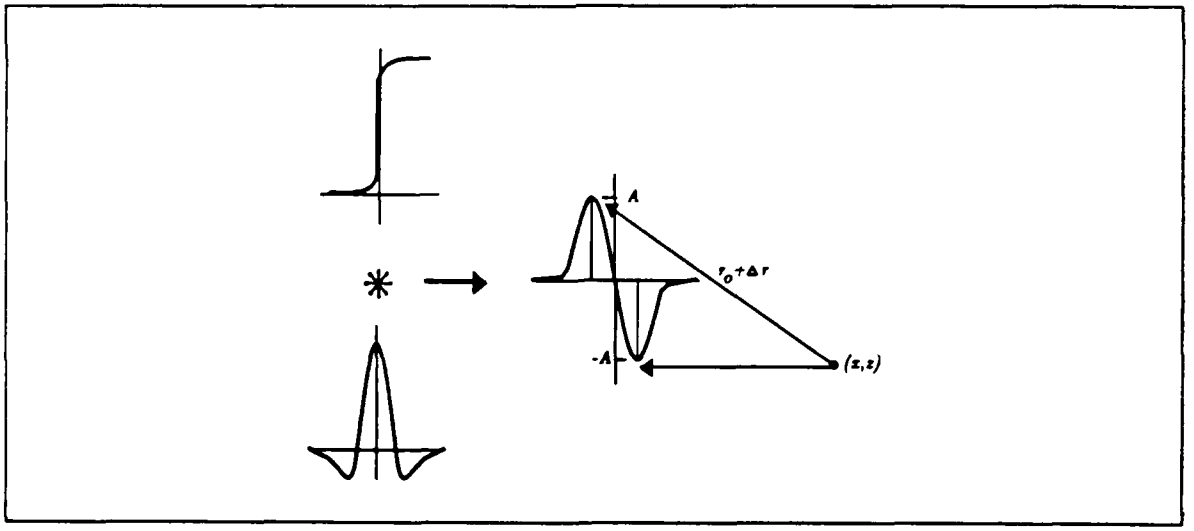

Figure 8.4: An intensity edge convolved with a DOG filter, yielding a strong positive result on one side of the edge and a strong negative result on the other.

The gradient operator, in contrast to the Laplacian, takes the difference in intensity signals across an edge, and thus sends a large, robust signal to (x, y). But

this signal is weighted by $\frac{1}{r}$, which falls off rapidly for points far away from (x, y). Beyond a critical distance r, additional contributions from the gradient operator will therefore not change significantly the lightness at (x, y), so, again, nearby edges will contribute most.

Both local operators thus imply an upper limit on the necessary length of cell-to-cell connections. Interestingly, weighting of nearby edges may offer a partial explanation of the phenomenon of simultaneous contrast, which has elsewhere been used as an argument against Land's original retinex algorithm.

The recent discovery of double opponent cells in primate V1, V2 and V4 points to the Laplacian operator as the most likely candidate in the physiological computation of color [66]. (Interestingly, Land's most recent retinex operator is a center-surround filter.) The double opponent cell, as do other center-surround cells, performs not the Laplacian in the formal statement of Horn's algorithm (Equation 4.19) but the difference-of-Gaussians (DOG) operation, equivalent to blurring the intensity signal through a Gaussian filter before applying the Laplacian. In this respect, double opponent cells perform exactly the same operation as in Horn's iterative implementation and in the multiple scales algorithm, as Equations (4.9) and (4.25) make clear. In fact, the application of the Gaussian serves an essential computational purpose in eliminating noise from the intensity signal and thereby stabilizing the implementation.

Double-opponent cells differ in another way from the simplest operator that Horn's algorithm would predict. The latter would perform the Laplacian on the intensity signal in one photoreceptor channel. Double-opponent cells have a center-surround receptive field organization with opposite pairs of opponent cone inputs to the center and surround (e.g., L+M− in the center, and L+M− in the surround) and thus perform spatial differentiation on a new channel formed by linear combination of the photoreceptor channels. Because a threshold operation is applied to the result of the Laplacian, the D-O cell operation is not equivalent to performing the Laplacian on the intensity signal in the photoreceptor channel and then taking linear combinations.

That is,

$$(T\nabla^2[S^i - S^j]) \neq (T\nabla^2 S^i) - (T\nabla^2 S^j)$$

Thus, the axes on which the lightness triplets are plotted in the original retinex algorithms (e.g., long- , middle- and short- wave axes) cannot simply be rotated into the physiological color-opponent axes.

At least one of the types of oriented wavelength-selective cells described above could act as a gradient operator.

The above arguments suggest that if the two types of operators coexist, they may be localized in different areas corresponding to different regions of the visual field. Because the effective weight of each operator falls off with distance, neither would be expected to contribute much from the periphery to the computed lightness in the central visual field. Thus, both may be expected to concentrate in the representations of the foveal and parafoveal regions. But because the gradient operator is more sensitive to the difference in intensity signals across an edge, it may also be found in the peripheral field representation, where it may correct for offsets in the surround reflectance.

Spatial integration

The different lightness algorithms predict long-range connections of different types. Horn's iterative method for integrating the Laplacian could be implemented, for example, by an array of interconnected center-surround cells, the outputs of which feed back into and are propagated across the array until a stable solution is reached, without explicit long-range connections. The sum over the responses of gradient operators, in contrast, would be difficult to achieve without explicit long-range connecting axons or dendrites.

The multiple scales method for integrating the Laplacian suggests a third way to implement the global process of spectral normalization. Rather than perform

an iteration over time, as in Horn's method, in the multiple scales method center-surround cells may perform an integration over space, again without explicit long-range connections. The sum is taken over a collection of center-surround cells of different sizes covering the same central point in the visual field. If the multiple scales are collected in a single anatomical area, cells with larger receptive fields would necessarily be built up by cells with smaller receptive fields from stages lower in the visual pathway. Alternatively, multiple scales may exist at different anatomical levels in the visual pathway; each successive scale could be built up by filtering the one at the previous stage through a Gaussian. The final-stage cell would be expected to have a relatively narrow center, and a large, shallow surround (a "non-classical" receptive field; see Allman et. al. [2]).

There is evidence of long-range connections in several cortical areas. In V1, the longest connections appear to be from blob to blob in layers II and III [66], and the typical length of 2 mm in layers II and III suggests that a single blob projects only to its nearest neighbor blobs. [3] In V2, intrinsic connections extend 2-3 mm, and in V4, 3.5-4 mm [97].

Each blob in the foveal representation covers only a piece of the cortical area devoted to a single point in the visual field. The absence of explicit connections between blobs more distant from each other leaves open the possibility that the pattern of blobs and interblob connections forms an interconnected array of cells performing a Horn-type iteration.

Alternatively, blobs may implement a multiple scales algorithm, by deliberately clustering cells with a range of receptive field sizes centered on the same point. The present lack of data on the variation of receptive field size within a blob makes it difficult to judge whether such a sum is possible. The connections between blobs could in this scheme serve other functions, such as to compare responses between cells of the same receptive field size across the field in order to compute an appropriate

[3] Rockland and Lund [97] also report that connections in V1 attain a length of 3 mm in the Stria of Gennari, but it is not known whether these are also interblob connections.

threshold.

Appendix A

Integrating Image Irradiance in Sensor Channels

This appendix illustrates how, by transforming the photoreceptor activities, to decompose the image irradiance equation into a sum of two components representing the surface reflectance and the effective surface irradiance.

In most natural scenes, the light reflected from objects includes both diffuse and specular components. To simplify the image irradiance equation, we assume that all reflection is Lambertian, or that there are no specularities. In this case, the light reflected by a surface is solely the product of its surface spectral reflectance (i.e. body reflectance) and the irradiance it receives. The amount of reflected light that reaches the eye further depends on the angles between the illumination source, the reflecting surface, and the eye. The response of the eye to the light depends on the spectral sensitivity of its photoreceptors. We may therefore write the light signal registered by the eye as:

$$S^{\nu}(x) = \log \int a^{\nu}(\lambda)\rho(\lambda, x)E(\lambda, x)d\lambda, \qquad (A.1)$$

where ν labels the spectral type of the photoreceptor ($\nu = 1, ...4$ for humans, counting 3 cone types and 1 rod type), $a^{\nu}(\lambda)$ is the spectral sensitivity of the νth-type photoreceptor and $S^{\nu}(x)$ its activity at image location x. $\rho(\lambda, x)$ is the surface spectral reflectance and $E(\lambda, x)$ the *effective irradiance*. We group together the geometric

factors influencing the intensity signal by defining the effective irradiance as the illumination modified by the orientation, shape, and location of the reflecting surface. The effective irradiance is the intensity that the surface would reflect if it were white, that is, if $\rho(\lambda, x) = 1$.

Equation A.1 is not separable into a product of reflectance and illumination components in a single color channel. To make it separable, we may make a transformation to new color channels that are combinations of the photoreceptor activities. We first choose basis functions $p^i(\lambda)$ and $q^i(\lambda)$ such that for most naturally occurring illuminants and surface reflectances

$$E(\lambda, x) \approx \sum_i p^i(\lambda) e^i(x)$$

$$\rho(\lambda, x) \approx \sum_i q^i(\lambda) r^i(x). \qquad (A.2)$$

The basis transformation (for a review of the origins of this idea see Maloney [70]; see also Buchsbaum [15] and Yuille [115]) leads to the following equation

$$S^\nu(x) = T^\nu_{ij} e^i(x) r^j(x), \qquad (A.3),$$

where the tensor T is defined as

$$T^\nu_{ij} = \int d\lambda a^\nu(\lambda) p^i(\lambda) q^j(\lambda),$$

where $\nu = 1, ...4$ and $i, j = 1, ...N$, the p's and the q's are the basis functions for the illuminant and for the albedo, respectively, and the sum is taken over repeated indices.

To simplify further analysis, we impose the conditions that the $p^i = q^i$ and that the $p^i(\lambda)$ are orthogonal with respect to the $a^\nu(\lambda)$ under integration over λ. That is, $T^\nu_{ij} = \int a^\nu(\lambda) \delta^{ij}(\lambda) d\lambda$. This orthogonality is insured if, for example, the $p^i(\lambda)$ do not overlap with respect to λ. In the simplest but least physiological case, the basis functions may be monochromatic: $p^i(\lambda) = \delta(\lambda - \lambda_i)$.

Substituting A.2 into A.1 then yields:

$$S^{\nu}(x) = \log \sum_{i} e^{i}(x) r^{i}(x) \int d\lambda a^{\nu}(\lambda) p^{i}(\lambda) p^{i}(\lambda). \qquad (A.4)$$

If we define the matrix $T_i^{\nu} = \int d\lambda a^{\nu}(\lambda) p^{i}(\lambda) p^{i}(\lambda)$, where $\nu = 1, ...4$ and $i = 1, ...N$, we obtain

$$\exp\left[S^{\nu}(x)\right] = \sum_{i} T_{\nu i} e^{i}(x) r^{i}(x). \qquad (A.5)$$

If the $p^{i}(\lambda)$ are suitably chosen, $T_{\nu i}$ is invertible. Then the linear equations represented in A.5 yield the following solution:

$$(T_{\nu i})^{-1} \exp\left[S^{\nu}(x)\right] = e^{i}(x) r^{i}(x) \qquad (A.6)$$

or,

$$s^{i}(x) = e^{i}(x) r^{i}(x) \qquad (A.7)$$

where $s^{i}(x) = (T_{\nu i})^{-1} \exp\left[S^{\nu}(x)\right]$. Taking logarithms of A.7 yields

$$\log\ s^{i}(x) = \log\ e^{i}(x) + \log\ r^{i}(x)$$

or

$$\tilde{s}^{i}(x) = \tilde{e}^{i}(x) + \tilde{r}^{i}(x) \qquad i = 1, \ldots, N, \qquad (A.8)$$

where $\tilde{s}^{i}(x) = \log\ s^{i}(x)$ and so on, which is the desired equation. The extension to the two-dimensional case is clear.

$\tilde{s}^{i}(x)$ is a linear combination of the activity of different types of photoreceptors. It is important to note that the index i labels not the color channels associated with the spectral sensitivities of the different photoreceptor types, but new channels which may be similar to the biological color-opponent channels. There is no ***a priori*** limit on the number of new channels formed by linear combinations, but efficiency of information transmission would require it to be close to the number of photoreceptor

types. Furthermore, although we could use the pseudoinverse of T rather than the inverse, a necessary condition for the latter to exist is that there are exactly as many new channels as photoreceptor types.

Appendix B

Channels for Spectral Normalization

Each narrow-band light in McCann et. al.'s [79] illuminant may be approximated by a monochromatic light: $E(\lambda_S)$ at 450 nm stimulates the short-wavelength-sensitive (S) cone only; $E(\lambda_M)$ at 530 nm stimulates the L and M cones only; and $E(\lambda_L)$ at 630 nm stimulates the L cone only. Thus the illuminant for one viewing may be written:

$$E^1(\lambda) = E^1(\lambda_S) + E^1(\lambda_M) + E^1(\lambda_L)$$

The photoreceptor response may be written as a function of the integral:

$$\int a^\nu(\lambda)E^1(\lambda)R(\lambda)$$

where $R(\lambda)$ is the surface reflectance of a Mondrian patch. The integral becomes a sum:

$$a^\nu(\lambda_S)E^1(\lambda_S)R(\lambda_S) + a^\nu(\lambda_M)E^1(\lambda_M)R(\lambda_M) + a^\nu(\lambda_L)E^1(\lambda_L)R(\lambda_L)$$

which we may write as:

$$\mathbf{a}^\nu \cdot \mathbf{e}^1 \cdot \mathbf{r}$$

where $\mathbf{r} = (r_S, r_M, r_L) = [R(\lambda_S), R(\lambda_M), R(\lambda_L)]$, $\mathbf{e} = (e_S, e_M, e_L)$ and so on. Von Kries adaptation says that for color constancy to hold under a change in illuminant from $\mathbf{e}^1$ to $\mathbf{e}^2$:

$$\frac{\mathbf{a}^\nu \cdot \mathbf{e}^1 \cdot \mathbf{r}}{k_1^\nu} = \frac{\mathbf{a}^\nu \cdot \mathbf{e}^2 \cdot \mathbf{r}}{k_2^\nu}$$

where each side of the equation represents the lightness in the νth channel for a different viewing.

McCann et. al. [79] choose $k_1^\nu = \int a^\nu(\lambda)E^1(\lambda)R_W(\lambda)$ to compute lightness and compare changes in the computed value with changes in perception. $R_W(\lambda)$ is the reflectance of a standard white paper, i.e. $R_W(\lambda_S) = R_W(\lambda_M) = R_W(\lambda_L) = r_W$. Thus $k_1^\nu = r_W(\mathbf{a}^\nu \cdot \mathbf{e})$. If lightness is indeed computed in the cone channels, then it is easy to see that multiplicative changes in $E(\lambda_S)$ alone will not affect the computed lightnesses at all. For example, the lightness in the S cone channel has no contributions from ϵ_M or ϵ_L:

$$\frac{a^S e_S^1 r_S}{r_W \cdot a^S e_S^1} = \frac{a^S e_S^2 r_S}{r_W \cdot a^S e_S^2} = \frac{r_S}{r_W}.$$

Since there is no change in illumination seen in the L and M channels, lightness does not change there either. Changes in the middle- or long-wavelength lights alone do not change the computed lightnesses in the M or S channels, but will in general change the lightness in the L channel, which is stimulated by both the changed and unchanged lights. This is in fact the pattern of changes in observers' color perception reported by McCann et. al., supporting the idea that normalization is performed on the original cone channels.

If normalization were done on the opponent-color channels instead, for example, we may expect a different pattern of color changes. For the blue-yellow (deuteranopic) channel, $a^D(\lambda) = a^M(\lambda) + a^L(\lambda) - a^S(\lambda)$. Under changes in the short-wavelength light alone, the computed lightnesses are generally not equal:

$$\frac{a^L e_L^1 r_L + a^M e_M^1 r_M + a^L e_M^1 r_M - a^S e_S^1 r_S}{r_W(a^L e_L^1 + a^M e_M^1 + a^L e_M^1 - a^S e_S^1)} \neq \frac{a^L e_L^1 r_L + a^M e_M^1 r_M + a^L e_M^1 r_M - a^S e_S^2 r_S}{r_W(a^L e_L^1 + a^M e_M^1 + a^L e_M^1 - a^S e_S^2)}$$

In general, changes in e_S will not be compensated by changes e_L and e_M. Changes in the middle-wavelength light alone will also tend to affect the blue-yellow channel

more than the red-green (L − M) because both the L and M channels see the change, whereas the S channel does not.

Appendix C

"Learning" Techniques Related to Optimal Linear Estimation

C.1 Recursive Estimation of L

It is of particular import for practical applications that the pseudoinverse can be computed in an adaptive way by updating it when new data become available[1]. Consider again Equation 5.5. Assume that the matrix Y consists of $n-1$ input vectors and Z of the corresponding correct outputs. We rewrite Equation 5.5 as

$$L_{n-1} = Z_{n-1}Y_{n-1}^{+}. \tag{C.1}$$

If another input-output pair $\mathbf{y}_n$ and $\mathbf{z}_n$ becomes available, we can compute L_n recursively using the formula

$$L_n = L_{n-1} + (Z_n - L_{n-1}\mathbf{y}_n)t_n^T, \tag{C.2}$$

where

$$t_n^T = \frac{\mathbf{y}_n^T(Y_{n-1}Y_{n-1}^T)^{-1}}{1 + \mathbf{y}_n^T(Y_{n-1}Y_{n-1}^T)^{-1}\mathbf{y}_n}, \tag{C.3}$$

provided that $(Y_{n-1}Y_{n-1}^T)^{-1}$ exists, for which it is necessary that the number of columns in Y is greater than or equal to the dimension of $\mathbf{y}$. The case in which

$(Y_{n-1}Y_{n-1}^T)^{-1}$ does not exist is discussed together with more general results in Albert [1]. Note that $(\mathbf{z}_n - L_{n-1}\mathbf{y}_n)$ in the updating equation (C.2) is the error between the desired output and the predicted one, in terms of the current L. The coefficient t_n is the weight of the correction: with the value given by Equation C.3 the correction is optimal and cannot be improved by any iteration without new data. A different value of the coefficient is suboptimal but may be used to converge to the optimal solution by successive iterations of Equation C.3 using the same data.

C.2 Connections to Other Statistical Techniques

The optimal linear estimation scheme described in Chapter 5 is closely related to a special case of Bayesian estimation in which the best linear unbiased estimator (BLUE) is found. Consider Equation 5.3: the problem is to construct an operator L that provides the best estimation $L\mathbf{y}$ of $\mathbf{z}$. We assume that the vectors $\mathbf{y}$ and $\mathbf{z}$ are sample sequences of Gaussian stochastic processes with, for simplicity, zero mean. Under these conditions the processes are fully specified by their correlation functions

$$E[yy^T] = C_{yy}, \quad E[zy^T] = C_{zy} \tag{C.4}$$

where E indicates the expected value. The BLUE of $\mathbf{z}$ (see [1]) is, given $\mathbf{y}$,

$$\mathbf{z}^{est} = C_{zy}C_{yy}^{-1}\mathbf{y}, \tag{C.5}$$

which can be compared directly with the regression equation

$$L\mathbf{y} = ZY^T(YY^T)^{-1}\mathbf{y}. \tag{C.6}$$

The quantities ZY^T and YY^T are approximations to C_{zy} and C_{yy}, respectively, since the quantities are estimated over a finite number of observations (the training examples). Thus there is a direct relation between BLUEs and optimal linear estimation. The learned operator captures the stochastic regularities of the input and output signals. Note that if the input vectors $\mathbf{y}$ are orthonormal, then $L = ZY^T$ and

the problem reduces to constructing a simple correlation memory of the holographic type (see [92, 91]). Under no restrictions on the vectors $\mathbf{y}$, the correlation matrix ZY^T may still be considered as a low-order approximation to the optimal operator (see Kohonen [55]).

If the operator L is space invariant, Equation C.5 is equivalent to the formulation of an optimal linear filter also called the *matched filter*. Equation C.5 becomes

$$Z^{est} = H^{opt} \cdot Y, \tag{C.7}$$

with

$$H^{opt} = \frac{\Phi_{zy}}{\Phi_{yy}},$$

where $\Phi_{zy}$is the ensemble cross power spectrum of output and inputs and Φ_{yy} is the power spectrum of the inputs. Y and Z are the Fourier transforms of the inputs and estimated outputs, respectively, and H is the transfer function of the optimal filter. Both Φ_{zy} and Φ_{yy} can be estimated from the set of $\mathbf{z}$ and $\mathbf{y}$ example pairs.

Bibliography

[1] Arthur Albert. *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, New York, 1972.

[2] John Allman, Francis Miezin, and EveLynn McGuinness. Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area MT. *Perception*, 14:105–126, 1985.

[3] Lawrence E. Arend and Robert Goldstein. Lightness models, gradient illusions, and curl. *Perception and Psychophysics*, 42(1):65–80, 1987.

[4] Lawrence E. Arend and Adam Reeves. Simultaneous color constancy. *Journal of the Optical Society of America*, 3(10):1743–1751, 1986.

[5] Denis A. Baylor. Photoreceptor signals and vision. *Investigative Ophthalmology and Visual Science*, 28(1):34–49, 1987.

[6] K. T. Blackwell and G. Buchsbaum. Quantitative studies of color constancy. *Journal of the Optical Society of America*, 5(10):1772–1780, 1988.

[7] Andrew Blake. On lightness computation in Mondrian world. In T. Ottoson and S. Zeki, editors, *Central and peripheral mechanisms of colour vision*, pages 45–49. Macmillan, 1985.

[8] Andrew Blake and Andrew Zisserman. *Visual Reconstruction.* MIT Press, Cambridge, Mass, 1987.

[9] J.K. Bowmaker. Trichromatic color vision: why only three receptor channels? *Trends in Neurosciences*, 6(2), 1983.

[10] J.K. Bowmaker and H. J. A. Dartnall. Visual pigments of rods and cones in a human retina. *Journal of Physiology*, 298, 1980.

[11] B. B. Boycott and H. Wässle. The morphological types of ganglion cells of the domestic cat's retina. *Journal of Physiology*, 240:397–419, 1974.

[12] D. H. Brainard and B. A. Wandell. Analysis of the retinex theory of color vision. *Journal of the Optical Society of America*, 3:1651–1661, 1986.

[13] M. H. Brill. Further features of the illuminant-invariant trichromatic photosensor. *J. Theoretical Biology*, 78:305–308, 1979.

[14] Philippe Brou, Thomas R. Sciascia, Lynette Linden, and Jerome Y. Lettvin. The colors of things. *Scientific American*, 255(3):84–91, 1986.

[15] G. Buchsbaum. A spatial processor model for object color perception. *J. Franklin Institute*, 310:1–26, 1980.

[16] C.M. Cicerone and J. L. Nerger. Relative numbers of l and m cones in human fovea centralis. *Vision Research*, 29(1):115–128, 1989.

[17] Hewitt D. Crane and Thomas P. Piantanida. On seeing reddish green and yellowish blue. *Science*, 221:1078–1080, 1983.

[18] O. Creutzfeldt, B. Lange-Malecki, and K. Wortmann. Darkness induction, retinex and cooperative mechanisms in vision. *Experimental Brain Research*, 67:270–283, 1987.

[19] M. Critchley. Acquired anomalies of colour perception of central origin. *Brain*, 88:711–724, 1965.

[20] A. Damasio, T. Yamada, H. Damsio, J. Corbett, and J. McKee. Central achromatopsia: Behavioral, anatomic, and physiologic aspects. *Neurology*, 30:1064–1071, October 1980.

[21] N. W. Daw. Colour-coded ganglion cells in the goldfish retina: Extension of their receptive fields by means of new stimuli. *Journal of Physiology*, 197:567–592, 1968.

[22] F. M. De Monasterio and P. Gouras. Functional properties of ganglion cells of the rhesus monkey retina. *Journal of Physiology*, 251:167–195, 1975.

[23] R. L. De Valois, I. Abramov, and G. H. Jacobs. Analysis of response patterns of LGN cells. *Journal of the Optical Society of America*, 56:966–977, 1966.

[24] R. L. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1988.

[25] A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357:241–265, 1984.

[26] A. M. Derrington and P. Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357:219–240, 1984.

[27] R. Desimone, S. J. Schein, J. Moran, and L. G. Ungerleider. Contour, color and shape analysis beyond the striate cortex. *Vision Research*, 25(3):441–452, 1985.

[28] E. A. DeYoe and D. C. Van Essen. Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 1988.

[29] B. M. Dow. Functional classes of cells and their laminar distribution in monkey visual cortex. *J. Neurophysiology*, 37:927–946, 1974.

[30] Ralph M. Evans. *An Introduction to Color.* John Wiley and Sons, New York, 1948.

[31] D. Geiger and F. Girosi. Mean field theory for surface reconstruction and integration. Artificial Intelligence Laboratory Memo 1114, Massachusetts Institute of Technology, 1989.

[32] Stuart Geman and Don Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, 1984.

[33] Alan L. Gilchrist. Perceived lightness depends on perceived spatial arrangement. *Science*, 195:185–187, 1977.

[34] Alan L. Gilchrist, Stanley Delman, and Alan Jacobsen. The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception and Psychophysics*, 33(5):425–436, 1983.

[35] P. Gouras. Identification of cone mechanisms in monkey ganglion cells. *Journal of Physiology*, 199:533–547, 1968.

[36] P. Gouras. The function of the midget cell system in primate color vision. *Vision Research*, 3:397–410, 1971.

[37] P. Gouras and E. Zrenner. Color vision: A review from a neurophysiological perspective. In D. Ottoson, editor, *Progress in Sensory Physiology 1.*, pages 139–179. Berlin: Springer, 1981.

[38] C. H. Graham and J. L. Brown. Color contrast and color appearances: Brightness constancy and color constancy. In *Vision and Visual Perception.* Wiley, New York, 1965.

[39] G. Healey and T. O. Binford. The role and use of color in a general vision system. Artificial Intelligence Lab Technical Report, Stanford University, 1987.

[40] H. Helson and D. B. Judd. An experimental and theoretical study of changes in surface colors under changing illuminations. *Psychology Bulletin*, 33:740–741, 1936.

[41] E. Hering. *Outlines of a theory of the light sense.* Harvard University Press, Cambridge MA, 1964/1874.

[42] C. A. Heywood, B. Wilson, and A. Cowey. A case study of cortical colour "blindness" with relatively intact achromatic discrimination. *Journal of Neurology Neurosurgery and Psychiatry*, 50:22–29, 1987.

[43] B. K. P. Horn. Determining lightness from an image. *Computer graphics and image processing*, 3:277–299, 1974.

[44] B. K. P. Horn. On lightness. Artificial Intelligence Laboratory Memo 295, Massachusetts Institute of Technology, 1974.

[45] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–231, 1977.

[46] B. K. P. Horn. *Robot Vision.* M.I.T. Press and McGraw-Hill, 1985.

[47] D. H. Hubel and M. S. Livingstone. Complex-unoriented cells in a subregion of primate area 18. *Nature*, 315:325–327, 1985.

[48] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195:215–243, 1968.

[49] R. W. G. Hunt. *The Reproduction of Colour.* Fountain Press, Tolworth, England, 1947.

[50] A. C. Hurlbert. Formal connections between lightness algorithms. *J. Optical Society of America A*, 3:1684–1693, 1986.

[51] A. C. Hurlbert and T. A. Poggio. Synthesizing a color algorithm from examples. *Science*, 239:482–485, 1988.

[52] Anya C. Hurlbert and Tomaso A. Poggio. Learning a color algorithm from examples. In Dana Z. Anderson, editor, *Neural Information Processing Systems*. American Institute of Physics, 1988.

[53] C. R. Ingling and E. Martinez. The spatiochromatic signal of the r-g channel. In J. Mollon and L. T. Sharpe, editors, *Colour Vision*. Academic Press, 1983.

[54] K. Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace and Co., 1935.

[55] T. Kohonen. *Associative Memory*. Springer Verlag, Heidelberg, 1977.

[56] A. Kyrala. *Applied Functions of a Complex Variable*. Wiley-Interscience, 1972.

[57] Edwin H. Land. Color vision and the natural image. *Proceedings of the National Academy of Science*, 45:115–129, 1959.

[58] Edwin H. Land. Recent advances in retinex theory and some implications for cortical computations: colour vision and the natural image. *Proceedings of the National Academy of Science*, 80:5163–5169, 1983.

[59] Edwin H. Land. Recent advances in retinex theory. In T. Ottoson and S. Zeki, editors, *Central and peripheral mechanisms of colour vision*, pages 5–17. Macmillan, 1985.

[60] Edwin H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Science*, 83:3078–3080, 1986.

[61] Edwin H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61:1–11, 1971.

[62] H.-C. Lee, E. J. Breneman, and C. P. Schulte. Modeling light reflection for computer color vision. Technical report, Eastman Kodak Company, 1987.

[63] Hsien-Che Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America*, 3:1694–1699, 1986.

[64] P. Lennie, P. W. Haake, and D. R. Williams. Chromatic opponency through random connections to cones. *Supplement to Investigative Ophthalmology and Visual Science*, 30(3):322, 1989.

[65] Lynette L. Linden and Jerome Lettvin. Colors that come to mind. Research Lab of Electronics Technical Report, Massachusetts Institute of Technology, 1982.

[66] M. S. Livingstone and D. H. Hubel. Anatomy and physiology of a color system in the primate visual cortex. *J. Neuroscience*, 4:309–356, 1984.

[67] M. S. Livingstone and D. H. Hubel. Segregation of form, color and stereopsis in primate area 18. *J. Neuroscience*, 7:3378–3415, 1987.

[68] N. K. Logothetis, P. H. Schiller, E. R. Charles, and A. C. Hurlbert. Perceptual deficits and the activity of the color opponent and broad band pathways at isoluminance in monkeys. *Submitted for publication*, 1989.

[69] G. MacKay and J.C. Dunlop. The cerebral lesions in a case of complete aquired color-blindness. *Scott. Med. Surg. J.*, 5:503–512, 1899.

[70] Lawrence T. Maloney. Computational approaches to color constancy. Technical Report 1985-01, Stanford University, 1985.

[71] Lawrence T. Maloney and Brian Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3:29–33, 1986.

[72] R. E. Marc and H. G Sperling. Chromatic organization of primate cones. *Science*, 195:454–456, 1977.

[73] D. C. Marr, T. Poggio, and E. Hildreth. Smallest channel in early human vision. *Journal of the Optical Society of America*, 70(7):868–870, 1980.

[74] David Marr. *Vision, A computational investigation into the human representation and processing of visual information.* W. H. Freeman, San Francisco, 1982.

[75] David Marr and Tomaso Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[76] David Marr and Tomaso Poggio. Analysis of a cooperative stereo algorithm. *Biological Cybernetics*, 28:223–239, 1978.

[77] Jose L. Marroquin. *Probabilistic Solution of Inverse Problems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1985.

[78] J. J. McCann. The role of simple nonlinear operations in modeling human lightness and color sensations. Technical report, Polaroid Corporation, Vision Research Laboratory, 750 Main Street, Cambridge, MA 02139, 1989.

[79] J. J. McCann, S. McKee, and T. H. Taylor. Quantitative studies in retinex theory – a comparison between theoretical predictions and observer responses to the colour mondrian experiments. *Vision Research*, 16:445–458, 1976.

[80] J. J. McCann, R. L. Savoy, and J. A. Hall. Visibility of low-frequency sine-wave targets: Dependence on number of cycles and surround parameters. *Vision Research*, 18(7):891–894, 1978.

[81] J. C. Meadows. Disturbed perception of colours associated with localized cerebral lesions. *Brain*, 97:615–632, 1974.

[82] C. R. Michael. Color vision mechanisms in monkey striate cortex: Dual-opponent cells with concentric receptive fields. *J. Neurophysiology*, 41:572–588, 1978.

[83] C. R. Michael. Color vision mechanisms in monkey striate cortex: Simple cells with dual opponent-color receptive fields. *J. Neurophysiology*, 41:1233–1249, 1978.

[84] John Mollon. Studies in scarlet. *The Listener*, 1985.

[85] B. Moore and T. Poggio. Representation properties of multilayer feedforward networks. *Abstracts of the First Annual INNS Meeting*, page 203, 1988.

[86] A.L. Nagy. K.F. Purl, and J.S. Houston. Cone mechanisms underlying the color discrimination of deutan color deficients. *Vision Research*, 25(5):661–9, 1985.

[87] J. Nathans, D. Thomas, and D. S. Hogness. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science*, 232:193–202, 1986.

[88] O. Packer, D. R. Williams, N. Sekiguchi, N. J. Coletta, and S. Galvin. Effect of chromatic adaptation on foveal acuity and aliasing. *Supplement to Investigative Ophthalmology and Visual Science*, 30(3):53, 1989.

[89] C. A. Pallis. Impaired identification of faces and places with agnosia for colours. *Journal of Neurology Neurosurgery and Psychiatry*, 18:218–224, 1955.

[90] B. T. Phong. *Illumination for Computer Generated Images.* PhD thesis, Univ. of Utah, 1973.

[91] T. Poggio. On optimal discrete estimation. In G. D. McCann and P. Z. Marmarelis, editors, *Proceedings of the first symposium on testing and identification of nonlinear systems*, pages 30–37. California Institute of Technology, 1975.

[92] T. Poggio. On optimal nonlinear associative recall. *Biological Cybernetics*, 19:201–209, 1975.

[93] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[94] Tomaso Poggio, J. Little, E. Gamble, W. Gillett, D. Geiger, D. Weinshall, M. Villalba, N. Larson, T. Cass, H. Bülthoff, M. Drumheller, P. Oppenheimer, W. Yang, and A. Hurlbert. The MIT Vision Machine. In *Proceedings Image Understanding Workshop*, Cambridge, MA, April 1988. Morgan Kaufmann, San Mateo, CA.

[95] Tomaso Poggio and the staff. MIT progress in understanding images. In *Proceedings Image Understanding Workshop*, pages 25–39, McLean, VA, December 1985. Scientific Applications International Corporation.

[96] V. S. Ramachandran. Perception of shape from shading. *Nature*, 331(6152):163–166, 1988.

[97] K. S. Rockland and J. S. Lund. Intrinsic laminar lattice connections in primate visual cortex. *J. Comparative Neurology*, 216:303–318, 1983.

[98] John Rubin and Whitman Richards. Colour vision: representing material categories. Artificial Intelligence Laboratory Memo 764, Massachusetts Institute of Technology, 1984.

[99] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[100] O. Sacks and R. Wasserman. The painter who became colorblind. *The New York Review of Books*, 34(18):25–34, November 1987.

[101] P. Sallstrom. Colour and physics: Some remarks concerning the physical aspects of human colour vision. Institute of Physics Report 73-09, University of Stockholm, 1973.

[102] Steven A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4):210–218, 1985.

[103] V.C. Smith and J. Pokorny. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, 15:161–171, 1975.

[104] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems.* Winston Sons, Washington, D.C., 1977.

[105] D. Y. Ts'o and C. D. Gilbert. The organization of chromatic and spatial interactions in the primate striate cortex. *The Journal of Neuroscience*, 8(5):1712–1727, 1988.

[106] A. Valberg and B. Lange-Malecki. Mondrian complexity does not improve "color constancy". *Investigative Ophthalmology and Visual Science Supplement*, 28:92, 1987.

[107] R. L. P. Vimal, J. Pokorny, V. C. Smith, and S. K. Shevell. Foveal cone thresholds. *Vision Research*, 29(1):61–78, 1989.

[108] H. L. F. von Helmholtz. *Handbuch der Physiologischen Optik.* J. P. Southall, Editor, Dover Publications, 1909/1962.

[109] G. Westheimer. Diffraction theory and visual hyperacuity. *J. Optom. Physiol. Opt.*, 53:362–364, 1975.

[110] H. M. Wild, S. R. Butler, D. Carden, and J. J. Kulikowski. Primate cortical area V4 important for colour constancy but not wavelength discrimination. *Nature*, 313:133–135, 1985.

[111] D. R. Williams. Seeing through the photoreceptor mosaic. *Trends in Neurosciences*, 9:193–198, 1986.

[112] J. A. Worthey. Limitations of color constancy. *Journal of the Optical Society of America*, 2:1014–1026, 1985.

[113] A. L. Yarbus. *Eye movements and vision*. Plenum Press, New York, 1967.

[114] R. A. Young. Principal-component analysis of macaque lateral geniculate nucleus chromatic data. *Journal of the Optical Society of America*, 3:1735–1742, 1986.

[115] Alan L. Yuille. A method for computing spectral reflectance. Artificial Intelligence Laboratory Memo 752, Massachusetts Institute of Technology, 1984.

[116] S. M. Zeki. Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, 9:741–765, 1983.

[117] S. M. Zeki. Colour coding in the cerebral cortex: the responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, 9:767–781, 1983.

[118] Steven W. Zucker and Robert A. Hummel. Receptive fields and the reconstruction of visual information. Computer Vision and Robotics Laboratory Technical Report 83-17, McGill University, 1983.