Hue Segmentation, Color Circuitry, and the Mantis Shrimp

Thesis by

Frank A. Perez

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1995

(Defended May 25, 1994)

Acknowledgements

I am deeply indebted to Christof Koch for the enthusiasm and the intellectual independence he gave me throughout my studies. I am also grateful to the other members of my thesis committee, past and present, John Allman, Al Barr, Fred Culick, Rod Goodman, Carver Mead, Pietro Perona, and David Van Essen. Special thanks to Carver for helping me clarify analog VLSI spectral processing. For their past encouragement, I owe a heartfelt thanks to Eric Antonsson, Frank Marble, and Theodore Wu. Also, I would like to express my gratitude to Yu-Chong Tai for support and collaborations.

I gratefully acknowledge the financial support provided by the James S. McDonnell Foundation, the National Science Foundation, the Office of Naval Research, the Rockwell International Science Center, and the Hughes Aircraft Company.

I would like to thank all my colleagues, too numerous to mention, at Caltech for stimulating discussions and friendship. Special thanks to Tim Horiuchi for layout of the current fuse chip, Tobi Delbrück for assistance in measuring spectral response curves, Alan Corcorran for help in manufacturing color slides as spectral filters, Chang Liu for fruitful discussions on filter applications, and Tom Cronin of University of Maryland for use of the Mantis shrimp spectral data. For reviewing and improving earlier drafts of this manuscript, I thank Jinko Gotoh, Jörg Kramer, and Ana Millaruelo. Finally, I thank Ana and Jinko for emotional support, and my mother and father for their lifelong support and encouragement.

Credits

I recognize the contributions of the following:

- Ron Alley of the Jet Propulsion Labs for for the solar spectral data shown in Fig. 1.4.
- Cindy Grove of the Jet Propulsion Labs for measurement of the spectral reflectivity curves shown in Fig. 1.5.
- Ricardo Zucca of the Science Center of Rockwell for help in measuring spectral reflectivity curves shown in Fig. 1.6 and Fig. 3.2 and for the data given in Fig. 6.9.
- Steve Shafer of Carnegie Mellon University's Calibrated Imaging Laboratory (which is sponsored by NSF, DARPA, and NASA) for the original data that was used to generate the images in Fig. 3.1.
- Garry Mines for assistance in the spectrophotometer transmissivity measurements for Fig. 6.11.

Abstract

The focus of this thesis is on engineering effective color vision subsystems for object segmentation based on hue and discontinuities in hue. Hue, which encodes color, is a psychophysical scalar variable defined on the ring $[0, 2\pi]$. A computational theory justifying the use of hue for distinguishing material differences is established and novel algorithms are developed to detect its discontinuity. Although the focus of this thesis has been on anthropocentrically based trichromatic systems, some effort is placed in exploring the multi-dimensional spaces of more than three primary colors. A hypothetical explanation is proposed of the 11 spectral sensors of the Mantis shrimp, which performs functionally as a biological spectrum analyzer. The importance of the opponency calculation is emphasized. Its role in optimum filter design offers *hyperacuity* in the spectral domain.

This work encompasses spectral filter design, color space evaluation, computer vision algorithm development, and hardware implementation in custom analog VLSI circuitry. A one pixel (zero dimension) "intensity/normalized-color/hue" sensor is built based on a trichromatic system. For the basic analog circuit element required in hue segmentation, the current-fuse is developed. Guidelines for building higher dimensional sensors in both spatial and spectral domains are presented. Practical offshoots from this research range from color quality sensors for inexpensive printing to vision systems for robotics and autonomous vehicles.

Contents

Acknowledgements iii

Abstract v

List of Figures ix

Glossary of Abbreviations xii

1 Introduction 1

1.1 Color Vision 4

1.2 Color Segmentation 6

1.3 Using Hue for Image Processing 9

1.4 Analog VLSI Implementation 11

1.5 The Number of Color Sensors 13

2 Color Space Comparison 19

2.1 RGB Space 19

2.2 Normalized RGB Space 20

2.3 Opponent Color Spaces 23

2.4 HSI Space 27

2.5 CIE Spaces 28

3 Properties of Hue **35**

3.1 Additive/Shift and Multiplicative/Scale Invariance 35

3.2 Hue in CIE Spaces 36

3.3 Problem of Singularities 37

3.4 Hue Discounts Confounding Cues 38

3.4.1 Discounting Transparency 38

3.4.2 Discounting Highlights and the **Integrated White** Condition 40

3.4.3 Discounting Shading 45

3.4.4 Discounting Shadowing 46

3.5 Rubin and Richards Representation 50

3.6 Physical Basis for Hue's Modulo Nature 52

4 Algorithms for Hue Segmentation **55**

4.1 Finding Edges of a Circular Variable 56

4.2 Performance Comparisons 58

4.2.1 Conventional Edge Detection 58

4.2.2 Results of Modulo Discontinuity Operator 59

4.2.3 Comparisons 59

4.3 Improvement with Regularization 65

4.3.1 Markov Random Field Formulation 65

4.3.2 Results 67

5 Analog VLSI Chip Implementation **69**

5.1 The Analog Representation 70

5.2 Chip Background and Experimental Setup 71

5.3 Normalized Color Sensor 74

5.4 Hue Sensor 82

5.5 Segmentation Circuitry 86

5.6 Future Hardware Issues 92
5.6.1 Color Arrays 92
5.6.2 Improved Hue Sensor 96

6 Beyond Trichromacy: The Mantis Shrimp **103**
6.1 Trichromacy: Sampling versus Encoding 104
6.2 Trichromacy: Principal Components 110
6.3 Color Systems Reviewed 111
6.4 A Model for the Mantis Shrimp 113
6.5 Importance of Opponency 116
6.6 Optimum Spectral Response Filter Design 119
6.7 Improving the Spectral Response of Intrinsic Devices 120
6.8 Proposed 6 Spectral Sensor System in Analog VLSI 124

7 Conclusions **127**
7.1 Future Directions 129

A Other Models for Discounting highlights **131**
A.1 Cook-Torrance Model 131
A.2 Dichromatic Model 134

B Modulo Algorithms for Digital Hardware **136**

C Spectral Filter Design **139**
C.1 Filter Parameters that Affect Hue 139
C.2 Optimum Separation of Gaussian Filters 144

References **147**

List of Figures

1.1 Natural scene as viewed with and without spectral filters 5
1.2 The geometry of image formation 7
1.3 Hue representation in neurobiology. 10
1.4 Solar and reflected sky natural illumination 14
1.5 Spectral reflectance of various colored papers 15
1.6 Measured radiance of various colored papers 16
1.7 Spectral sensitivity of the human eye 18

2.1 Color images of Mondrian and bell peppers 21
2.2 RGB components of Mondrian and peppers 22
2.3 Normalized RGB of Mondrian and peppers 24
2.4 Details of hue space 29
2.5 HSI of Mondrian and peppers 30
2.6 CIE L*a*b* components of Mondrian and peppers 34

3.1 Benefit of hue on plastic objects 43
3.2 Measured radiance of green paper under natural illumination 49
3.3 Rubin and Richards' material representation in trichromatic space . . 51
3.4 Absorption band of materials justifies the Hue measure 54

4.1 RGB edges of Mondrian and peppers 60

4.2 Normalized RGB edges of Mondrian and peppers 61
4.3 HSI edges of Mondrian and peppers 62
4.4 CIE L*a*b* edges of Mondrian and peppers 63
4.5 Smoothed hue edges via regularization 67

5.1 Filter spectral transmissivity 73
5.2 Intensity variation in experimental setup 75
5.3 Normalization circuit examples and function 79
5.4 Measured response of normalized color circuit 80
5.5 Measured response of improved normalized color circuit 81
5.6 Block diagram of hue circuit 84
5.7 Measured response of the hue circuit 85
5.8 Comparison of measured hue circuit output and other hue measures . 87
5.9 Current-fuse circuit element 89
5.10 Current-fuse chip layout 90
5.11 Current-fuse chip test data 91
5.12 Simulated reconstructed hue and hue edges from subsampled image . 94
5.13 Color filter array geometries 95
5.14 Comparison of CMOS and Bipolar Translinear Circuits 98
5.15 Comparison of calculated hue of various Gaussian filter configurations 101
5.16 Computed hue currents from intrinsic diodes 102

6.1 Fourier transform of the human receptor cones and psychophysical response to "comb-filtered" stimuli. 105
6.2 Signal gamuts within human visual space 108
6.3 Sampling versus encoding 109
6.4 Mantis shrimp architecture 114
6.5 Spectral response curves of Mantis shrimp photoreceptors 115

6.6 Simulated spectral response curves for an opponent system based on the Mantis shrimp 117
6.7 Spectral sharpening afforded by the opponency operation 118
6.8 Spectral responses for 5 cone butterfly 120
6.9 Spectral response of intrinsic silicon devices 121
6.10 Spectral response of intrinsic silicon devices sharpened with film . . . 123
6.11 Green-red color mixing on Kodak slide filters 125
6.12 Artificial multispectral system based on Kodak slide filters 126

C.1 Two cones for hue 140
C.2 Calculated hue from 2 cones of varying widths 141
C.3 Hue span reduction 142
C.4 Figure of merit for spectral widths 143
C.5 Offset Gaussians 146

Glossary of Abbreviations

A/D	Analog to Digital
BiCMOS	Bipolar CMOS
CCD	Charge-Coupled Device
CMOS	Complementary MOS
CIE	Commission International de L'Eclairage
HSI	Hue, Saturation, and Intensity
L*a*b*	CIE color space
L*u*v*	CIE color space
MOS	Metal Oxide Semiconductor
MOSIS	MOS Implementation System
	(Government subsidized chip fabricator)
MRF	Markov Random Field
NTSC	National Television Systems Committee
	(or affectionally known as Not Twice the Same Color)
Nrgb	Normalized Red, Green, and Blue
THz	Terahertz (10^{12} cycles/second)
V4	A visual cortex area in primate
VLSI	Very Large Scale Integration
XYZ	CIE color space
YIQ	NTSC color space

Chapter 1

Introduction

Although color is one of many cues that help discriminate one object from another, it plays an important and unique role. While texture, intensity, depth, and motion are other cues that contribute to the object segmentation task, they differ from color in one important respect: color sensing requires extra "front-end" equipment. Spectral processing requires special sensing hardware (cones) and circuitry (neurons), thus demanding more resources and power. Nevertheless, the evolution of biological visual systems shows a trend toward an increase in and refinement of multispectral processing. In primates we see the recent evolution of the middle wavelength cone from the long wavelength cone (Nathans et al., 1986), thus creating a trichromatic system from a dichromatic one. Then there is the case of the Mantis shrimp, with at least 11 spectral sensors (Cronin and Marshall, 1989b). Why does it need so many cones? What can we learn from its multi-spectral processing?

The underlying theme of this thesis is the design and manufacture of color sensors, which with further refinement can be integrated into effective visual hardware. We envision utilizing analog VLSI systems because of the speed, cost, power, and efficiency advantages they offer over other alternatives. To date, only monochromatic

implementations of these analog VLSI systems have been realized. The addition of color sensing to artificial analog systems offers new representations of the visual environment and boosts performance levels. This parallels the evolution of multi-spectral sensing in biological systems. Suitable engineering by-products from this research range from color quality sensors for inexpensive printing to vision systems for robotics and autonomous vehicles.

Overall, this thesis addresses color segmentation issues within Marr's framework:

- What is the goal of color segmentation computation and why is it appropriate (Chapter 3)?
- What algorithm will implement this color computation (Chapter 4)?
- What hardware can be built for this (Chapter 5)?

This paradigm for information processing encompassing theory, algorithm, and hardware (Marr, 1982) is a powerful tool for carrying out any vision-understanding task. Although many representations and algorithms can be utilized to realize the goal of color segmentation, the physical embodiment in hardware limits and defines the algorithm.

At the core of this thesis is the utilization of the psychophysically based hue parameter and its role in segmentation. A recurrent theme in this endeavor is the imitation of biological systems. In particular, the hardware implementation of the hue sensor (Chapter 5) was made viable only after the utilization of the biologically-based opponent color system. An outline of this thesis is as follows:

- Chapter 1[1] introduces color segmentation issues and history and motivates the utilization of hue. Extension of color segmentation to multispectral domains is

[1]Portions of this chapter (1.1 to 1.4) in addition to portions of chapters 2, 3, 4 and 5 have already appeared in publication(Perez and Koch, 1994).

suggested by the Mantis shrimp.

- Chapter 2 defines nomenclature and commonly used color spaces.

- Chapter 3 evaluates the advantages and disadvantages of the hue space, and how it compares to normalized color. Important properties that relate color variables to the underlying material properties are emphasized.

- Chapter 4 develops novel operations for convolution and regularization to find discontinuities in modulo hue space. Performance is compared to standard methods applied to real images.

- Chapter 5 describes the hardware implementation of normalized color and hue sensors. Spectral tuning curves are measured and compared to biological data. Enhancements of these sensors are discussed. Also, a current-fuse circuit, which has direct application in hue segmentation, is described and tested.

- Chapter 6 focuses on multi-spectral sensor systems and speculates on the 11-cone Mantis shrimp function. Filter optimization and sampling is discussed. An artificial 6 cone silicon-based sensor is proposed.

- Chapter 7 summarizes the thesis and indicates future directions for research.

The novel contributions of this research include: (1) Establishing a connection between the hue variable and physical properties of materials, (2) Developing modulo operators to detect discontinuities in the hue variable, (3) Design and testing of biologically motivated analog VLSI color sensors, and (4) Proposing multi-spectral vision hardware systems based on a study of the spectral filter design of the Mantis shrimp.

1.1 Color Vision

The primary goal of color segmentation is to determine where changes of material occur in a visual scene. While it is true that most of the information content in an image is contained in the achromatic intensity channel[2] (Buchsbaum and Gottschalk, 1983), it is clear that there is some merit to evaluating spectral differences in scenes. For example, locating the red berries in the achromatic scene of Figure 1.1a is made easier through the use of a red filter as shown in the filtered scene of Figure 1.1b. Such a simple task as finding red berries amongst green foliage suggests that biological systems obtain an advantage, through evolution, from the addition of extra chromatic sensors.

While a material's surface properties are associated with its spectral reflectance signature (also called the "bidirectional spectral reflectivity" (Horn, 1986; Siegel and Howell, 1981) or "albedo" when independent of viewing geometry such as for lambertian surfaces), it is only the image radiance—spectral reflectance multiplied by the illumination or spectral irradiance—that is available for data processing at the image sensors.

$$\text{Radiance}(\lambda) = \text{Irradiance}(\lambda)\ \text{Albedo}(\lambda)$$

This relationship is illustrated in Figure 1.2. Thus, the surface spectral reflectance, which is a material property, is independent of illumination, while the image radiance is not. The measurement of surface color from the image radiance is an approximation to obtaining the spectral reflectance signature[3]. This absolute measurement of surface

[2]For human vision, 97% of the signal energy is contained in the luminance channel. See Chapter 2 for further discussion.

[3]The problem is underconstrained and requires some assumptions for its solution. The typical

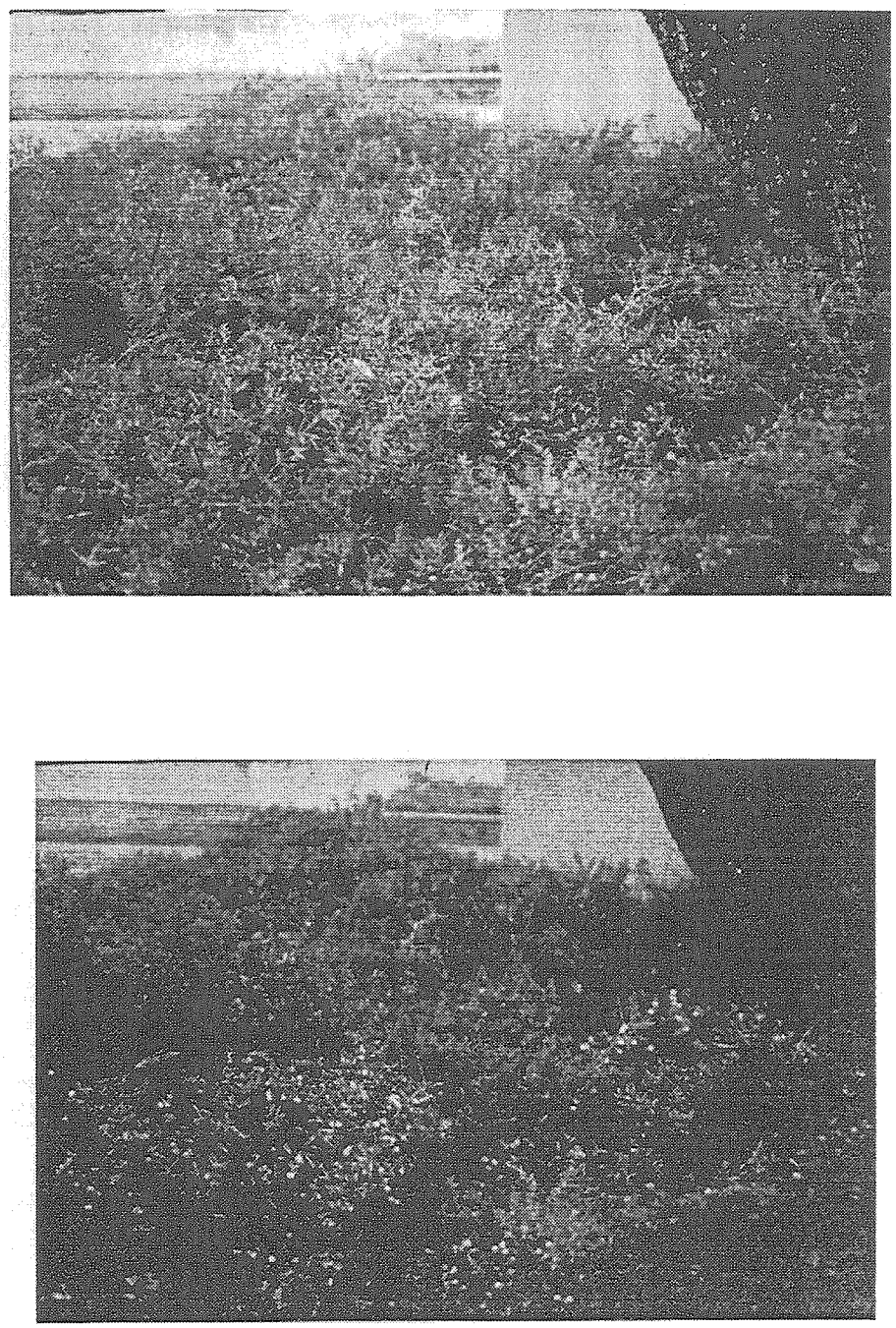


Figure 1.1: Natural scene photographed with a standard 35 mm camera using red and grey filters. Notice that the reddish objects (berries) in the scene are accentuated when the two images are compared.

color, which is an illuminant-independent property, is known as "color constancy" (D'Zmura and Lennie, 1986; Hurlbert, 1989; Maloney and Wandell, 1986) and falls outside the scope of this study. What is of interest here is the measurement of relative color differences calculated from the image radiance. It is this task that we call color segmentation, which in general, is associated with distinguishing material differences and thus object boundaries.

1.2 Color Segmentation

Recent work on color segmentation has been based on physically based models in multispectral environments. Rubin and Richards describe an approach based on assigning material differences to spectral crosspoints (Rubin and Richards, 1982) and to the opposite spectral slope condition (Rubin and Richards, 1984). Gershon, Jepson, and Tsotsos (1986) incorporate that idea in a double-opponent center surround operator to distinguish material changes from shadow boundaries. Other approaches involve separating image radiance into surface reflection and body reflection. For example, Klinker, Shafer, and Kanade (1988, 1990) discuss a physical-based color reflection model (the dichromatic model (Shafer, 1985)) that accounts for highlight reflection and matte shading to improve segmentation. Their method evaluates planar clusters in three-dimensional color space while considering camera limitations. Similarly, Healy and Binford (1987, 1989) describe a reflectance model of materials to classify metals and dielectric surfaces.

Non-physically based segmentation models, in particular those that utilize hue, have been few and far between. Ohlander (1976) uses recursive thresholding on histograms of a 9-parameter color space (RGB, HSI, YIQ) for his segmentation algo-

assumption is that the illumination and the spectral reflectance of objects in the scene can be constructed with a few orthogonal basis functions.

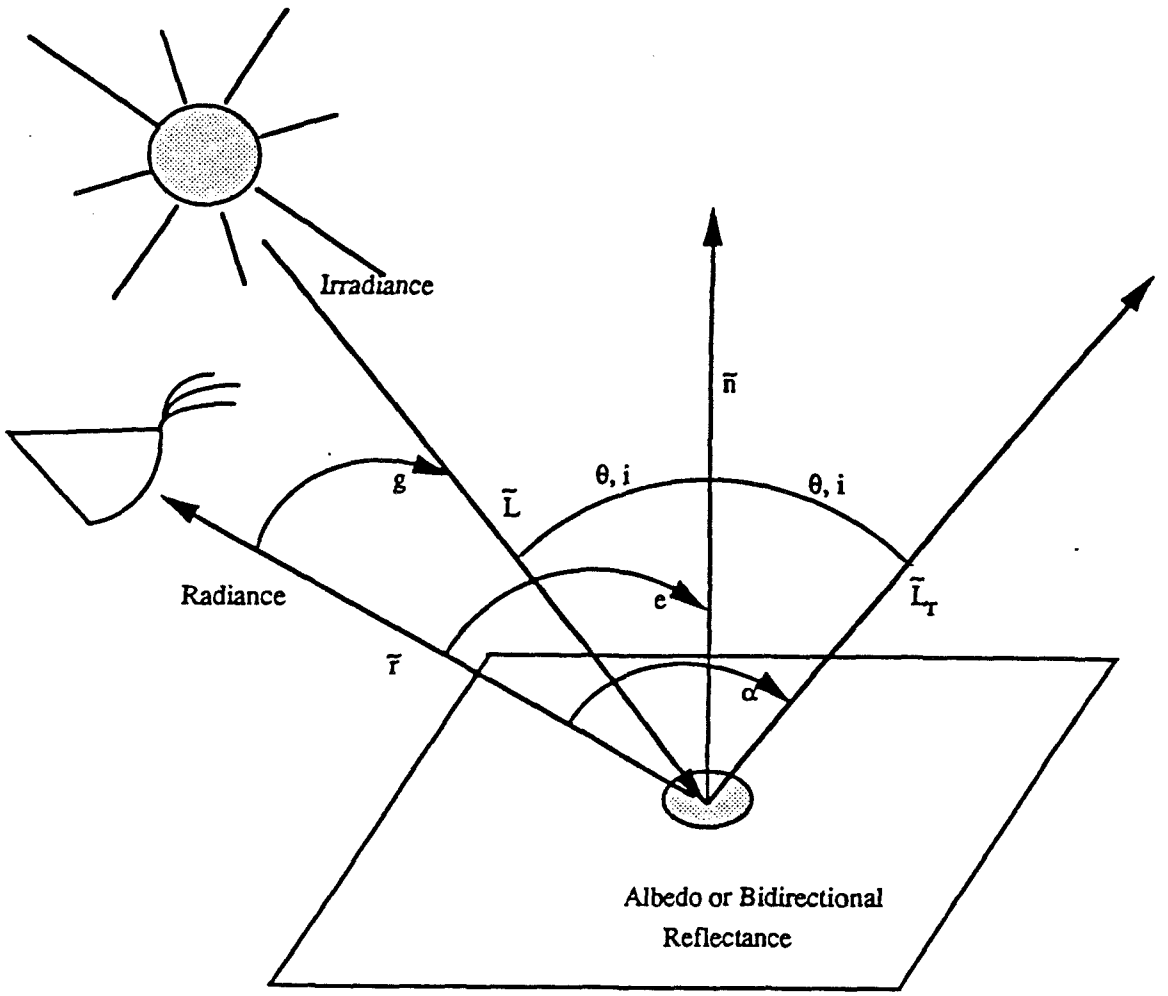


Figure 1.2: The Geometry of Image Formation. The surface property, albedo, is contained in the radiance signal–viz. Radiance(λ)=Irradiance(λ) Albedo(λ). (Alternate variables i, e, g are utilized in the Dichromatic Model (Shafer, 1985).

rithm. Ohta (1980) evaluates the performance of 7 different color spaces (21 parameters, 3 of which were HSI) utilizing Ohlander's algorithm. In recent times, Celenk (1990) utilizes the peaks of the 1-D histograms of the equivalent hue-saturation-intensity (HSI) coordinates in the CIE (L*a*b*) uniform color coordinate system to identify cluster regions and then projects the clusters onto a line for 1-dimensional thresholding. Similarly, Tominaga (1990) utilizes iterative histogramming on the principal components of the CIE (L*a*b*) color space to identify cluster regions, and then applies a grouping operation dependent on hue difference thresholds.

Bajcsy (1990) uses a physically based model to construct HSI from an orthogonal basis space whereupon shading, highlights, shadows, and inter-reflections are discounted. Nevertheless, clustering, histogramming, and thresholding are applied to the hue parameter. As an extension, Lee (1991) improves segmentation performance by considering multiple views to compensate for viewpoint-dependent specular reflections.

With the exception of Rubin and Richards' approach these methods use classical pattern recognition techniques—histogram thresholding, linear discriminant function, recursive region masking, clustering, etc.—which require heavy computational resources. These techniques represent neither plausible schemes for naturally-evolved, efficient, biological systems nor lend themselves very easily to dedicated analog "vision chips." This research is fueled by the desire to understand the basis of color segmentation in primate cortex as well as to imitate those structures in analog CMOS VLSI circuits (see for example (Harris, 1990b; Koch et al., 1991; Perez and Koch, 1992)). In particular, the low-accuracy and inhomogeneity of the circuit components we are using (Mead, 1989) forces us—as in biological evolution—to consider simple and robust implementations of vision algorithms. Accordingly, we reevaluate and justify a simplified hue description, propose an algorithm for hue segmentation and demonstrate its adaptation to the first version of an analog VLSI hue chip.

1.3 Using Hue for Image Processing

In this study, the psychophysically based parameters hue, saturation, and intensity were used as the starting point for color segmentation. Functional studies suggests that hue is computed at a high level in the nervous system (Desimone et al., 1985; Schein and Desimone, 1990; Zeki, 1983). An examination of the spectral response curves in area V4 of the monkey visual cortex bears this out. For example, we partially reproduce Schein and Desimone's (1990) Fig 2C and Fig. 4 in Fig. 1.3. Fig. 1.3a shows the measured response of what can be labeled a "purple" cell, since it responds maximally to both red and blue stimuli. Fig. 1.3b shows the distribution of 71 cells as a function of peak wavelengths. Here, the whole visible spectral range is spanned by the V4 neurons. Neurobiology literally constructs dozen of color classes from three basic primary classes.

Furthermore, anthropological studies indicate that hue ordering and color naming evolution are universal and are not culturally unique (Berlin and Kay, 1969). Heider and Olivier (1972) investigated the short-term color memory of New Guinea Dani, a stone-age agricultural people whose color lexicon is only limited to two terms: *milli* for dark/cold hues and *molla* for bright/warm hues. Through psychophysical experiments they found that the Dani's color representation formed cylindrical structures very much akin to Western color scales.

In the parlance of system engineering, this suggests that hue is a high level variable. In addition to its circular (modulo) nature, it has invariance to illumination changes and white light additions. The motivation for its selection in image segmentation is that material boundaries correlate more strongly with hue than with intensity discontinuities. Shadow boundaries are strongly associated with intensity edges, and less so with hue boundaries. The same is said for highlight, transparency, and shading boundaries. These conjectures are evaluated in Chapter 3. (See also the

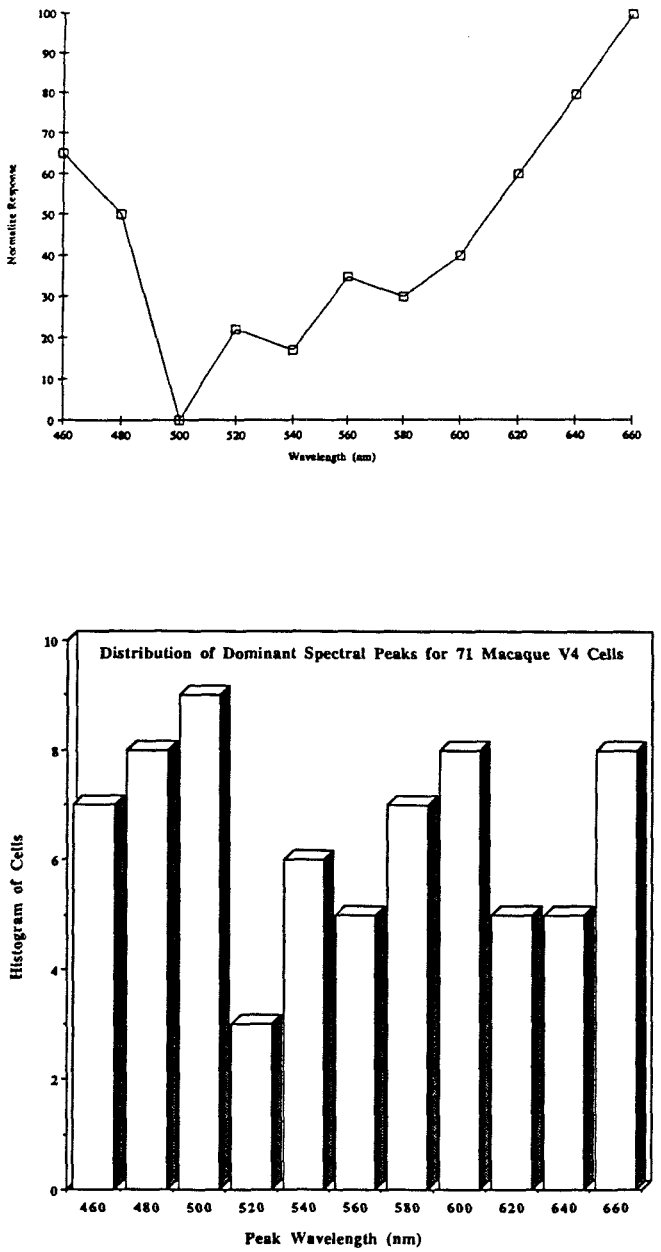


Figure 1.3: Partial Reproduction of Schein and Desimone's (1990) Fig. 2C and Fig. 4 which shows the spectral properties of V4 neurons in the Macaque. (a) Normalized response of a 'purple" cell. (b) Histogram distributions of peak wavelengths for 71 cells which have mean half-bandwidth of 27 nm.

study of (Bajcsy et al., 1990)). Furthermore, segmentation in the one-dimensional hue parameter space is computationally less expensive than in the three-dimensional RGB space. In comparison to other techniques, the method developed here is simpler yet as effective in removing confounding cues as the color clustering and histogram thresholding methods developed for color image segmentation. Also, because hue has basis in biological hardware, it is readily implemented in electronic analog hardware.

The history of hue as a computational variable for image segmentation has not been favorable. Kender's 1976 study showed that nonlinear color transforms such as HSI and normalized color have essential singularities, and spurious modes due to the digitizing nature of the nonlinear transforms. Kender's recommendation was to use linear spaces. Still, researchers (Healey, 1989; Hurlbert and Poggio, 1989; Nevatia, 1977; Poggio et al., 1988; Rubin and Richards, 1984) have used normalized color as the basis of an illumination independent color space.

Few studies exist that address the special considerations of modulo computation. (Winfree (1990) is a rare example.) Perhaps this is due to the limited application of modulo variables (phase, orientation, and hue are the only variables that come to mind). Chapter 4 describes the necessary tools to segment images in hue space. Modulo convolutions and regularization utilizing the Markov Random Field approach are developed. This can be compared to Yoon (1991), who describes a computationally-extensive geodesic-based clustering/segmentation algorithm in the hue-saturation space.

1.4 Analog VLSI Implementation

After the algorithms were developed and tested, the next phase was their implementation onto special purpose hardware. In Chapter 5 we describe a zero-dimension (one pixel) intensity/normalized-color/hue chip we manufactured, complete with on-

board photoreceptors and custom analog VLSI circuitry. The spectral differentiation of photoreceptor elements was achieved by the manual placement of RGB gelatin filter elements directly on the chip die photoreceptors[4]. This manufacture of a neurologically based integrated color sensor lacks precedents. On the one hand there are analog vision chips emulating various neurobiological components and circuits, such as the silicon retina (Mead, 1989), pyramidal cells (Mahowald and Douglas, 1991), and superior colliculus (Horiuchi et al., 1994). On the other hand there are color imagers (Parulski et al., 1992; Watanabe et al., 1984) utilizing digital circuits and RGB CCD arrays for the purpose of replicating color images. In this thesis we combine the efficient computing of analog VLSI with the color sensing approaches of digital CCD imagers to make an effective color vision sensor. Closest in spirit to this work is the color constancy analog computer (Moore et al., 1991), which utilizes analog VLSI resistive grids to compute color corrected RGB channels from a CCD camera.

Additionally, in Chapter 5 we describe a segmentation (current-fuse) circuit that complements our hue sensor. The future integration of this segmentation circuit with the hue sensor is the fundamental element for the *smart color sensor.*

Manufacturing one and two-dimensional color systems was not explored in this thesis due to the apparent technical limitations encountered earlier on in the research. The manual placement of spectral filters directly on the chip precludes any dense spatial positioning. This difficulty can now be overcome with color array filter mask additions discussed in Chapter 5. Extensions to one and two-dimensional color systems are now forthcoming.

[4]Utilizing intrinsic silicon devices such as buried diodes at varying depths to exploit the differential absorption of light in silicon apparently is not beneficial within the visible spectrum according to Delbrück (1993). However, we re-evaluate the use of intrinsic diodes in Chapters 5 and 6.

1.5 The Number of Color Sensors

As mentioned earlier, image radiance is the multiplication of the surface spectral reflectance by the spectral illumination or irradiance. Figure 1.4 shows the prototypical natural illumination available during a midlatitude summer. The figure shows illumination as a function of wavelength for the overhead sun and various portions of the sky. We notice that sky-light spectrum has a noticeable blue shift when compared to the solar spectrum.

Measured illumination-independent surface spectral reflectances of 4 colored papers are shown in Figure 1.5. The corresponding image radiance of these papers under the natural illumination is shown in Figure 1.6. At first glance, reconstruction of the surface spectral reflectance from the image radiance looks like a foreboding project considering the spectral diversity of materials. Fortunately, principal component analysis of surface reflectance signatures suggests that objects may be accounted for by a small number of components. For example, Maloney (1986) shows that a three parameter linear model will empirically fit a large number of surface reflectances.

Obviously, three is not a magic number. Numerous biological systems have four, five, and six cone systems. Recently, the Mantis shrimp (Cronin and Marshall, 1989b) was found to contain 11 color sensors. What governs this diversity is the system requirement of accurate spectral discrimination for the job at hand. For example, remote sensing artificial systems, such as an airborne radiometer with 512 spectral sensors covering the 0.4 to 1.1 μm region (Chang and Collins, 1983), have been used to detect metal sulfides in soils by examining for a tell-tale spectral blue shift (7 to 10 nm) of the chlorophyll red-edge in coniferous and deciduous trees. To paraphrase a maxim, this application "sees the *leaves* for the trees."

Thus, the measurement of surface color is an approximation to an object's spectral signature. It is an approximation because the measurement of color is obtained by

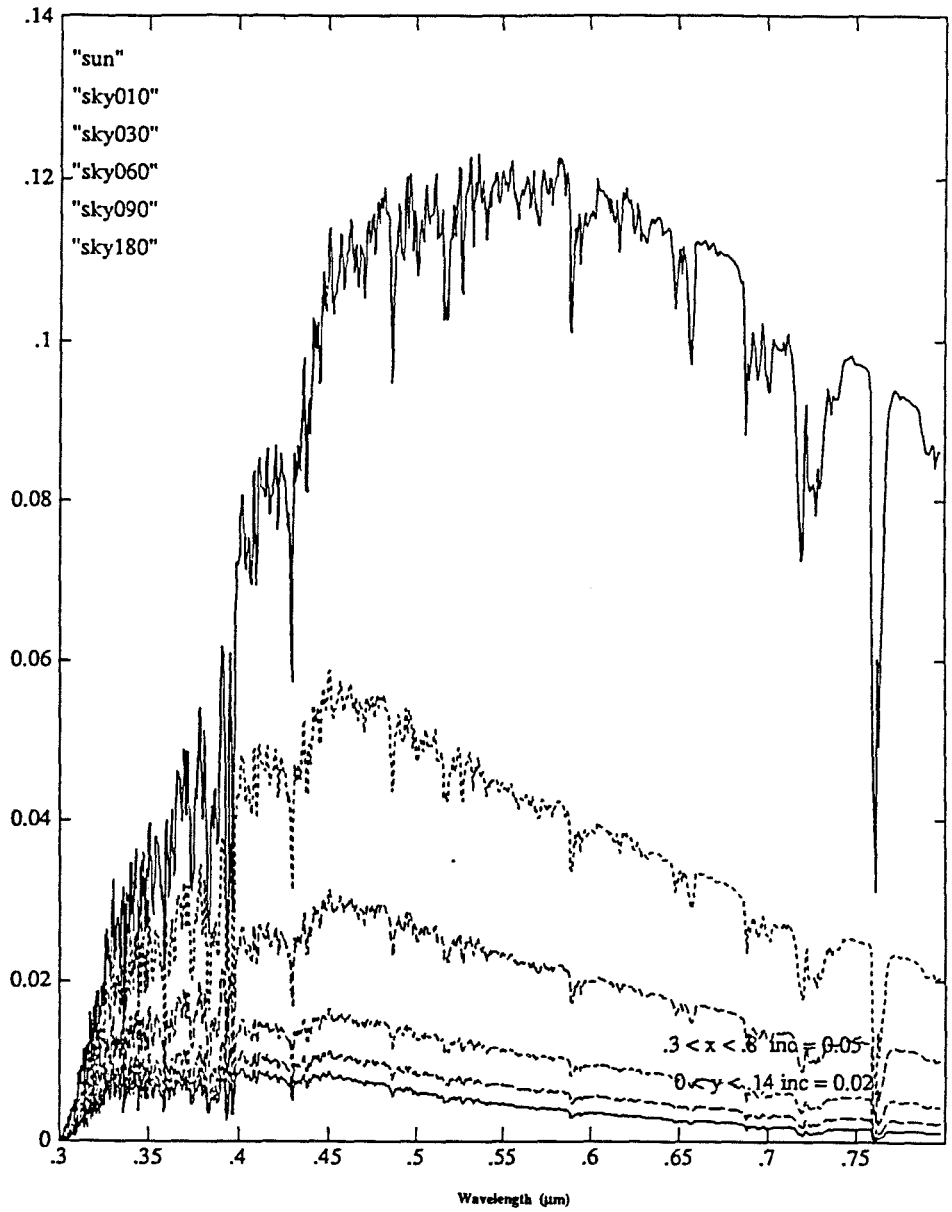


Figure 1.4: Typical illumination from various portions of the sunlit sky based on a LOTRAN-7 software simulation courtesy of Ron Alley of Jet Propulsion Laboratory. The reflected sky radiance is based on a model midlatitude summer atmosphere with rural aerosols and assumes Mie scattering. Notice that the ambient sky illumination has a noticeable blue shift when compared to the solar spectrum. Also notice that to a first order approximation, the scattering from various portions of the sky differs by an *additive shift* in spectral intensity.

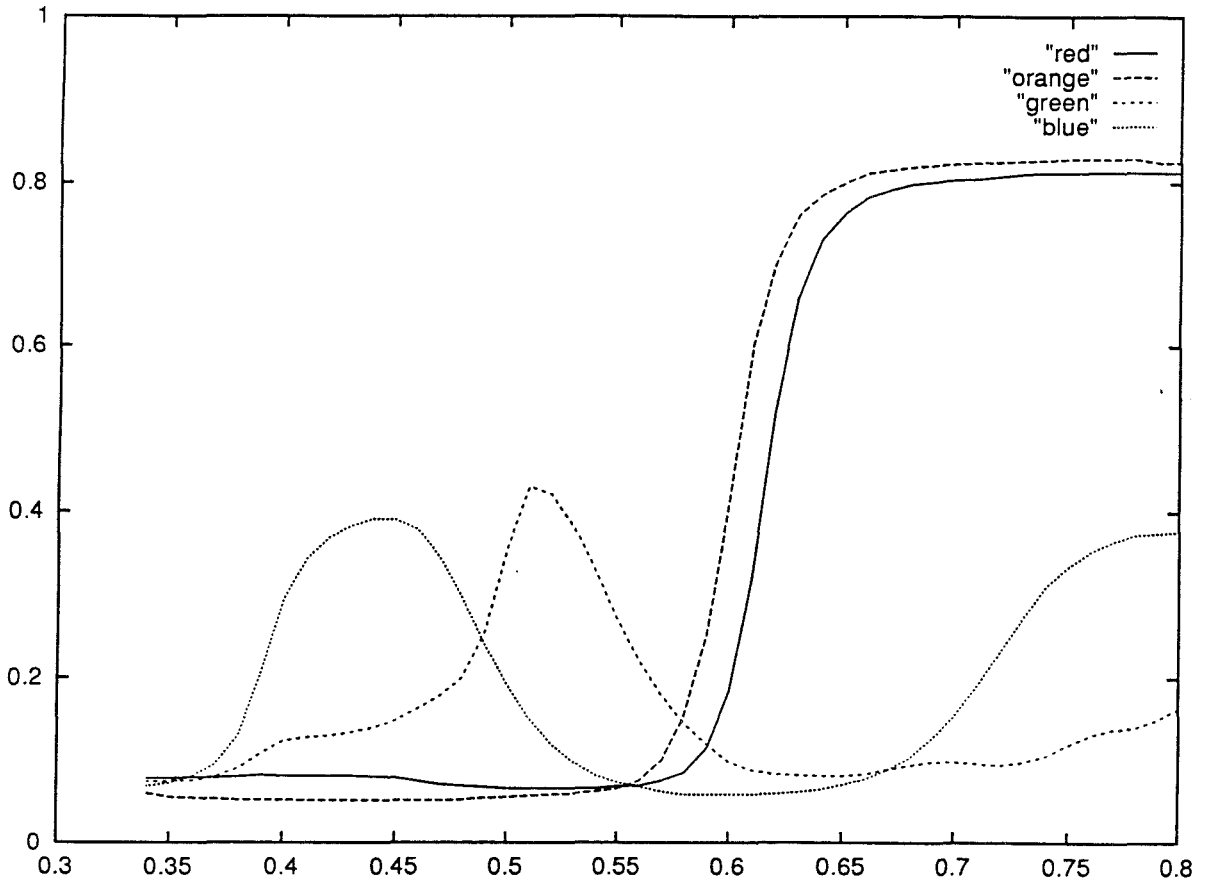


Figure 1.5: Surface Spectral Reflectance of Red,Orange,Green,and Blue Papers. This was measured using JPL's Beckman UV5240 Spectrophotometer.

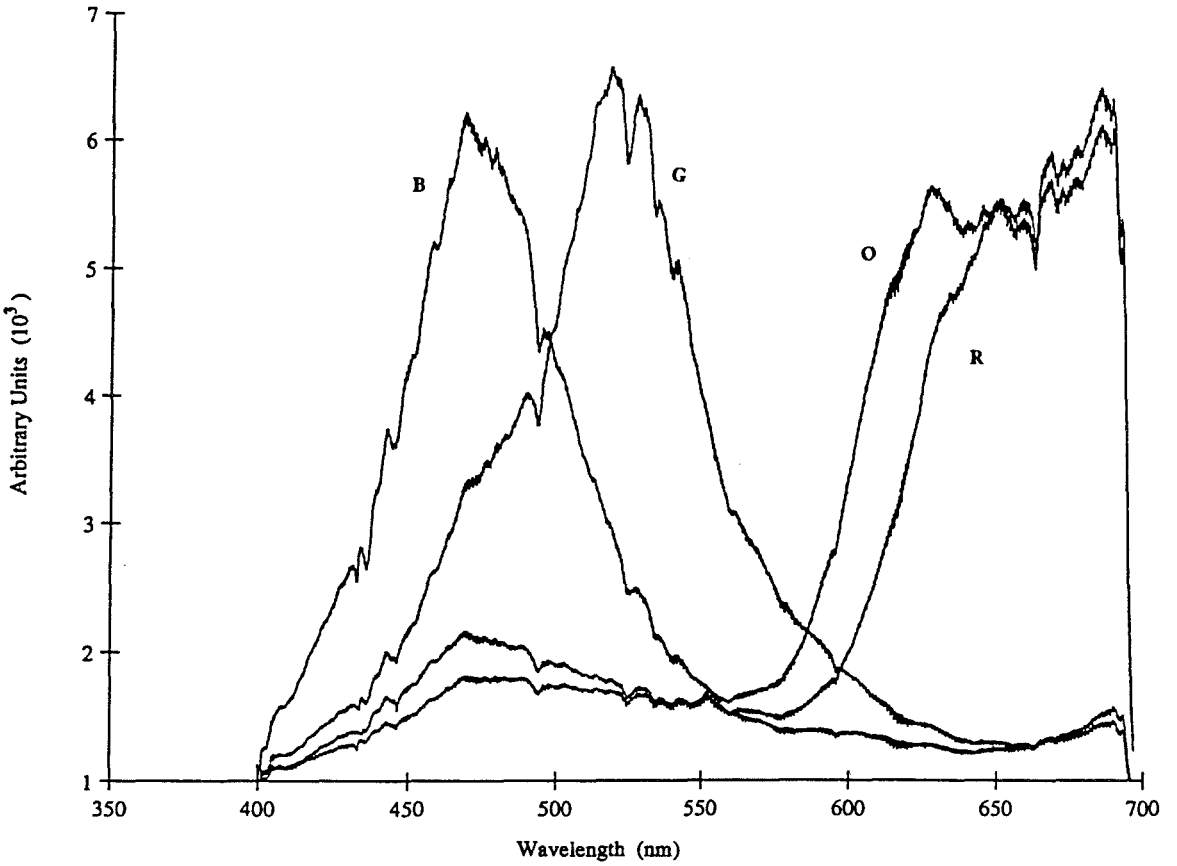


Figure 1.6: Measured radiance of red, orange, green, and blue papers from a typical March Southern California early afternoon sunlight utilizing a Jarrell Ash 275 cm spectrometer with a Tracor Northern Optical Multichannel Analyzer courtesy of Rockwell International.

a small and finite number of spectrally integrating sensors, usually on the order of three sensors for many natural and artificial systems. As an example, the spectral response for human cones is shown in Figure 1.7 and is similar to the RGB system of many color camera systems.

Based on the Nyquist sampling theorem, justification for trichromacy has been proposed (Barlow, 1982; Bowmaker, 1983). The sampling theorem establishes the number of independent channels that can be constructed given the available spectral range and sensor spectral bandwidth. In the case of the primate visual system the number of independent channels is calculated to be three. Although this appears as a justification for trichromatic systems, it cannot explain the 11-cone Mantis shrimp. Also, sampling theorem arguments appear to directly conflict with the requirements for accurate spectral encoding operations. To sample accurately, more narrowly-tuned sensors are required; on the other hand, encoding requires broadly-tuned overlapping spectral responses. Obviously, these contradictory requirements are resolved in biology. Chapter 6 explores some of the pertinent issues related to this topic within the Mantis shrimp.

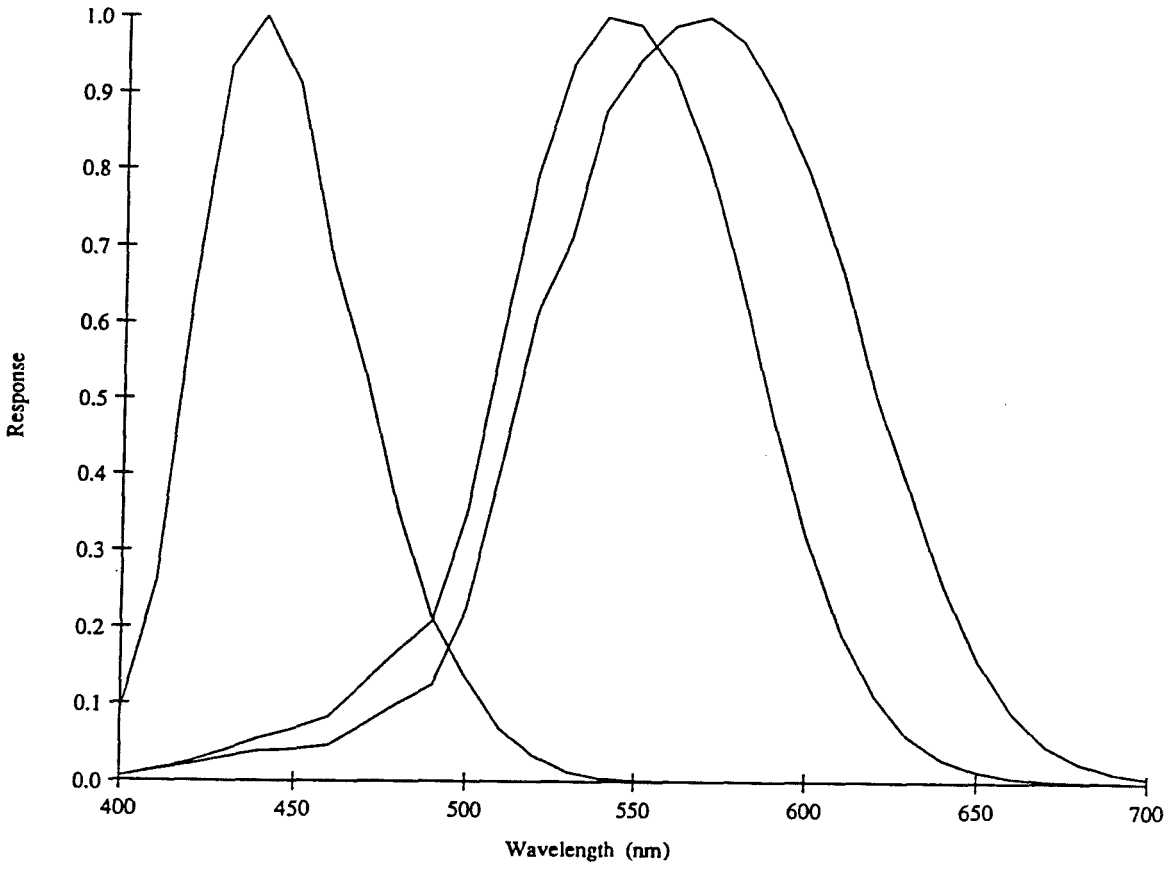


Figure 1.7: Spectral sensitivity of the human eye (Ingling and Tsou, 1977). (Compare this to the spectral response curves for the artificial color sensor in Chapter 5.) The substantial spectral overlap of the long- and middle-wave cones supports their recent evolution (Nathans, et al., 1986) but contains suboptimal spectral crossing patterns (see Chapter 6). It has been argued (Barlow, 1982; Mollon, 1990) that form vision and thus spatial resolution would be compromised if the sensitivity curves became too separated thus resulting in an increase in chromatic aberration.

Chapter 2

Color Space Comparison

RGB, normalized RGB (Nrgb), YIQ, HSI, Opponent color, Munsell, and various CIE spaces (Foley et al., 1990; Joblove and Green, 1978; Ohta et al., 1980; Schwarz et al., 1987; Smith, 1985; Wyszecki and Stiles, 1982) are a few of the most widely used color spaces. The existence of many different color spaces is largely a result of color scientists attempting to construct perceptually uniform color organization. The Munsell and the CIE spaces in particular fall into this subjective category. The spaces introduced in this chapter and examined in the thesis are the RGB, Nrgb, Opponent color, YIQ, HSI, and CIE (L*a*b*) color spaces[1].

2.1 RGB Space

The Red, Green, and Blue colors represent the tristimulus components and define the basic color space(Wyszecki and Stiles, 1982). Each of these components corresponds to a filtered spectral mapping from image space to a 3-D sensor space. The equation

[1]Portions of this chapter have already been published (Perez and Koch, 1994).

governing this transformation is:

$$C = \int_{\lambda} E(\lambda) S_C(\lambda) d\lambda \quad \text{for } C = (R, G, B) \tag{2.1}$$

where C are the tristimulus values, $E(\lambda)$ is the incoming light intensity or radiance, and S_C are the three hypothetical color filters.

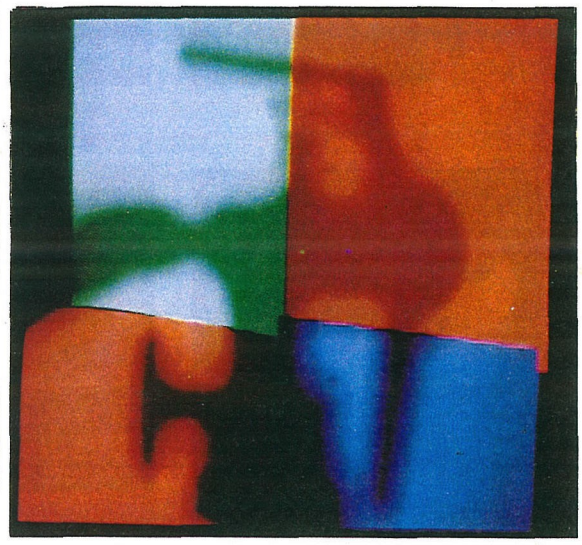
The original color images in Fig. 2.1 were obtained in the laboratory utilizing a Panasonic WV-D5000 RGB camera. Separate RGB components of that image are shown in Fig. 2.2. These images were chosen because they contained some attributes of shading, transparency, highlights, and shadowing. Fig. 2.1a is a simplified "Mondrian" consisting of four patches with the shadow of a camera tripod head cast upon it. Fig. 2.1b is a group of peppers in a semi-transparent bowl. In both images, multi-light sources are evident. A projector halogen lamp was used for direct illumination, and ambient fluorescent light provided background illumination.

The major problem of the RGB space is that segmentation needs to be performed in 3-D space. Along these lines, Wright fused R, G, and B images using Markov Random Fields (Wright, 1989). The results show that each of the components in the RGB space are highly correlated and not independent of each other. This result is also confirmed in the principal component analysis study of (Ohta et al., 1980) which will be discussed in Section 2.3.

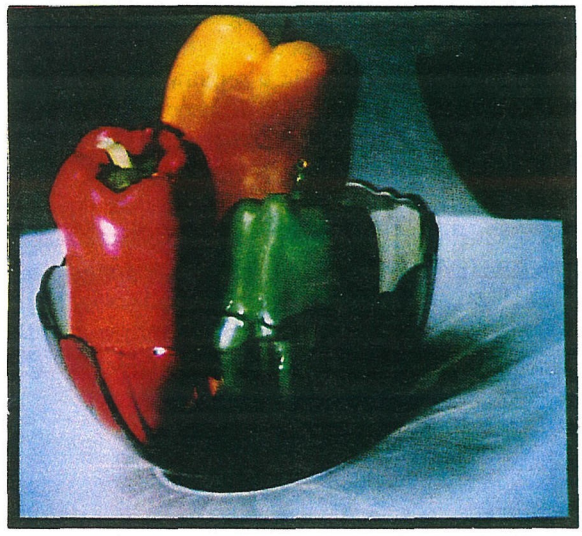
2.2 Normalized RGB Space

Nrgb gives a space that is independent of uniformly varying lighting levels. The transformation to normalized colors is given by:

$$N_C = \frac{C}{(R + G + B)} \quad \text{for } C = (R, G, B) \tag{2.2}$$



(a)



(b)

Figure 2.1: Original color images of (a) shadowed Mondrian and (b) peppers. In both images a projector lamp was used for direct illumination while fluorescent light provided ambient illumination.

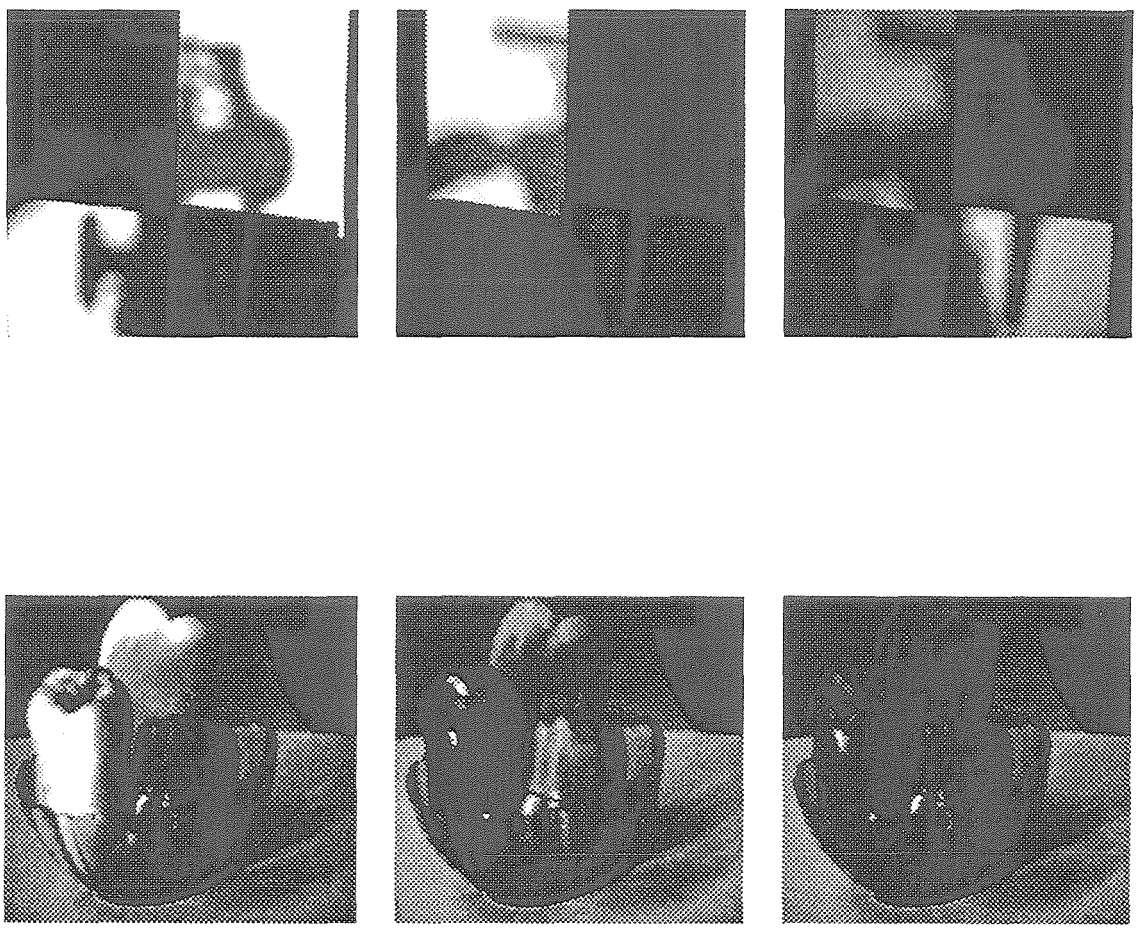


Figure 2.2: Red, green, and blue components of (a) Mondrian and (b) peppers of Fig. 2.1.

But, since Nrgb in Equation 2.2 is redundant (viz. $N_B = 1 - N_R - N_G$), the preferred normalized color space is typically formulated as (Hurlbert and Poggio, 1989; Kender, 1976; Nevatia, 1977):

$$\begin{aligned} Y &= c_1 R + c_2 G + c_3 B \\ T_1 &= \frac{R}{(R+G+B)} \\ T_2 &= \frac{G}{(R+G+B)} \end{aligned} \tag{2.3}$$

where c_1, c_2, and c_3 are chosen constants such that $c_1 + c_2 + c_3 = 1$. Y is interpreted as the image luminance of the image pixel and T_1 and T_2 are chromatic variables which are approximately independent of illumination (Nevatia, 1977). The Nrgb images of the Mondrian and peppers are shown in Fig. 2.3. When we compare this figure to Fig. 2.2 we note that illumination dependence is less for normalized color than for RGB images.

2.3 Opponent Color Spaces

Opponent color spaces can be considered as intermediate transforms of the RGB system to other systems. The simplicity of the linear operations from the opponency transformation is a plus[2]. However, the most important benefit is the construction of the principal components in the RGB color space. This has been used to advantage in applications such as computer vision and color television signal encoding as discussed below.

One of the originally proposed color opponent transform (see for example (Hurvich and Jameson, 1957)) that was promulgated by Hering was based on psychophysical

[2]This benefit is not directly exploited in our segmentation algorithms of Chapter 4 but is explicitly utilized in our hardware implementations of Chapter 5.

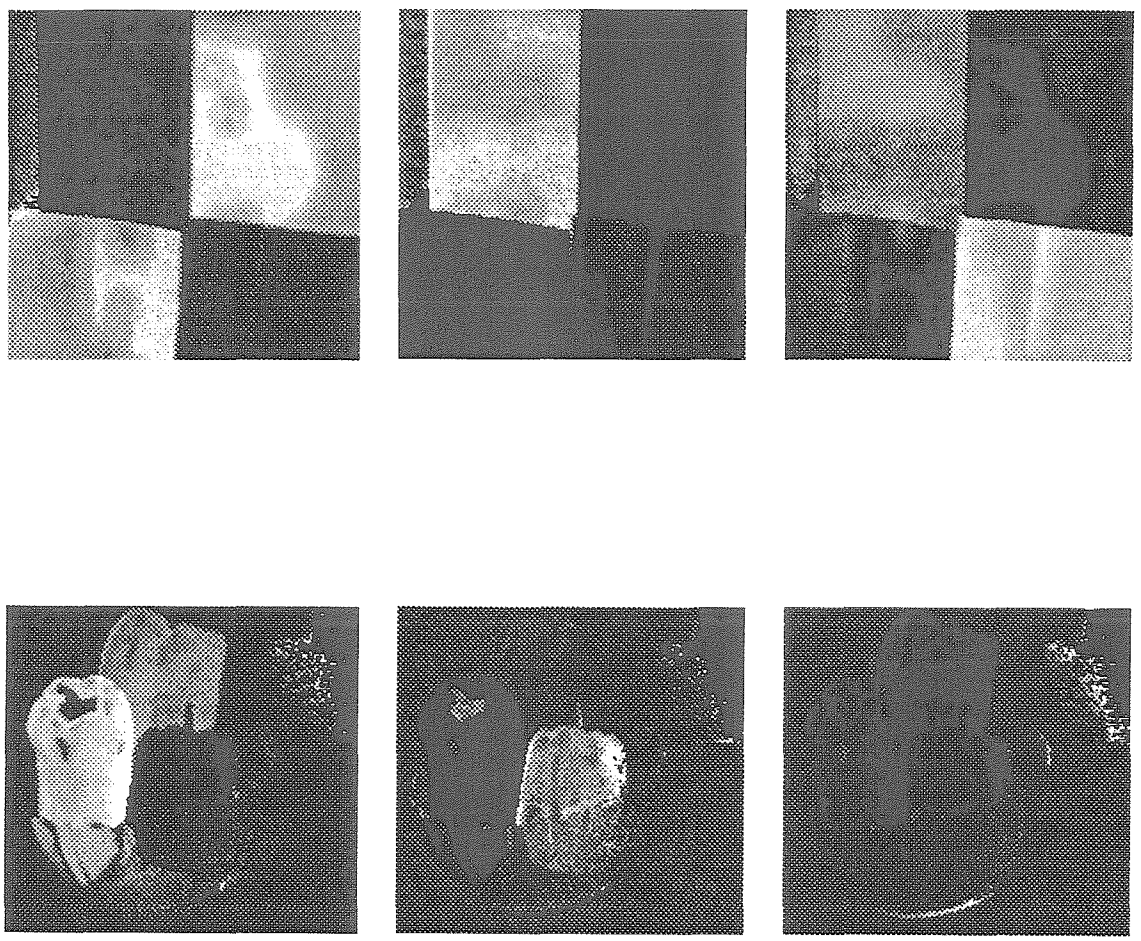


Figure 2.3: Normalized red, green, and blue components of (a) Mondrian and (b) peppers of Fig. 2.1.

observations and is given by

$$\begin{pmatrix} Wh - Bl \\ RG \\ BY \end{pmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & 2 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.4}$$

Usually, this transform matrix is modified with other values to account for psychophysical data. We present this equation both for historical reasons and as an introduction to other opponent color spaces. In the 1880's Hering qualitatively proposed the opponent system so that various psychophysical phenomena could be easily explained. Later in 1957, Hurvich and Jameson quantified and championed the opponent representation over the three component Young-Helmholtz RGB model. Eventually, we utilize a variant of this simple and elegant formulation of the opponent color space and construct high-level color variables directly within silicon hardware.

In an another study, an empirically derived opponent color space was derived based on the Karhunen Loeve transformation on RGB data of various images (Ohta et al., 1980). By analyzing more than 100 color features obtained during segmentation of eight kinds of color pictures, the following color features, ordered by their principal components, I_1 I_2 I_3, were found to be most effective.

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.5}$$

Another practical opponent color system is the YIQ system (Buchsbaum, 1987)

defined by

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.6}$$

This space is utilized in the optimum transmission of color television signals by the National Television Systems Committee (NTSC). The YIQ standard was designed to provide the best[3] possible color television signal within the existing monochromatic standard. Y stands for the luminance (It is the Y in the CIE XYZ) and is a major portion of the signal energy. At small portions of the signal energy, I, or in-phase, is psychophysically related to *red−cyan*, while Q, or quadrature, is related to *magenta−green*. Y is transmitted at a bandwidth of 4.2 Mhz, while I and Q are transmitted at 1.3 Mhz and 0.7 Mhz respectively. That the system works within the bandwidth constraints of the existing monochromatic standard can be attributed to exploiting human color encoding.

In general, opponent color systems are utilized to reduce the information content of the signal. For human color vision, Buchsbaum shows that efficient information transmission is achieved by transforming the initial RGB signal into one achromatic and two opponent chromatic channels which are all uncorrelated and have maximum variability. These three principal channels are computed from the Vos-Walraven human psychophysical response curves (Buchsbaum and Gottschalk, 1983) and have signal energy of 97.2%, 2.78%, and 0.015% respectively. For color TV images the typical signal distribution for YIQ is 93%, 5%, and 2% respectively. Thus, most of the information content of an image is in the achromatic signal.

[3]In this sense, best signal means best compromise for "standard" observers.

2.4 HSI Space

Of the many similar spaces that achieve Hue-Saturation-Intensity (HSI) characteristics, i.e., color ordering systems that are based on human color perception, the Munsell color system is remarkably popular (Wyszecki and Stiles, 1982). This system characterizes color in terms of Hue, Chroma (or Saturation), and Value (or Intensity) components and has been shown to achieve favorable hue segmentation (Tominaga, 1987).

Many transforms from RGB to HSI type spaces have been presented in the Computer Graphics (Foley et al., 1990; Joblove and Green, 1978; Smith, 1985) and Computer Vision (Gershon, 1985; Jain, 1989; Kender, 1976; Tominaga, 1987) literature. The transformation to HSI from RGB that is used in this study is currently implemented with special purpose digital hardware (Genz, 1990) and is given by:

$$\begin{aligned} Int &= \frac{(R+G+B)}{3} \\ Sat &= 1 - \frac{\min(R,G,B)}{Int} \\ Hue &= Arctan\left[\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right] \end{aligned} \tag{2.7}$$

where $Arctan(y/x)$ utilizes the signs of both y and x to determine the quadrant in which the resulting angle lies[4]. Generally, hue is thought of as the angle between a reference line and the color point in the RGB system. The physical model used to determine the hue angle is based on the diagram shown in Fig. 2.4a. If the R, G, B radial basis vectors are equally spaced $\frac{2}{3}\pi$ apart on the unit circle, the x and y component of an arbitrary point in the plane can be calculated from basic trigonometry

[4]Note that embedded within this hue definition are opponency calculations. Also note that hue becomes undefined at $G = R = B$. (See Section 3.3)

and are given by

$$\begin{aligned} x &= R - \frac{G+B}{2} = \frac{1}{2}[(R-G)+(R-B)] \\ y &= \frac{\sqrt{3}}{2}(G-B) \end{aligned}$$

This results in the hue angle shown in Equation 2.7. Conceptually, one can think of the HSI space as a cylindrical one (Fig. 2.4b), where the coordinates r, θ, z respectively correspond to saturation, hue, and intensity with $0 \leq r \leq 1$. The resulting HSI components for the Mondrian and the peppers are shown in Fig. 2.5[5]. Discussion of all these figures follows in the next chapter. For now, notice that the hue map of the Mondrian gives a false visual impression that there are different values within the upper color patch. Because of hue's modulo nature, the opposite is true, that "dark" and "light" hue values are very similar to each other. Notice also that the low saturation and low intensity background of the pepper image has inconstant hue values.

2.5 CIE Spaces

For completeness we describe the CIE "uniform" perceptual spaces that are involved in human color perception. The concept of "uniform perceptual distance" is an anthropomorphic one in that these spaces were set up so that traversals of a unit distance in any direction in the space is perceived by human observers to have the same "color difference." Computing the CIE representation of color involves a linear intermediate

[5]Also see Chapter 5 for a comparison of hue for human visual data and measurements from the color VLSI chip.

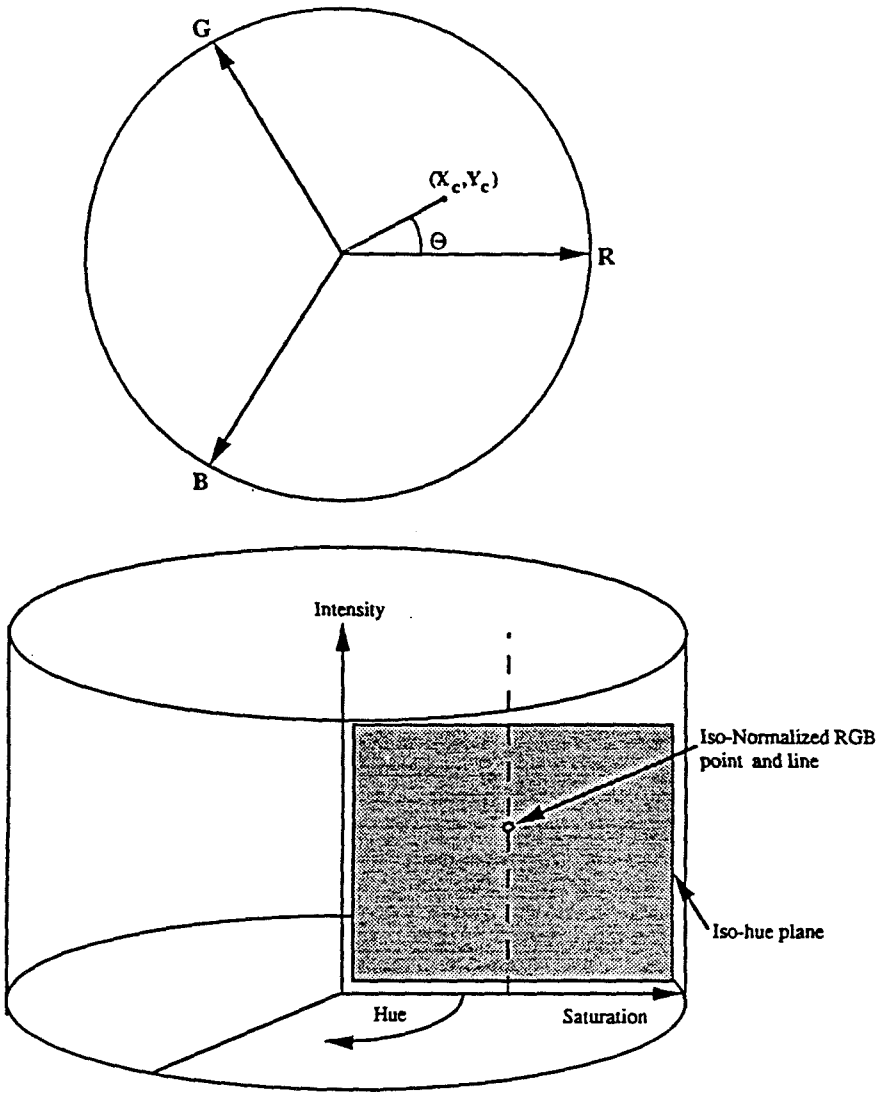


Figure 2.4: Details of hue space. (a) Physical model for simplified hue based on weighted average of RGB vectors. (b) Comparison of normalized color in HSI space. The non-tilted hue planar structure is a simplification of the dichromatic planar hypothesis in (Klinker et al., 1990) without the added complexity of color clustering and histogramming.

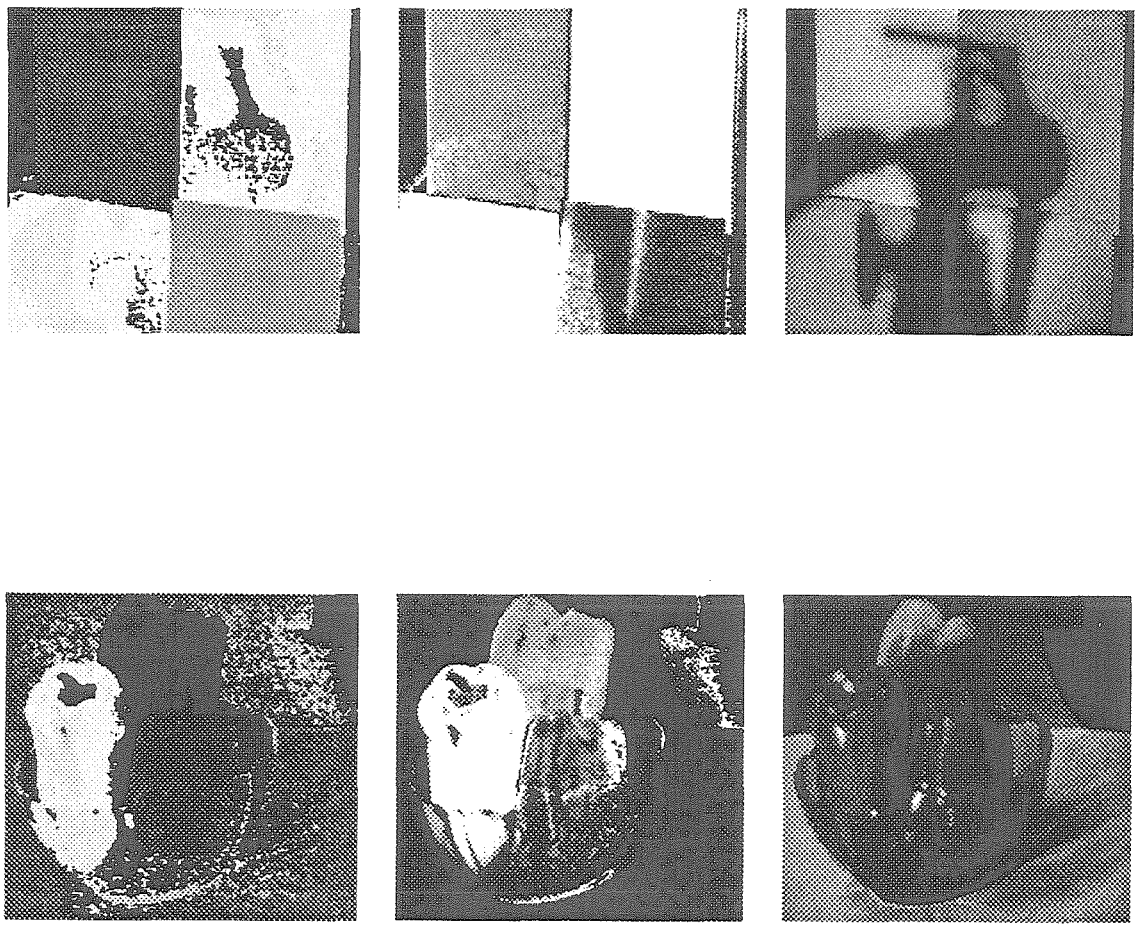


Figure 2.5: Hue, saturation, and intensity components from left to right of (a) Mondrian (top) and (b) peppers (bottom) of Fig. 2.1.

transformation, followed by a nonlinear transformation. First, the tristimulus values R, G, and B are transformed to another tristimulus set: X, Y, and Z.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.8}$$

For other arbitrary RGB sensor inputs, the transfer matrix must be determined empirically. In particular, the transform for the NTSC receiver primary system is

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.9}$$

Once the XYZ tristimulus coordinates are computed, a number of different CIE spaces can be constructed. One of these spaces, the CIE (L*u*v*) space can be described by

$$\begin{aligned} L^* &= 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 \\ u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \end{aligned} \tag{2.10}$$

Where

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z} \end{aligned}$$

and u'_n and v'_n are the nominally white object-color stimulus. Equation 2.10 is only valid for $Y/Y_n > 0.01$. If Y/Y_n is equal to or less than 0.008856 then

$$L^*_m = 903.3\frac{Y}{Y_n}$$

Another important CIE space, the CIE (L*a*b*) coordinates, is usually selected because it appears to have more uniform perceptual properties and gives better results than the CIE (L*u*v*) space in segmenting color pictures (Ohta et al., 1980). The CIE (L*a*b*) space is represented by

$$\begin{aligned} L^* &= 116\left(\frac{Y}{Y_n}\right)^{1/3} - 16 \\ a^* &= 500\left[\left(\frac{X}{X_n}\right)^{1/3} - \left(\frac{Y}{Y_n}\right)^{1/3}\right] \\ b^* &= 200\left[\left(\frac{Y}{Y_n}\right)^{1/3} - \left(\frac{Z}{Z_n}\right)^{1/3}\right] \end{aligned} \tag{2.11}$$

with the constraint that $X/X_n, Y/Y_n, Z/Z_n > 0.01$. Formulas that do not satisfy that constraint are given in (Wyszecki and Stiles, 1982).
Chroma for the L*u*v* and the L*a*b* spaces is respectively given by

$$\begin{aligned} C^*_{uv} &= \left[(u^*)^2 + (v^*)^2\right]^{1/2} \\ C^*_{ab} &= \left[(a^*)^2 + (b^*)^2\right]^{1/2} \end{aligned} \tag{2.12}$$

Similarly, hue is given by

$$hue_{uv} = Arctan\left[\frac{v^*}{u^*}\right], \text{ and } hue_{ab} = Arctan\left[\frac{b^*}{a^*}\right] \tag{2.13}$$

The CIE (L*a*b*) hue, chroma, and lightness images of the "Mondrian" and peppers are shown in Fig. 2.6.

Both CIE (L*u*v*) and CIE (L*a*b*) spaces require an intermediate transform to the XYZ system from the system dependent RGB system, and then either a normalization or a cube-root transformation. In comparison, the hue transformation given by Equation 2.7 is substantially simpler. This added complexity is not warranted since we will show that the simplified hue formulation gives satisfactory results. Furthermore, CIE spaces were developed for the psychophysical need to have perceptual uniformity for the standard human observer. We are concerned with the use of color segmentation for analog vision sensors and not in matching human perception.

When we compare the figures, we see that qualitatively CIE (L*a*b*) hue, chroma, and lightness are similar to HSI. We note that the hue measure is least correlated to intensity, that normalized color is somewhat correlated to intensity, and that RGB is strongly correlated to intensity. Noisy hue regions within images are indicative of low saturation or low intensity regions. Their treatment requires special consideration and is described in the following chapters.

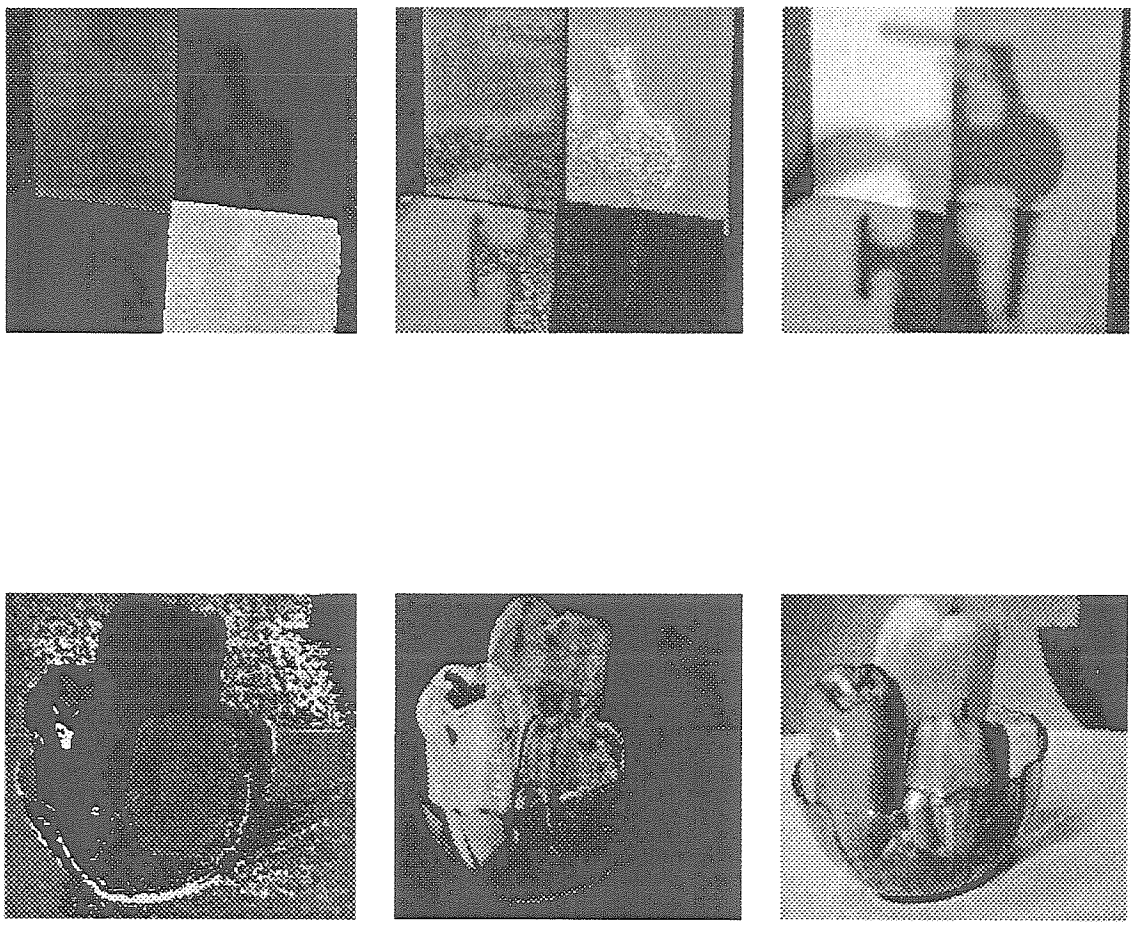


Figure 2.6: CIE (L*a*b*) hue, chroma, and lightness components of (a) Mondrian and (b) peppers of Fig. 2.1. RGB inputs are based on the NTSC receiver primary system. While hue and lightness in this figure are directly related to hue and intensity of Fig. 2.5, chroma is unrelated to saturation. Note also that the CIE hue angle is slightly shifted when compared to simplified hue of Fig. 2.5.

Chapter 3

Properties of Hue

In this chapter we describe the fundamental properties of hue and compare it to normalized color. We show some of its problems, and some of its advantages for locating material changes in images[1].

3.1 Additive/Shift and Multiplicative/Scale Invariance

Both Nrgb and HSI have the desirable property of *multiplicative/scale invariance*, that is, uniform variation of the tristimulus components will not change the measured quantity. This property is illustrated by the following relationships:

$$\begin{aligned} \text{hue}(R,G,B) &= \text{hue}(\alpha R, \alpha G, \alpha B) \\ N_C(R,G,B) &= N_C(\alpha R, \alpha G, \alpha B) \\ & \quad (\forall \alpha > 0) \end{aligned} \tag{3.1}$$

[1]Portions of this chapter have already been published (Perez and Koch, 1994).

These results can be verified by examining the constitutive Equation 2.2 and Equation 2.7.

Additionally, hue has *additive/shift invariance* which Nrgb lacks:

$$\begin{aligned} \text{hue}(R,G,B) &= \text{hue}(R+\beta, G+\beta, B+\beta) \\ N_C(R,G,B) &\neq N_C(R+\beta, G+\beta, B+\beta) \\ &\quad (\forall \beta \text{ such that } (R+\beta, G+\beta, B+\beta) \in [\Re > 0]^3) \end{aligned} \tag{3.2}$$

This property of *additive/shift invariance* gives iso-hues a greater span in the color space. The comparison shown in Fig. 2.4b between normalized color and hue in the HSI space illustrates that iso-hues occupy plane segments while iso-Nrgb occupy line segments. An equivalent statement to Equation 3.2 is that hue is invariant under saturation changes, while normalized RGB is not. Another interpretation is that Hue is invariant to white color vector additions. (We generalize this to the *Integrated White* condition in Section 3.4.2.) Originally, these *multiplicative/scale* and *additive/shift* invariance properties were used by (Kender, 1976) to analyze hue instabilities. In this study, we show their advantages.

3.2 Hue in CIE Spaces

Similarly, hue in the CIE (L*a*b*) space defined by Equation 2.13 show *multiplicative/scale* invariance. This property follows from a straightforward evaluation of Equation 2.13 in the invariance relationship of Equation 3.1. However, the CIE hue lacks strict *additive/shift* invariance. It only approximates *additive/shift* invariance due to its non-linear cube-root transformation and normalization. Our approach is to avoid the CIE color spaces because of their added complexity in matching human perception.

3.3 Problem of Singularities

Kender's 1976 study of the properties of nonlinear color transforms revealed certain problems associated with the use of HSI space. In particular, the HSI transform has the unfortunate property of an unremovable singularity at the axis of the color cylinder, where R=G=B (Saturation = 0). We can see this by examining the hues of totally saturated pixels along the red-green segment of the color wheel, i.e., $hue\{R = (1-r)x; G = rx; B = 0\}$, where $0 \leq r \leq 1$. This situation corresponds to $Int = \frac{x}{3}$ and $Sat = 1$). It then follows from Equation 2.7:

$$\lim_{x \to 0} hue\,\{(1-r)x, rx, 0)\} = \arctan(\frac{\sqrt{3}r}{2-3r})$$

Thus, hue varies continuously from 0 (when r=0) to $2\pi/3$ (when r=1). In addition, hue near its singularities is intrinsically unstable. Take for example, pixels whose values are $hue\{x, x, 0\}$, where $x \geq 0$. For 8 bit digital implementation (i.e., hue $\in$ [0..255]) a minimal digital perturbation gives $hue\{x+1, x, 0\}$ which can result in changes up to $\pi/3$. This characteristic *noise* is prominent in the hue images in Fig. 2.5 (and in the hue edges shown in Fig. 4.3). Thus, the mapping from RGB to HSI is ill-conditioned near the central axis (low saturation) and ill-posed at the axis (zero saturation). Consequently, hue values will vary dramatically from one location to the next within black or white areas within an image.

The Kender study went as far as recommending not to use nonlinear color transforms such as HSI and normalized color spaces but to use linear transforms such as YIQ and opponent color spaces. This recommendation has been heeded in the applied computer vision literature (Barth et al., 1986; Ohta et al., 1980).

3.4 Hue Discounts Confounding Cues

In this section, we show that assuming reasonable material properties and lighting conditions, hue information will discount intensity edges due to transparency, highlights, shading, and shadowing.

3.4.1 Discounting Transparency

The first question that comes to mind: why is discounting transparency important? Possibly the underwater medium could be transparent. But, when we examine the data for pure water, the absorption coefficient (α) exhibits varying attenuation throughout the visible range. In particular, the data shows (Wolfe and Zissis, 1985)

λ,μm	0.40	0.44	0.50	0.54	0.60	0.65	0.70
α, m^{-1}	0.072	0.023	0.016	0.024	0.125	0.210	0.840
τ	0.931	0.977	0.984	0.976	0.882	0.811	0.432

where the transmission coefficient, τ, at depth, x= 1 meter is calculated from

$$\tau(\lambda) = \exp(-\alpha(\lambda)x)$$

Although we use the data for pure water, the attenuation coefficients for other types of waters, such as *deep oceanic* and *clear coastal* waters, also vary within the same spectral range (Wolfe and Zissis, 1985).

For small distances through the water environment, the uniform spectral transmission assumption is valid. Additionally, a reasonable assumption for an object is that its material points are physically close to each other, thus, the transparency effects for viewing objects in aqueous environments will be negligible because relative transmission distances are small. In general, the tristimulus equation with and

without an absorbing media is given by the following:

$$\begin{aligned} X_k &= \int_\lambda E(\lambda)S_k(\lambda)d\lambda \\ XT1_k &= \int_\lambda E(\lambda)S_k(\lambda)\tau^{D1}(\lambda)d\lambda \\ XT2_k &= \int_\lambda E(\lambda)S_k(\lambda)\tau^{D2}(\lambda)d\lambda \end{aligned} \tag{3.3}$$

where $X_k, k = R, G, B$ are the tristimulus values, $XT1_k$ and $XT2_k$ are tristimulus values through transparent medium, $E(\lambda)$ is the incoming light intensity defined as in Equation 2.2, S_k are the three hypothetical color filters, $\tau(\lambda)$ is the transmittance per unit distance in the medium, and $D1$ and $D2$ are the distances in the medium for points 1 and 2.

For a spectrally uniform transmitting medium,

$$XT_k = (\tau_0)^D \int_\lambda E(\lambda)S_k(\lambda)d\lambda = (\tau_0)^D X_k$$

This case has the property of multiplicative scale invariance. Evaluating Equation 2.2 for normalized color and Equation 2.7 for hue results in:

$$\begin{aligned} Hue(X_R, X_G, X_B) &= Hue(XT_R, XT_G, XT_B) \\ nRGB(X_R, X_G, X_B) &= nRGB(XT_R, XT_G, XT_B) \end{aligned}$$

Both normalized colors and hue discount image transparency through a spectrally uniform medium.

3.4.2 Discounting Highlights and the Integrated White Condition

In this section we describe the Phong shading model, how it accounts for highlights and how utilizing hue can discount highlights. A similar analysis in Appendix A on other advanced models of shading such as the Cook-Torrance Model and the Dichromatic Reflection model gives comparable results.

The Phong shading model (Phong, 1975) is based on the empirical observation that the radiance from a highlight reflection falls off sharply with increasing α, where α is the angle between the reflection vector and the viewpoint vector as shown in Fig. 1.2. The form adopted by Phong was $\cos^n(\alpha)$, where n varies from 1 to 200 depending on the surface. The full equation for Phong shading is given by:

$$I_C = I_{aC}k_{aC} + \frac{I_{pC}}{r+k}[k_{dC}\cos(\theta) + k_s\cos^n(\alpha)] \quad \text{for } C = (R,G,B) \tag{3.4}$$

where, I_a and k_a are the ambient intensity and ambient reflection coefficient. I_p is the intensity of a point light source, r is the distance from the perspective viewpoint to the surface, k is a constant, k_d is the diffuse reflection coefficient, and k_s is a specular reflection coefficient. Both reflection coefficients k_a and k_d have three components for the tristimulus equations and assume a constant value between 0 and 1. In the Phong model the specular reflection coefficient, k_s, is assumed to be constant and independent of surface color. The first two terms in Equation 3.4 represent the non-highlight condition, while the third term contains the highlight contribution.

Explicitly, for the non-highlighted (NH) region, the equivalent RGB tristimulus values become

$$C_{NH} = I_{aC}k_{aC} + \frac{I_{pC}}{r+k}[k_{dC}\cos(\theta)] \quad \text{for } C = (R,G,B) \tag{3.5}$$

Then, if we assume a white light source ($I_{pR} = I_{pG} = I_{pB}$), the equivalent RGB tristimulus for the highlighted (H) region becomes

$$C_H = C_{NH} + \overbrace{\frac{I_{pC}}{r+k}[k_s \cos^n(\alpha)]}^{\beta} = C_{NH} + \beta \quad \text{for } C = (R, G, B) \tag{3.6}$$

We immediately see that this equation satisfies the *additive/shift invariance* condition. Thus, a computation based on hue will discount highlights due to the Phong shading model, while a computation on normalized color, in general, will not. An exception is the condition that the ambient illumination is proportional to the source illumination (i.e., $I_a(\lambda) = cI_p(\lambda)$). Now *multiplicative/scale invariance* holds and both normalized color and hue will discount highlights.

The Phong shading model has been criticized within the computer graphics community because it gives an object a "plastic" appearance. Ironically, the figures used in the study by Klinker, Shafer, and Kanade (Klinker et al., 1988; Klinker et al., 1990) were composed of plastic objects. Furthermore, a simple hue transform on the image in Fig. 3.1 gives accurate segmentation while discounting highlights. The measurements for a horizontal slice through two highlight regions in the subsampled image are shown in the following table.

Top left doughnut (y=36)				Orange highlit cup (y=57)			
x	Hue	Sat	Intensity	x	Hue	Sat	Intensity
33	3	126	77	63	7	73	109
34	2	98	88	64	2	19	196
35	0	56	110	65	225	6	245
36	253	57	111	66	249	36	163
37	1	111	78	67	4	64	117
38	3	140	71	68	6	71	112
% span	2.4	32.9	15.7	% span	14.9	27.1	53.3

For the central orange cup, a typical slice through the dominant highlight shows that hue varies by 14.9 percent while intensity varies by 53.3 percent of the total range. As we cross the highlight, hue changes 9.4 percent while intensity changes by 32.2 percent[2].

The hue map shown in Fig. 3.1 discounts most of the highlights in the intensity map. The reason for this is that typical plastic has embedded pigment particles within a substrate that is transparent or white (Cook and Torrance, 1981). Thus, light reflected off plastics will have a diffuse colored component and a specular white component.

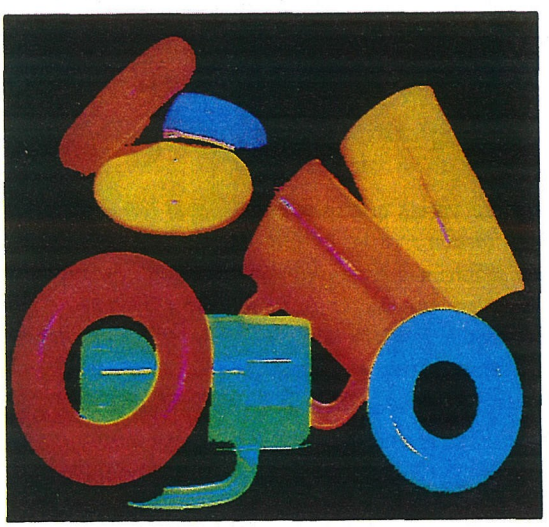
If we generalize the Phong shading model by assuming that the reflectance coefficients are not constant but vary with wavelength we have:

$$I(\lambda) = I_a(\lambda)k_a(\lambda) + \frac{I_p(\lambda)}{r+k}[k_d(\lambda)\cos(\theta) + k_s\cos^n(\alpha)] \tag{3.7}$$

This implies that from Equation 2.1 the tristimulus values in the nonhighlighted

[2]Note that due to the wrap-around nature of hue, low values of hue are near high ones. For example, a hue value of 0 is 3 units removed from 253.

(a)



(b)

Figure 3.1: (a) Plastic Objects with highlights (Klinker et al., 1988; Klinker et al., 1990), and (b) hue transform of the same image for intensities above 20/255. Note that highlights are removed for most of the plastic objects. Although not considered in this paper, we clearly see yellow interreflection on the blue doughnut within the hue space.

region becomes

$$C_{NH} = \int_\lambda I_a(\lambda)k_a(\lambda)S_C(\lambda)d\lambda + \frac{\cos(\theta)}{r+k}\int_\lambda I_p(\lambda)k_d(\lambda)S_C(\lambda)d\lambda \quad \text{for } C = (R,G,B) \tag{3.8}$$

For the highlighted region the equivalent tristimulus values become

$$C_H = C_{NH} + \frac{k_s\cos^n(\alpha)}{r+k}\int_\lambda I_p(\lambda)S_C(\lambda)d\lambda \quad \text{for } C = (R,G,B) \tag{3.9}$$

We see that hue invariance due to highlights will work exactly if and only if the *integrated white* condition[3] holds, namely if

$$\int_\lambda I_p(\lambda)S_R(\lambda)d\lambda = \int_\lambda I_p(\lambda)S_G(\lambda)d\lambda = \int_\lambda I_p(\lambda)S_B(\lambda)d\lambda \tag{3.10}$$

This follows from an inspection of Equation 3.9 and the fact that the *additive/shift invariance* condition of Equation 3.2 is reached. The *multiplicative/scale invariance* condition of Equation 3.1 can not in general be reached for the generalized Phong shading model. The assumption of ambient illumination being proportional to source illumination will not simplify matters much.

But, how reasonable is the *integrated white* assumption? For white light illumination, $I_p(\lambda)$ is independent of λ. This implies that the *integrated white* condition simplifies to

$$\int_\lambda S_R(\lambda)d\lambda = \int_\lambda S_G(\lambda)d\lambda = \int_\lambda S_B(\lambda)d\lambda \tag{3.11}$$

For artificial systems, this feature may be designed in with carefully selected spectral filters. Additional compensation of non-white illumination to a white standard

[3]Bajcsy (1990) utilizes a similar concept called *illumination whitening.*

perhaps utilizing "color constancy" techniques is also necessary. Nevertheless, the general *integrated white* condition for non-uniform spectral intensities requires that Equation 3.10 must hold[4] in order for hue to discount highlights completely. Anything less will give an approximate hue invariance to highlights. Within the framework of the Phong shading model, normalized color will not generally discount highlights.

3.4.3 Discounting Shading

Shading or surface orientation changes is another confounding cue that will confuse an achromatic vision system. A simple analysis by Rubin and Richards (1985) for a single light source situation shows that

$$\begin{aligned} I_1 &= \rho(\lambda) I_p(\lambda) \cos(\theta_1) \\ I_2 &= \rho(\lambda) I_p(\lambda) \cos(\theta_2) \end{aligned} \tag{3.12}$$

where I_1 and I_2 are the image radiance intensities for two points that differ only by a surface orientation change, $\rho(\lambda)$ is the spectral albedo (will be similar to the bidirectional reflectance of a lambertian surface), and θ is the angle between the surface normal and the illumination direction. The tristimulus equations (Equation 2.1) immediately implies that the RGB ratios for these two points become

$$\frac{R_1}{R_2} = \frac{\cos(\theta_1)}{\cos(\theta_2)}, \qquad \frac{G_1}{G_2} = \frac{\cos(\theta_1)}{\cos(\theta_2)}, \qquad \frac{B_1}{B_2} = \frac{\cos(\theta_1)}{\cos(\theta_2)} \tag{3.13}$$

For this simple model, the *multiplicative/scale invariance* condition holds. Thus, both

[4]This is viable in industrial applications where the illumination and the filter characteristics are controlled such that the conditions of Equation 3.10 are satisfied.

normalized color and hue will discount this model of shading.

If ambient light were added to the model, the *integrated white* condition (such as those defined by Equation 3.10) must hold for hue to discount confounding shading cues[5]. The analysis for this is similar to the above calculations for highlight invariance for the Phong shading model and the Cook-Torrance model. The Dichromatic Reflection model requires a slight modification of the addition of an ambient illumination term.

3.4.4 Discounting Shadowing

For analyzing shadowed regions, we require two sources of illumination: one to cast the shadow on an object and the other to lightly illuminate the darkened region. In practice, the second light source often results from interreflections or scattering from the principal light source. Our starting point is taken from Rubin and Richards' analysis (1985) for shadowed regions. (See also (Gershon, 1985) for a discussion on distinguishing material from shadow boundaries.) The governing equation for this simple model of an image point with and without a shadow is given by

$$\begin{aligned} I_{lit} &= \rho(\lambda)[I_s(\lambda) + I_d(\lambda)] \\ I_{shade} &= \rho(\lambda)[I_d(\lambda)] \end{aligned} \tag{3.14}$$

where I_{lit} and I_{shade} are the radiance intensities in the non-shadowed and shadowed region respectively, ρ is the albedo as defined before, I_p is the illumination intensity. and I_d is the diffuse ambient intensity. The relevant tristimulus equations for the

[5]This cue can also be discounted when the ambient light differs from the direct light by an additive white light shift (which is approximately true for natural illumination). This is discussed in section 3.4.4 below.

shade region becomes

$$C_{shade} = \int_{\lambda} I_d(\lambda)\rho(\lambda)S_C(\lambda)d\lambda \quad \text{for } C = (R, G, B) \tag{3.15}$$

While the relevant tristimulus equations for the lit region become

$$C_{lit} = C_{shade} + \int_{\lambda} I_s(\lambda)\rho(\lambda)S_C(\lambda)d\lambda \quad \text{for } C = (R, G, B) \tag{3.16}$$

In the most general case, hue will discount shadows if and only if

$$\int_{\lambda} I_s(\lambda)\rho(\lambda)S_R(\lambda)d\lambda = \int_{\lambda} I_s(\lambda)\rho(\lambda)S_G(\lambda)d\lambda = \int_{\lambda} I_s(\lambda)\rho(\lambda)S_B(\lambda)d\lambda$$

This equation seems coincident with the *integrated white* condition of Equation 3.10. However, this condition depends on the surface albedo $\rho(\lambda)$ which depends on material type[6].

A further simplification is possible if we assume that the ambient lighting and the main lighting are related. In particular, if the ambient lighting tracks the main lighting, that is $I_p(\lambda) = \alpha I_d(\lambda)$, then the following tristimulus equations hold

$$C_{lit} = (\alpha + 1)C_{shade} \quad \text{for } C = (R, G, B)$$

For this case, the condition of *multiplicative/scale invariance* holds. Thus, both normalized color and hue will discount this particular shadow cue. In the following table, we compare color values for typical points in the light and shadow regions of the lower

[6]The implication of this equation is that the diffuse illumination can be arbitrarily colored and not violate the *integrated white* condition. Segmentation will be achieved under these conditions.

right hand side of the Mondrian of Fig. 2.1a within different color spaces.

Color space	Δ light to shade		
R:G:B	80%	92%	68%
nR:G:B	16%	68%	24%
H:S:I	3%	55%	76%
(CIE)H:C:L	4%	11%	52%

Hue varies only 3% across the shadow edge while normalized color varies at least 16%. The performance of the CIE (L*a*b*) hue is similar to the simplified hue (for more details, see Chapter 4).

In most cases, however, $I_p(\lambda) \neq \alpha I_d(\lambda)$. The work of (Gershon et al., 1986) provide a complementary analysis for discounting shadow boundaries utilizing Rubin and Richards' spectral crosspoint concept. At the heart of the issue: intensities from within shadow regions will be illuminated by ambient sources that will not have the same spectral characteristics as the primary source. The example of natural illumination shown in Fig. 1.4 illustrates this clearly. First, there is a color shift in the ambient illumination when compared to the direct illumination. Second, there is an *additive shift* component to the overall ambient spectral radiance from different portions of the sky. Although complicating the analysis, the *additive shift* attribute of natural illumination complements the hue representation and supports discounting confounding cues cause by shadowing and shading[7]. The measured spectral radiance of green paper shown in Fig. 3.2, which also includes shadow and shading effects, supports the hypothesis that hue is conserved. Calculated values from the figure show that although the radiance value can shift by 90% the hue shifts by no more than 20 degrees or 5.6%

[7]This type of illumination will affect the measured RGB values from objects with different orientations, but will be discounted in hue

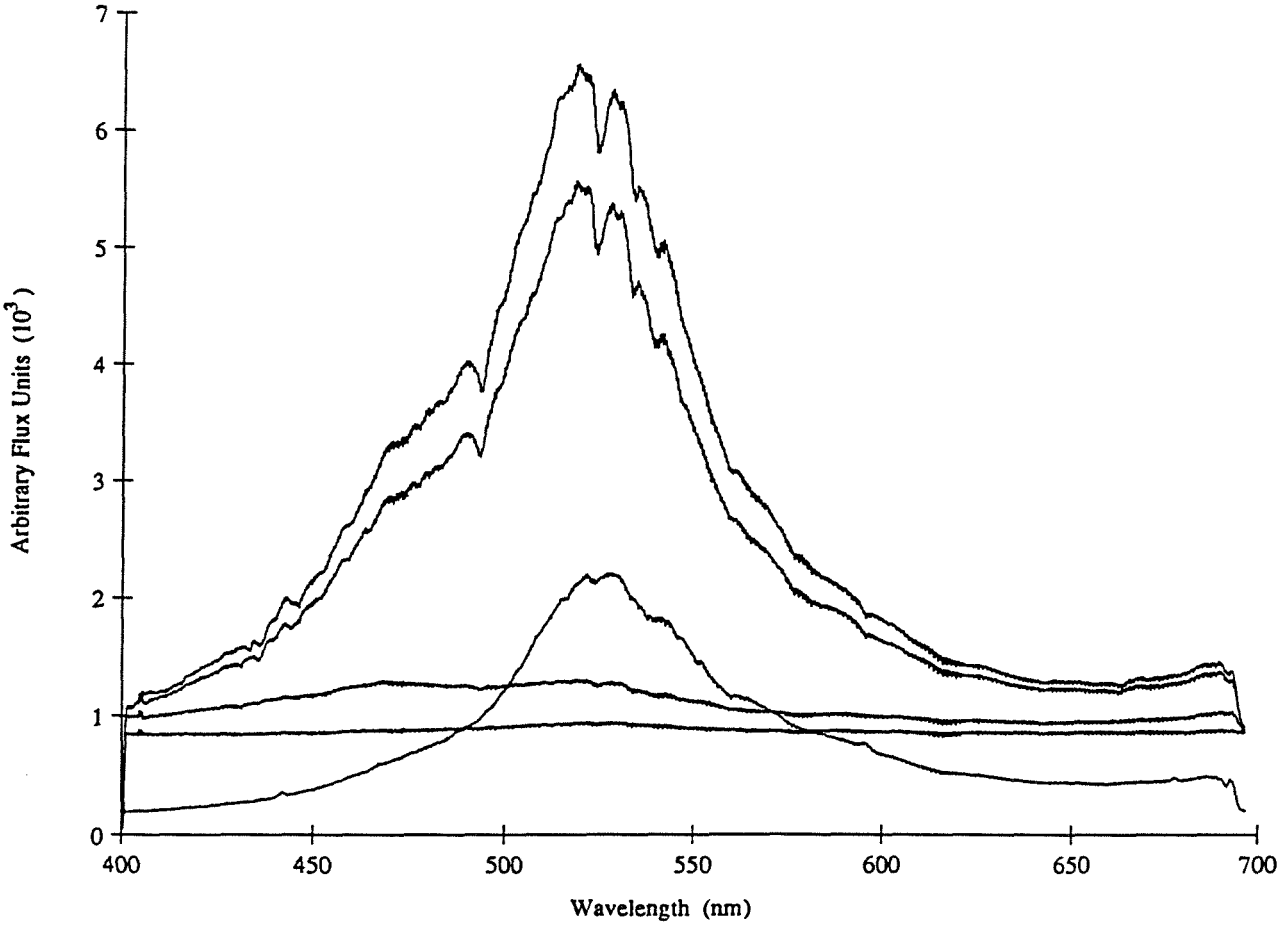


Figure 3.2: Radiance measured of green paper under natural illumination under varying orientations (shading) and shadowing. Shading changes cause an approximate linear shift in spectral intensity. Shadows have both *multiplicative scale* and *additive shift* changes in intensity. Calculated hue values for the shadow region differ by 16 to 20 degrees from the non-shadowed region.

3.5 Rubin and Richards Representation

Rubin and Richards (1984) describe a method of determining material boundaries by comparing the spectral slopes at candidate positions in an image. For a two sensor color system—say R and G—they conclude that no single continuous function of R and G will be invariant under multiplicative (shading or surface orientation and approximate shadows), exponential (pigment density), and additive (highlights) changes[8]. Therefore, they concentrate on one invariance property—multiplicative change—to facilitate discounting shadow and shading cues. For this particular case, they examine three candidate functions:

$$F_1(x) = \frac{R(x)}{G(x)}, \qquad F_2(x) = \frac{R(x)}{R(x)+G(x)}, \qquad F_3(x) = \frac{R(x)-G(x)}{R(x)+G(x)}$$

where x is a spatial location in the image

F_1 is discarded because it maps image regions onto the unbounded interval $[0, \infty]$, while F_2 and F_3 maps onto the closed intervals $[0, 1]$ and $[-1, 1]$ respectively. F_3 is selected because it has a simple condition for determining material boundaries, namely, the opposite slope condition:

$$Sign[F_3(x_1)] \neq Sign[F_3(x_2)]$$

This condition implies that the material at position x_1 differs from the material at x_2. In any case, both F_2 and F_3 can be considered respectively as normalized color and differences of normalized color. We know this will only account for *multiplicative*

[8]Fortuitously, the construction of hue in a trichromatic system is invariant to both multiplicative and additive changes. Constructing a function that includes invariance to exponential changes is more difficult and left as future work.

scale changes within an image.

For trichromacy, Rubin and Richards note that six basic material types are possible due to the 3 slope sign combinations. For a three sensor system—say R, G, and B—these become $(B-G, G-R, B-R)$. This information can be summarized in the two-dimensional plot reproduced in Fig. 3.3. The critical elements are the quadrant boundaries in the $(\frac{R-G}{R+G}, \frac{B-G}{B+G})$ space and the slope = 1 line in that space, which divides the material zones into 6 distinct regions. Image points that cross these boundaries signify material differences.

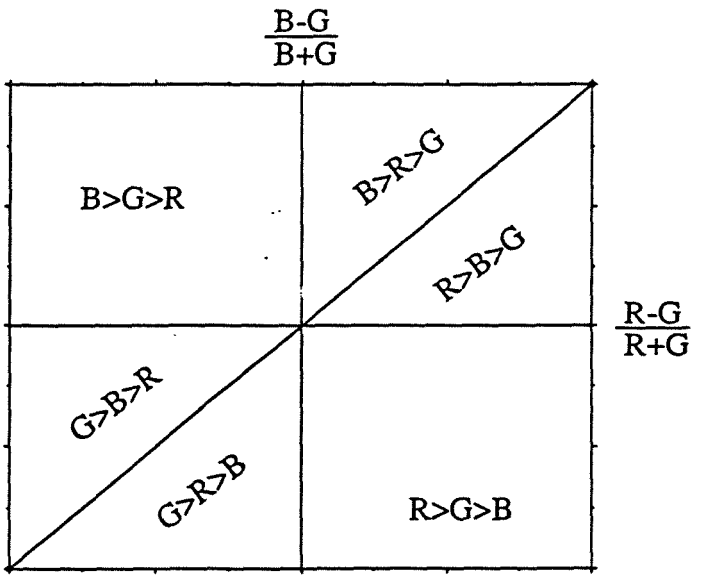


Figure 3.3: Rubin and Richards' (1986) material representation in trichromatic space. The axis lines and the unit slope line divide the color space into six regions, which correspond to 6 different material types.

In comparison, the candidate function we utilize in this thesis is the hue representation given by

$$Hue = \arctan(X), \qquad X = \frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}$$

The hue argument, X, functionally gives invariance to both multiplicative and additive changes. Furthermore, even though X maps onto $[0, \infty]$, the nonlinear "arctan" function compresses X to the range $(0, 2\pi)$. An added benefit is the wrap-around nature afforded by the "arctan" function, which is a measured characteristic of the psychophysical variable hue. Since the criterion for material changes is based on threshold conditions, Rubin and Richards' representation is a discrete valued variable (of six values). Our representation of material differences is hue, a continuous variable which has no set boundaries and offers higher resolution.

3.6 Physical Basis for Hue's Modulo Nature

Within Maloney's (1986) analysis of surface spectral estimation by a few parameters, he summarizes the physical basis for the properties of materials. We further summarize his work and justify the circular nature of hue, which has its basis in the physics and chemistry of materials.

For simple organic and inorganic molecules, light interaction within the visible spectrum results in a smooth broad shaped spectral absorptance curve[9] whose half-width size is on the order of 5,000 cm^{-1}, and caused primarily by low energy electronic transitions(Kauzmann, 1957; Nassau, 1983; Suzuki, 1967) [10]. The fact that the spectral curves are not spiky at these electronic transitions is due to electronic and molecular interactions which smooth out the discrete transitions of the high fre-

[9]For opaque materials, absorptance is the complement of reflectance. Conceptually, peaks and troughs are interchanged when switching from absorptance curves to reflectance curves.

[10]The conversion from wavenumber units to wavelength is $\bar{\nu}[cm^{-1}] = 10,000,000/\lambda[nm]$.

quency vibrational and rotational molecular transition energies. From a quantum physics standpoint, the low energy transitions are regularly repeated as the frequency space is traversed.

Since the energy separation of different electronic states is of the order of 10,000 cm^{-1} and since the width of the visible spectrum is 10,000 cm^{-1}, there can be a maximum of only one spectral absorptance peak within the visible range for this simple material. To first order, this peak absorptance corresponds to hue directly.

The concept illustrated in Fig. 3.4 shows the effect of shifting the position of the peak absorptance transitions thus resulting in a changing perception of color. Nevertheless, as the peak absorptance is shifted through the visible spectrum toward the left, for example, the tail of the next peak absorptance (nearby energy transition) appears toward the right of the spectrum. Further shifting will result in a repetition of the spectral pattern in the visible window. Thus, the hue variable will be cyclical as the absorption curve traverses the spectrum.

The measurement of spectral position for absorptance curves caused by electronic energy transitions requires, at a minimum, the use of 3 sensors to determine the relative placement (left, centered, or right) in the visible spectrum. Because molecular vibration energy transitions have a factor of 1/3 to 1/10 the separation of electronic energy states, a minimum of 9 spectral sensors are required to discriminate them. In the extreme, the discrimination of molecular rotational energy transitions, which have energy separations on the order of 1 to 100 cm^{-1}, require 100 to 1000s of spectral sensors.

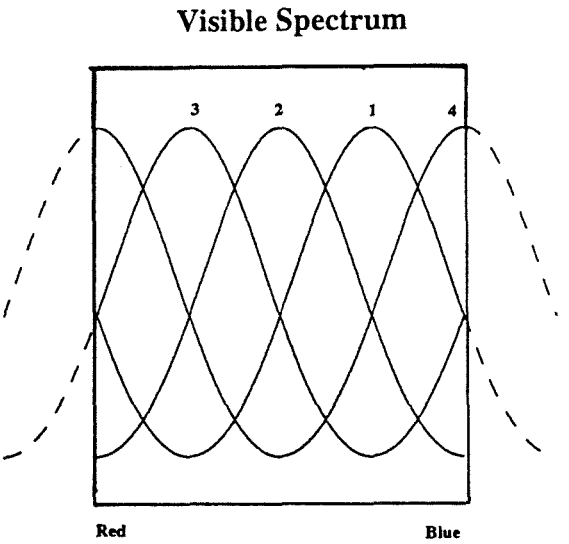


Figure 3.4: Absorption band of 4 different materials in relation to the visible spectrum. Each materials absorption band has energy spacings of 10,000 cm^{-1} and half-widths of 5,000 cm^{-1}. For different materials the absorption band locations can be shifted. In this example, materials 1, 2, 3, and 4 correspond to red, purple, blue, and green objects respectively.

Chapter 4

Algorithms for Hue Segmentation

Before we can utilize hue in image segmentation, we need to develop special tools to detect edges in hue space. Furthermore, because of the instabilities of hue in low saturation and low intensity regions (as described in Section 3.3 and indicated in Figs. 2.5 and 2.6), we need to temper our hue edge detection algorithm with some measure of "goodness." The natural candidate is the Markov Random Field (MRF) regularization technique which uses line discontinuities, when appropriate, to segment images. The adoption of the MRF regularization method will lead to improved color segmentation performance.

In this chapter we only present novel results. For edge detection, we assume the background of (Marr and Hildreth, 1980; Canny, 1986). For regularization and the MRF formulation we assume the background of (Geiger and Girosi, 1990; Geman and Geman, 1984; Poggio and Koch, 1985; Terzopoulos, 1986)[1].

[1]Portions of this chapter have already been published (Perez and Koch, 1994).

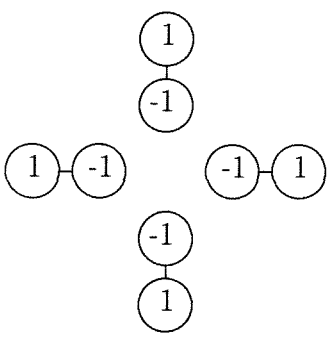
4.1 Finding Edges of a Circular Variable

Standard techniques of intensity edge detection (Canny, 1986) by convolving with various masks followed by thresholding or extremum detection such as by Canny's algorithm do not work with hue, because unlike intensity, hue is defined on the ring $\mathcal{S}^1$ rather than on the interval $\Re^1$. However, standard convolution techniques can be modified to work with these modulo variables to determine their spatial "edges." Conceptually, this is similar to a 2-D lattice populated with particles whose state is defined by a scalar variable defined on $\mathcal{S}^1$, such as orientation or phase angle θ. In the latter example, "edges" would correspond to all locations on the 2-D lattice across which the phase angle changes maximally.

Traditional edge detection entails finding the *zero-crossings* of the 2-D image convolved with a Laplacian. One of the simplest discrete approximations of a Laplacian is the following kernel:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

In the manner of Terzopoulos, this kernel can be decomposed into the following computational molecules (Dahlquist and Bjorck, 1974; Terzopoulos, 1985):



The individual computational molecules represent the distance measure between nearest pixel neighbors. For linear operations the centroid of the computational molecules can be summed and combined to give the Laplacian kernel. For non-linear modulo operations, this consolidation is incorrect. Nevertheless, it is this formulation that allows us to generalize the modulo Laplacian as the modulo distance of nearest neighbors from a central pixel. Modulo distance or angular distance, $\Phi(y-x)$, can be represented functionally as a linear saw tooth pattern with slope 1 and period 2π. It is given by the Fourier series expansion

$$\Phi(y-x) = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin[n(y-x)] \tag{4.1}$$

An equivalent form of this equation is given by

$$\Phi(y-x) = (y-x) - 2\pi k \tag{4.2}$$

where $k = \text{Floor}\left[\frac{1}{2}\text{Ceiling}[\frac{y-x}{\pi}]\right]$. For example, if $y - x = \pi + \delta$ with δ small, then $k = \text{Floor}\left[\frac{1}{2}\ \text{Ceiling}\ [1 + \frac{\delta}{\pi}]\right] = \text{Floor}\left[\frac{2}{2}\right] = 1$. Thus, $\Phi(\pi+\delta) = \pi + \delta - 2\pi = \delta - \pi$.

Since the hue values x, y are given in the principal range 0 to 2π, the angular distance formula can be simplified to

$$\Phi(y-x) = y - x + \begin{cases} +2\pi & \text{if } y - x < -\pi \\ -2\pi & \text{if } y - x > \pi \\ 0 & \text{otherwise,} \end{cases} \tag{4.3}$$

where positive values of $\Phi(y-x)$ indicate that y is positioned clockwise to x[2]. Therefore, the strategy for finding material boundaries in an image is to (1) convolve the hue image array with the modulo Laplacian, defined by the decomposed Laplacian kernel and the distance measure of Equation 4.3, and (2) identify the hue *zero crossings.*

We can compare our approach to Yoon (1991) who developes a mathematically elegant but numerically extensive algorithm in hue-saturation space. His method requires evaluating geodesics on the hue-saturation manifold to cluster image values about pre-selected cluster centers. Yoon's distance measure, which is used to mark hue-saturation cluster boundaries, requires solutions to a two-variable non-linear equation. In contrast, we can explicitly state the distance measure in our simplified hue space as Equation 4.3.

4.2 Performance Comparisons

In this section we compare conventional edge detection in the RGB and Normalized color spaces to the non-linear operator developed for hue space. A quantitative measure of performance is also computed.

4.2.1 Conventional Edge Detection

Edge detection is performed utilizing Canny's (1986) algorithm which is remarkably robust with noisy images. It achieves this by using an optimal detector that maximizes detection and localization performances with certain classes of edges. Canny's algorithm assigns edges to maxima in the gradient magnitude of a Gaussian-smoothed image.

[2]Appendix B presents a digital implementation of this algorithm that does not rely on conditional operations.

Fig. 4.1a shows the edges obtained from the RGB image of the Mondrian shown in Fig. 2.2a, while Fig. 4.1 (b) shows the edges obtained from the RGB image of the peppers shown in Fig. 2.2b. The edges obtained represent intensity changes, and thus, falsely indicate the shadow as a material boundary.

Notice that the edges shown in Fig. 4.2 for the normalized RGB image of Fig. 2.3 also show segmentation at the shadow edge. The reason for this has been discussed previously, and is a result of normalized color only factoring out uniform changes in color—what we call *multiplicative scale invariance.*

4.2.2 Results of Modulo Discontinuity Operator

Based on the algorithm developed in the previous section, we computed hue edges associated with both the HSI and the CIE (L*a*b*) color spaces. Figs. 4.3 and 4.4 display the hue edges as well as the saturation/chroma and intensity/lightness edges of both the Mondrian and the peppers.

Comparison of the Mondrian hue edge map with the intensity edge map in both HSI and CIE color spaces show an amelioration of the shadow edge effect. Hue edges are qualitatively superior to normalized RGB edges in discounting shadow cues. The reason for this is that the *additive scale invariance* property in HSI space offers the added benefit of cancelling the saturation changes to the RGB components of the image radiance. An evaluation of the performance of hue segmentation is discussed in the next section.

4.2.3 Comparisons

Qualitatively, a visual inspection of the Mondrian edges in the four color spaces used illustrates that hue discontinuities in both HSI or CIE spaces appear to be the best measure for detecting material changes, independent of shadow edges. In order for

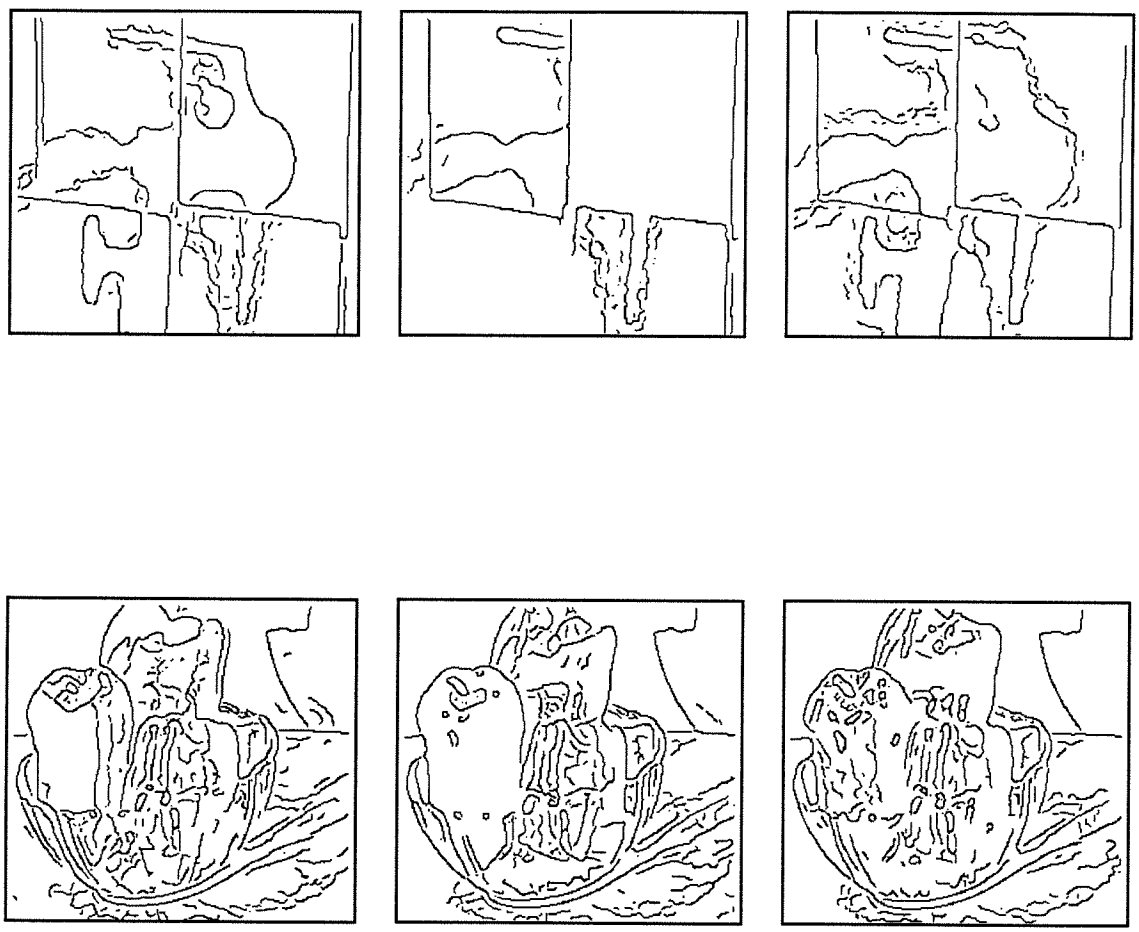


Figure 4.1: Red, green, and blue Canny edges of (a) Mondrian and (b) peppers of Fig. 2.1.

Figure 4.2: Normalized red, green, and blue Canny edges of (a) Mondrian and (b) peppers of Fig. 2.1.

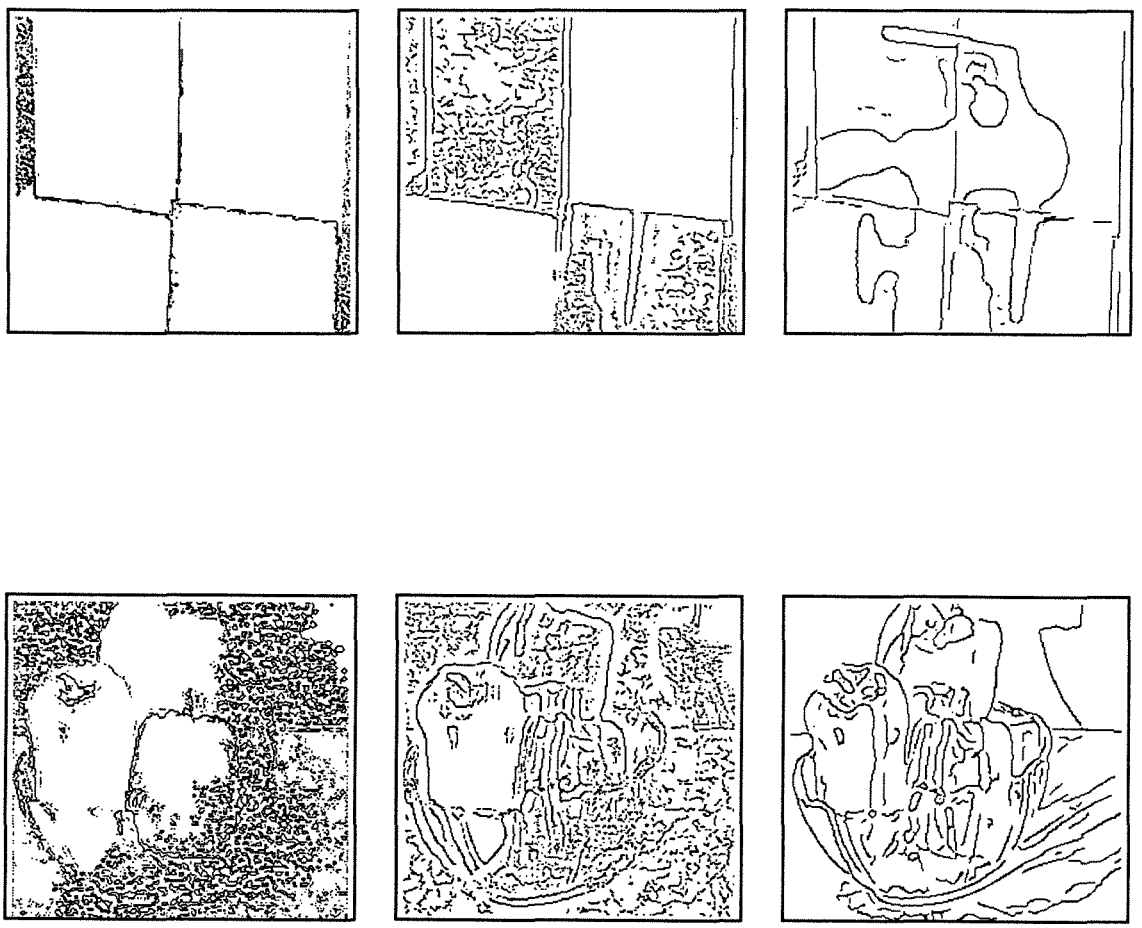


Figure 4.3: Hue, saturation, and intensity edges of (a) Mondrian and (b) peppers. Excellent performance is achieved in discounting the shadow boundary of the Mondrian hue edge map. The unsatisfactory performance for the peppers is due to the instabilities of the hue transform at low saturations and intensities associated with the shadows and background of Fig. 2.1.

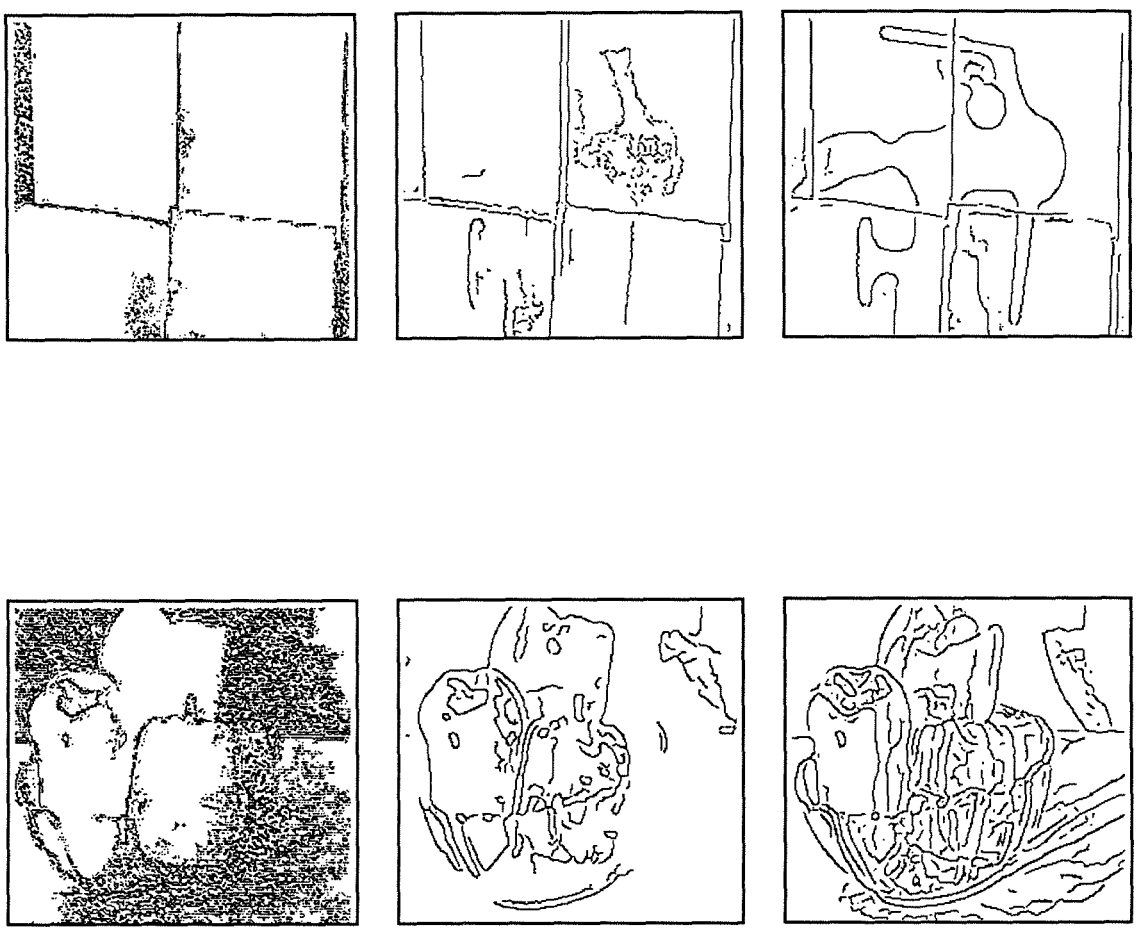


Figure 4.4: CIE L*a*b* Hue, chroma, and lightness edges of (a) Mondrian and (b) peppers. Results for the Mondrian are similar to Fig. 4.3. As in the previous figure, the hue edges were obtained with the modulo algorithm, while the chroma and lightness edges were obtained with the Canny edge operator.

us to form a more quantitative judgement of the relative performance in the different color spaces, we use a modified figure of merit, F, first proposed by (Abdou and Pratt, 1979; Pratt, 1978) to compare different edge detection schemes. This is given by

$$F = \frac{1}{\max(Ni, Na)} \sum_{i=1}^{Na} \frac{1}{1 + \alpha d^2} \tag{4.4}$$

where Ni and Na represent the number of ideal and actual edge map points, α is a scaling constant (typically $\alpha = \frac{1}{9}$ to penalize offset edges more than smeared edges), and d is the separation from the location of an actual edge point to the closest ideal point[3]. (In this study, the ideal edge map was hand-constructed piecemeal from the intensity map.) The figure of merit, F, defined by Equation 4.4 will penalize both non-localized edges and inaccurately positioned ones. Values of F close to 1 are ideal.

Applying this measure on the saturated portion (central half) of the Mondrian image gives the following results:

Color space	Figure of Merit		
R:G:B	0.2596	0.4126	0.2626
nR:G:B	0.1830	0.2509	0.2330
H:S:I	0.9158	0.1765	0.3093
(CIE)H:C:L	0.7421	0.3215	0.2733

In general, normalized color gives spurious edges within low intensity regions. That is why its figure of merit is penalized more than that of the RGB system. The best performance for object segmentation is achieved in hue space. For this particular

[3]In the original reference, Abdou (1979) evaluated edges within 1-dimensional images. In our extension to the 2-dimensional image of the Mondrian, our definition of d will give an upper bound on F. We adopt this viewpoint because the problem of finding the correct location of an actual edge point in a 2-D image is ill-posed.

example, hue segmentation in HSI space outperforms the CIE (L*a*b*) hue segmentation.

4.3 Improvement with Regularization

While the hue segmentation works adequately in some areas, it is clear that it performs poorly in other areas. As the image pixels approach low saturation and low intensity values, hue values become unstable. This is apparent in the upper left hand portion of the Mondrian-edge image in Fig. 4.3a and in the background regions of the pepper-edge image in Fig. 4.3b. This suggests hierarchical processing based on confidence values depending on saturation and intensity. That is, hue regions of low saturation and intensity should be smoothed before segmentation is performed. This strategy is the complement of combining all sensory data to find all "true" edges (Ohlander, 1976). The rationale behind this algorithm is that a cooperative pixel neighborhood scheme is appropriate for scene segmentation since object pixels will correspond more closely to nearby pixels. This neighborhood correspondence is a fundamental premise in Geman and Geman's Markov random field (MRF) formulation.

4.3.1 Markov Random Field Formulation

Smoothing unstable hue regions is performed by utilization of a regularization technique based on MRFs. While other researchers (Hurlbert and Poggio, 1989; Wright, 1989) have used the MRF formulation in color segmentation, the research of (Daily, 1989) comes closest in spirit to this work in the selection of hue as a useful measure for image segmentation. Although our use of the MRF formulation requires heavy digital computational resources, its implementation of minimizing a non-convex functional is directly transferable to analog VLSI hardware.

In this study, we use deterministic hue discontinuities to segment hue regions smoothed by a first-order membrane type stabilizer. This represents a deterministic approximation to the underlying stochastic MRF algorithm of (Geman and Geman, 1984) (see also (Geiger and Girosi, 1990).) The advantages of using such a deterministic approach are simplicity, speed, and the fact that hue values are smoothed while hue discontinuities are computed at the same time. Noise is eliminated while discontinuities are preserved. The algorithm utilizes a first-order Tikhonov stabilizing functional (Poggio et al., 1985) by minimizing the following "energy" function:

$$E_{tot} = E_{data} + E_v + E_h + E_{line},$$

where

$$E_{data} = \lambda \sum_{ij} [RelativeDist(Hue_{ij}, D_{ij})]^2$$

$$E_v = \sum_{ij} [RelativeDist(Hue_{ij+1}, Hue_{ij})(1 - v_{ij})]^2$$

$$E_h = \sum_{ij} [RelativeDist(Hue_{i+1j}, Hue_{ij})(1 - h_{ij})]^2$$

$$E_{line} = \alpha \sum_{ij} (h_{ij} + v_{ij})$$

Here, i, j are pixel locations in a rectangular lattice, λ is associated with the data confidence to smoothing ratio, v_{ij} and $h_{i,j}$ are vertical and horizontal hue discontinuity line processes which take on values of 0 or 1, and α is the energy penalty for forming these line processes. Because the formulation of this energy functional has many local minima, standard gradient-descent techniques can converge to non-global solutions. Finding the optimum solution requires utilizing an annealing process which allows convergence to unique global solutions during each annealing cycle. The deterministic annealing schedule we utilized was based on adjusting the λ parameter from very small values (high degree of smoothing) to higher values (small degree of smoothing).

The method we used in segmenting scene boundaries was to start with the intensity edge map generated from the Canny edge operator and to gradually eliminate those edges not due to hue differences. Our constraint is that hue discontinuity line processes would only form if pixel saturation values exceeded some specified minimum value (20% saturation).

4.3.2 Results

By utilizing a membrane type stabilizer to smooth low confidence regions due to low saturation or low intensity, improvement in object segmentation is obtained. The corresponding hue edges for regions away from unstable saturation are shown in Fig. 4.5 and shows significant improvement over the edges shown in Fig. 4.3b.

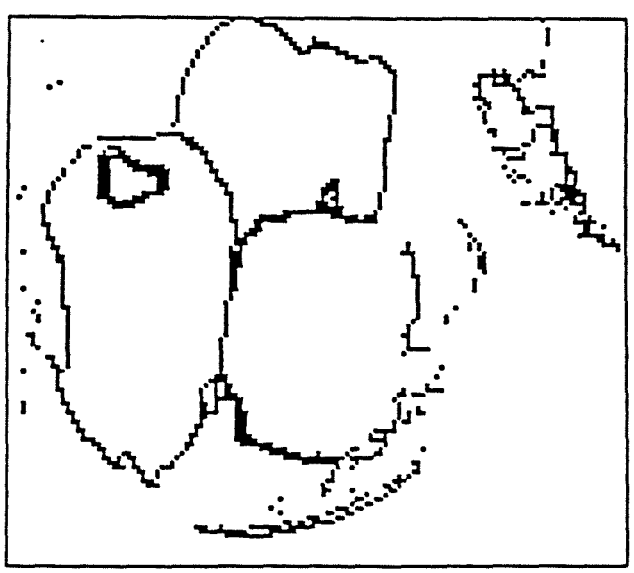


Figure 4.5: Smoothed hue edges away from low saturation regions. The threshold was 20% saturation.

The procedure used to generate this map was to start out with an initial edge map identical to the intensity edge map and to minimize the hue energy functional. Edges are created if both critical hue thresholds and critical saturation thresholds are exceeded, while intensity edges are eliminated if hue differences are small.

Notice that edges associated with intensity variations due to shading on the curved bell-pepper surface and intensity filtering through the translucent bowl have been eliminated in the hue edge map. This hue edge map correlates more strongly to the material boundaries than the intensity edge map.

In general, the MRF regularization formulation for global minimization requires the "tweaking" of parameters that cannot be derived *a priori*. Selecting proper values for the regularization parameters α and λ are problem specific, while the selection of the annealing procedure is strongly problem dependent bordering nearly on intuition and art. Automating this process of global optimization is an active research topic (Harris, 1991; MacKay, 1992) and lies beyond the scope of this thesis.

Chapter 5

Analog VLSI Chip Implementation

Based on the performance of our hue segmentation algorithm, we proceeded to the next step of our research program, implementing a hue sensor using integrated circuit technology. Because the underlying photoreceptor signals are continuous, it is appropriate to use the analog paradigm rather than the digital one, so that all the information of a spectral signal can be elegantly extracted. In this chapter we describe the one pixel (zero dimensional) intensity, normalized color, and hue sensors which were manufactured. This represents the first analog CMOS VLSI circuit that uses on-board photoreceptors responsive to different spectral components[1]. We also present a current segmentation circuit useful for detecting discontinuities. Our main objective in this research effort of achieving the first steps of on-chip color processing has been met. In our analysis of future hardware improvements (in this chapter and the next) we uncover the importance of spectral filter design on the assigned task. The ideal design requires narrow-band filters for precise spectral sampling and wide

[1]Portions of this chapter have already been published (Perez and Koch, 1994).

overlap for accurate encoding. Further work is required in optimizing spectral filters to improve measurement accuracy and system performance.

5.1 The Analog Representation

Our technology of choice is analog CMOS VLSI technology, as developed and applied to a range of neuromorphic systems by Carver Mead (1989) and his collaborators. A significant number of circuits have been successfully built with this technology, including a *silicon retina* with logarithmic photoreceptors (Mahowald, 1992; Sivilotti et al., 1987), resistive networks for smoothing and "fuses" for detecting discontinuities (Harris, 1990b; Harris, 1991). This will allow us in the future (Koch, 1989) to integrate smoothing and discontinuity detection circuits with our color sensors to build a single *smart sensor* for directly computing hue discontinuities.

The alternative CCD camera/color imaging system (D'Luna and Parulski, 1991; Khosla, 1992; Watanabe et al., 1984) is a power-hungry system requiring analog pixel scanning, A/D conversion, and digital computation. In comparison to CCD camera-based vision systems, integrated spectral sensors combined with analog CMOS circuits operating in the subthreshold regime offer low power consumption[2], real-time performance, illumination independent solution to color segmentation.

Although the previous chapter successfully demonstrates a number of machine vision algorithms to segment hue, these algorithms are not amenable to analog circuit implementation due to their complexity (witness, for instance, Equation 4.3). Furthermore, no direct analog circuit can be constructed for hue utilizing Equation 2.7 because the required division operation is not a functional analog computational unit. Our approach fortunately overcomes these obstacles. We report on two color

[2]The *silicon retina* requires less than 1 mW of power, most of which is used in the photoconversion stage.

circuits that perform the required computations using simplified algorithms. Because normalization[3] lends itself quite readily to an analog circuit implementation, we first built a circuit for computing normalized color with inputs provided by on-board red, green and blue photoreceptor sensors. In order to incorporate the *additive/shift* invariance property we utilize an opponency strategy to compute hue components. The final outputs of the hue circuit are the x and y components of hue rather than the single hue angle. It is interesting that this functional progression from three wavelength-selective signals to opponency to hue imitates the known stages of color computation in the primate visual system (De Valois and De Valois, 1975; Lenny and D'Zmura, 1988). These circuits should be viewed as exploratory designs, proving that analog VLSI hue chips are feasible. We first describe some of the practical aspects of chip design and fabrication.

5.2 Chip Background and Experimental Setup

Because all of our chips are fabricated using the government sponsored silicon foundry service MOSIS, we are restricted to using standard CMOS and BiCMOS processes. Over the visible range of light, the spectral sensitivities of the various kinds of photodiodes and phototransistors available in these processes is apparently insufficient for full color vision according to Delbrück (1993). Thus, we did not exploit any intrinsic wavelength filtering of silicon for this research effort[4].

Furthermore, we do not have access to the highly developed colored polymer film deposition technology found in the majority of modern solid-state commercial video

[3]Although normalization is a division operation, its output is restricted to the range of 0 to 1. The general division operation however has no restriction since large outputs can be generated from small denominator inputs.

[4]However, in Sections 5.6.2 and 6.7, we evaluate their advantages and suggest improvements of these devices.

cameras. We therefore had to manually deposit spectral filter over our phototransistors, a quite tedious process which does not lend itself to an extension to one- or two-dimensional arrays of photoreceptors.

We manually glued color gelatin filters (KODAK, 1981) on top of standard bipolar phototransistors, which were further covered by an IR filter so that we could concentrate our experiments in the visible range. Spectral transmissivity of these filters are shown in Fig. 5.1. (Also included in the figure is the spectral response curve for bipolar phototransistors.)

The areas of the phototransistors were approximately sized to accommodate the color filter transmissivity[5]. The phototransistors we use have logarithmic voltage responses over five orders of magnitude of intensity change (Mead, 1989; Sivilotti et al., 1987). In our application, we use photocurrent output that varies linearly with intensity. All chips were fabricated using the MOSIS 2 μm process. We used the standard 40 pin "Tinychip" die which has an effective 1.6 mm by 1.6 mm usable chip area.

To measure spectral tuning curves for our constructed color sensors, we used the calibrated prism monochromator and halogen incandescent lamp source setup of Delbrück (1993). The prism monochromator (Gaertner Scientific Corp.) has adjustable input and output slits which we use to regulate intensity levels. The light source is a 400 watt General Electric "quartzline" lamp. Because the light source used in this experiment is approximately a black body radiator with peak intensity in the infrared, the intensity varies 3 orders of magnitude over the spectrum of interest as shown in Fig. 5.2. Even though the light source exhibits these non-ideal spectral characteristics, our circuit compensates for this variation by performing a *multiplicative/scale* invariance operation, thus removing intensity contributions from

[5]The red cell was sized to 100 by 100 μm while the green and blue cells were sized a factor of 1.39 and 2.18 larger, respectively.

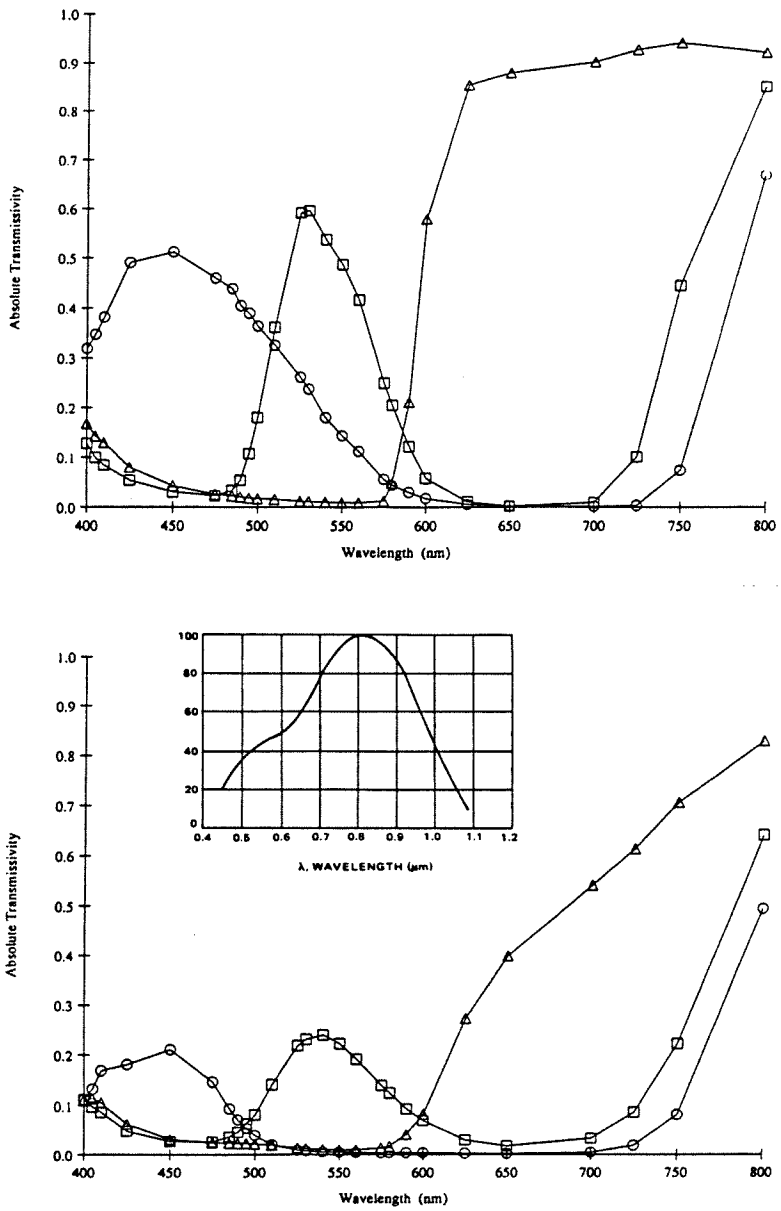


Figure 5.1: Filter spectral transmissivity of (a) Kodak Wratten gelatin filters and (b) Kodak Ektachrome 100 HC slide film. The curves were calculated from chip intensity measurements of filtered and un-filtered phototransistors, and thus, represents the total system spectral response. Kodak Wratten gelatin filters No. 24, 58 and 38A were used for the red, green, and blue photoreceptors in (a). Blue transmissivity for slide film in (b) attenuates illumination much greater than in (a) and must be compensated. Included in the insert is the spectral response of a typical bipolar transistor such as the Motorola MRD300 series (Bliss, 1993).

the chromatic computation. We report our results in the visible 400 to 700 nm wavelength range.

Instead of building our circuits using the prevalent voltage mode of operation, our circuits operate in the current mode[6], an approach that is gaining popularity (see (Toumazou et al., 1990)). Since *charge* rather than *voltage* is the active parameter, higher usable gains, accuracy, and bandwidth are expected. Also in the current representation, linear computations such as addition and subtraction are more easily achieved than in the voltage representation (for examples of available "building block" circuits see the compendium in (Seevink, 1988)). Even though, we lose the logarithmic compression (Mahowald, 1992) of intensity that the voltage-mode of operation affords, we achieve the same functionality with the current normalization circuit.

5.3 Normalized Color Sensor

Because of its relative design simplicity, we first designed and built a Normalized color sensor (Perez and Koch, 1992). The basic circuit design is based on Gilbert's translinear principle (Gilbert, 1975) and consists of a current-mode normalization circuit (described by (Seevink, 1988)) that has the desired scaling behavior. The input currents I_r, I_g, I_b from the three phototransistors produce normalized output currents N_r, N_g, N_b such that

$$\begin{aligned} N_r &= I_{\text{bias}} \left(\frac{I_r}{I_r + I_g + I_b} \right) \\ N_g &= I_{\text{bias}} \left(\frac{I_g}{I_r + I_g + I_b} \right) \end{aligned} \tag{5.1}$$

[6]In monochromatic image processing, a current-mode approach has also been used for modeling the synaptic interactions in a contrast sensitive silicon retina (Boahen and Andreou, 1991).

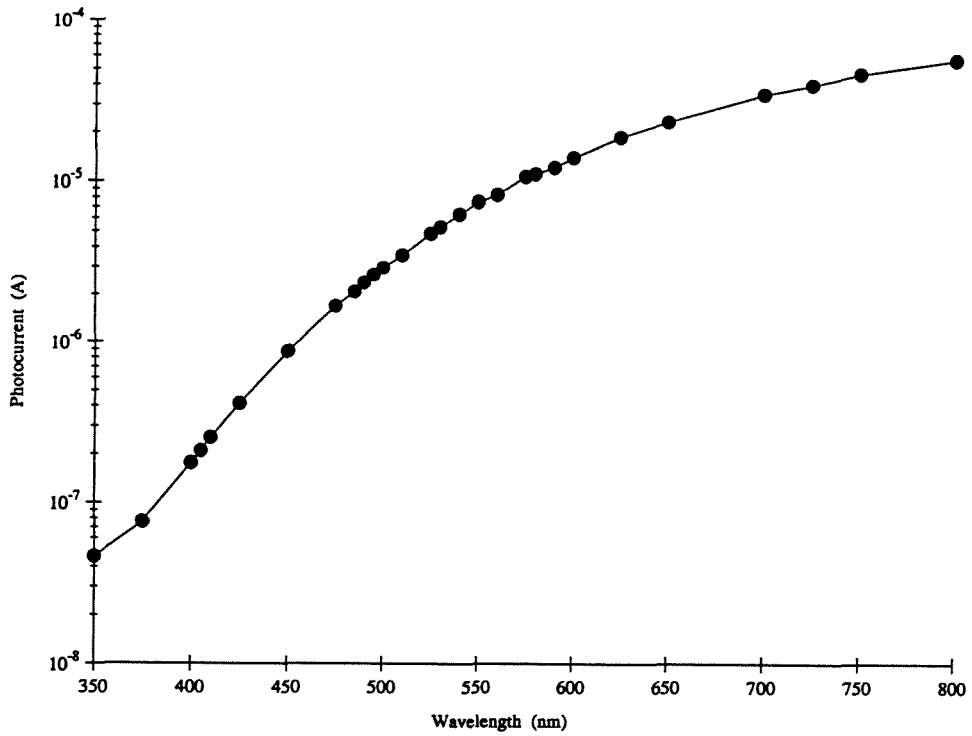


Figure 5.2: Typical intensity photocurrent output from the experimental setup used throughout this thesis as a function of wavelength. Because of the nature of the tungsten halogen light source, a variation of intensity of 3 orders of magnitude occurs across this spectrum.

$$N_b = I_{\text{bias}} \left(\frac{I_b}{I_r + I_g + I_b} \right),$$

where I_{bias} is set to operate in the subthreshold regime (Mead, 1989). Notice that this equation matches the normalized color equation given in Equation 2.2, and thus offers *multiplicative/scale* invariance. This normalization operation is achieved with 7 transistors and functions over 4 orders of magnitude of measured intensity change.

To confirm the behavior of the normalizing circuits shown in Fig. 5.3 we apply the analysis (and notation) of Mead (1989) (See also (Gilbert, 1984; DeWeerth, 1991)). For the simple color normalization circuit shown in Fig. 5.3a, if we assume that all transistors are saturated, then the input and output currents become

$$\begin{aligned} I_C^{in} &= I_0 e^{\kappa V_C} \text{ for } C = (R, G, B) \\ I_C^{out} &= I_0 e^{\kappa V_C - V} \end{aligned} \tag{5.2}$$

This directly implies that

$$I_C^{out} = I_C^{in} e^{-V} \quad \text{for } C = (R, G, B) \tag{5.3}$$

Summation over the output currents of Equation 5.3 and application of Kirchoff's current law gives

$$I_{\text{bias}} = \sum_{C=(R,G,B)} I_C^{out} = e^{-V} \sum_{C=(R,G,B)} I_C^{in} \tag{5.4}$$

Combining Equation 5.3 with Equation 5.4 results in the normalized color of Equa-

tion 5.1, which in the notation of the current analysis is given by

$$I_C^{out} = I_{\text{bias}} \frac{I_C^{in}}{\sum_C I_C^{in}} \quad \text{for } C = (R, G, B) \tag{5.5}$$

Similar analysis on the circuits shown in Fig. 5.3 will verify the associated equations. Of particular interest is the series diode connected normalizing circuit we call "square" normalization shown in Fig. 5.3b, which has the canonical equations given by

$$I_C^{out} = I_{\text{bias}} \frac{I_C^{in\,\frac{\kappa+1}{\kappa}}}{\sum_C I_C^{in\,\frac{\kappa+1}{\kappa}}} \quad \text{for } C = (R, G, B) \tag{5.6}$$

The advantage of this equation is not fully realized in CMOS devices because $\kappa = 0.5$ to 0.7. However, bipolar transistors have $\kappa = 1$ so that the denominator in Equation 5.6 exactly represents the distance measure, since its exponent $\frac{\kappa+1}{\kappa} = 2$. In general, for $M-1$ diodes in series, the exponent, P, is given by

$$P = 1 + \frac{1}{\kappa} + \frac{1}{\kappa^2} + \ldots = \sum_{m=1}^{M} \frac{1}{\kappa^{m-1}}$$

So generally we have

$$I_C^{out} = I_{\text{bias}} \frac{I_C^{in^P}}{\sum_C I_C^{in^P}} \quad \text{for } C = (R, G, B) \tag{5.7}$$

Fig. 5.3c shows a translinear circuit stacked with "m" diode connected transistors that obeys these relations. Obviously, there is a finite limit to how many diodes can

be connected in series before the voltage headroom is exhausted.

In the limit, we note that the "m-th" normalization approaches the "winner-take-all" normalization shown in Fig. 5.3d. Analysis of this circuit is presented elsewhere (DeWeerth, 1991; Lazzaro, 1990; Lazzaro et al., 1988).

Fig. 5.4 shows the experimental spectral tuning curves for the normalized color circuit at two different illumination levels: (a) an arbitrary baseline level and (b) 10 times that baseline level. The ten-fold increase in illumination uniformly across the entire spectrum was achieved by opening the monochromator slit, and confirmed by photocurrent measurement. The figure shows that for illumination level increases of one order of magnitude, the circuit output increases no more than 25% for N_r and N_g. That N_b is 60% off for the baseline illumination is attributed to circuit operation in the non-saturated regime for this particular experiment. Ensuring operation in the saturation regime will improve the performance of the normalization circuit. The circuit modification that will achieve this is a change in the source of the input transistor from ground to a higher voltage, V_{ref}. Experimental results of this simple circuit modification, with V_{ref} set to 0.6 volts, are shown in Fig. 5.5. The figure shows that for a 10-fold increase in illumination, the circuit output increases no more than 30% for N_r, N_g, and even N_b.

Since the translinear principle is based on Kirchhoff's current and voltage laws and the transistor voltage-current logarithmic relationship, the circuit normalizing behavior will hold over a large range (4 orders of magnitude from experiments) of input currents. The advantage of this circuit is that intensity gain control can be achieved electronically instead of by mechanical iris movements as is customary in standard camera systems.

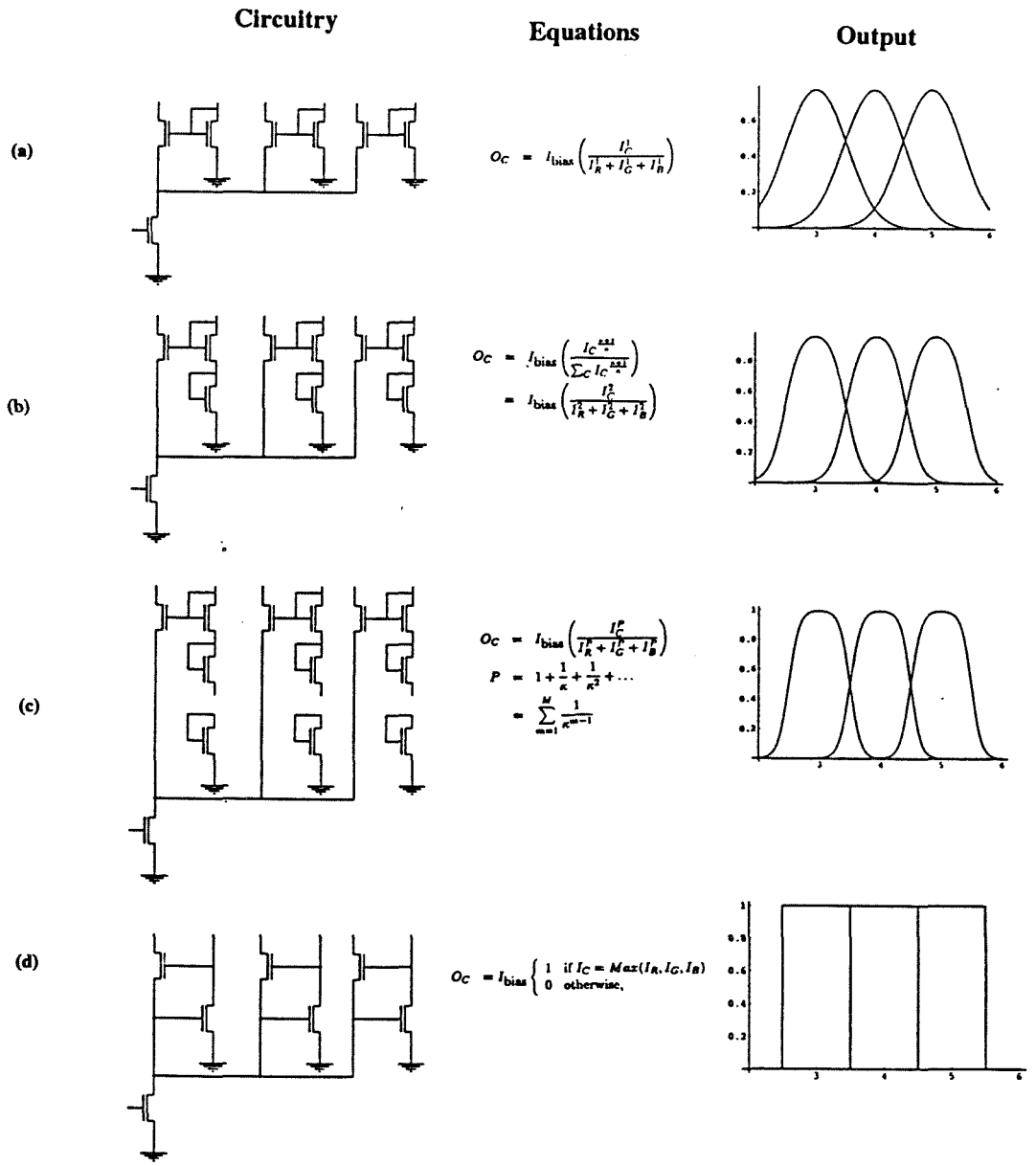


Figure 5.3: Normalization circuitry, canonical equations, and simulated output for (a) "linear" (b) "square" (c) "mth" and (d) "winner-take-all" normalizations. The input was a large array of evenly spaced Gaussians. Circuit equations are for bipolar rather than CMOS transistors, whose differences are further elaborated in Section 5.6.2.

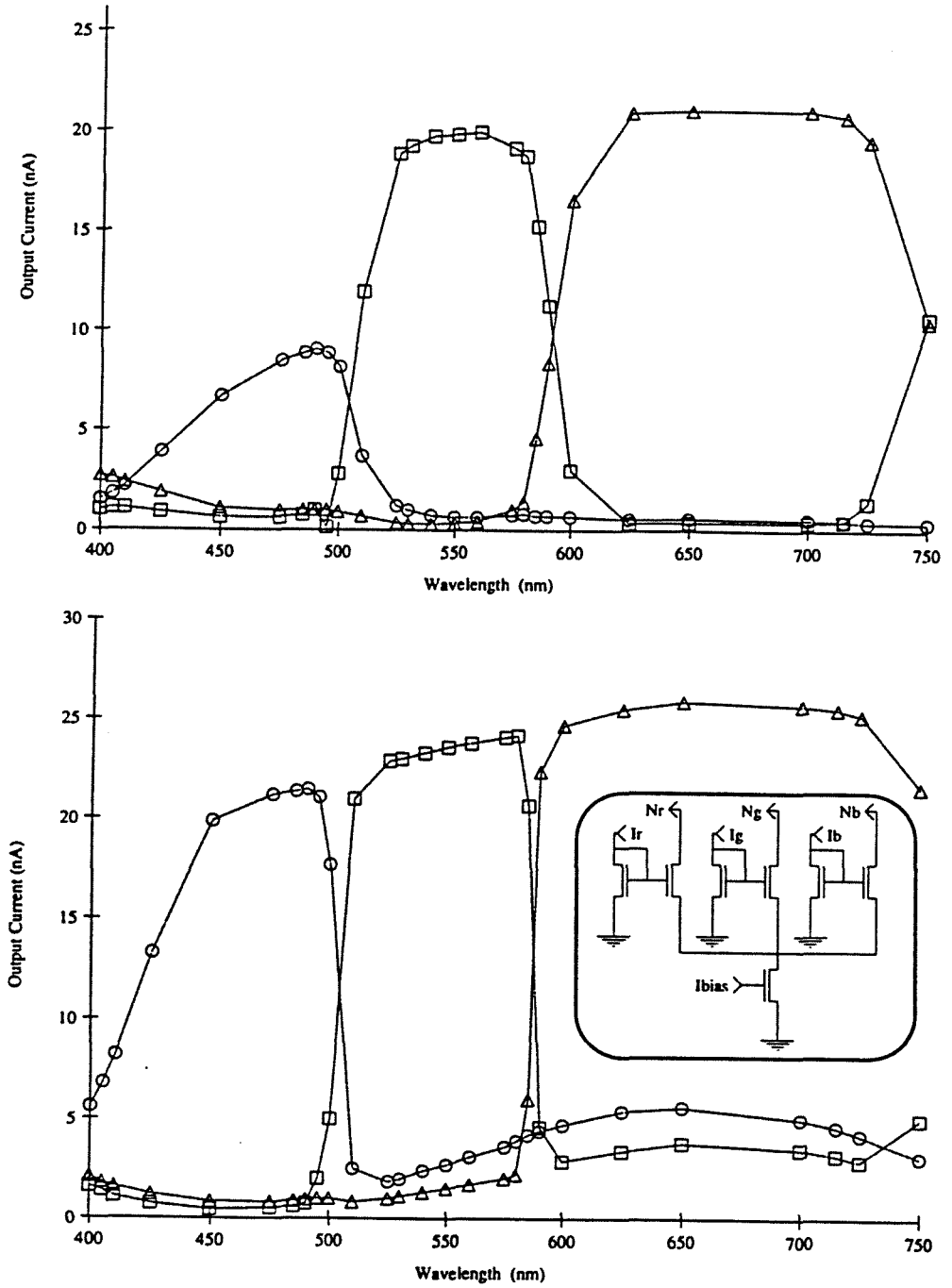


Figure 5.4: Spectral Tuning Curves for Normalized Color Circuit from (a) standard intensity (arbitrarily set) and (b) 10 times standard intensity. The normalization operation is demonstrated on $N_r(\Delta)$ and $N_g(\square)$ because a 10-fold increase in intensity causes a 1.25 increase in circuit output. The 60% attenuation of $N_b(o)$ is attributed to low S/N from the experimental setup and the resultant non-saturation circuit operation. In the central portion of the curve, the sum of the normalized color currents equals a constant I_{bias}. The circuit diagram used for normalized color is shown in the insert.

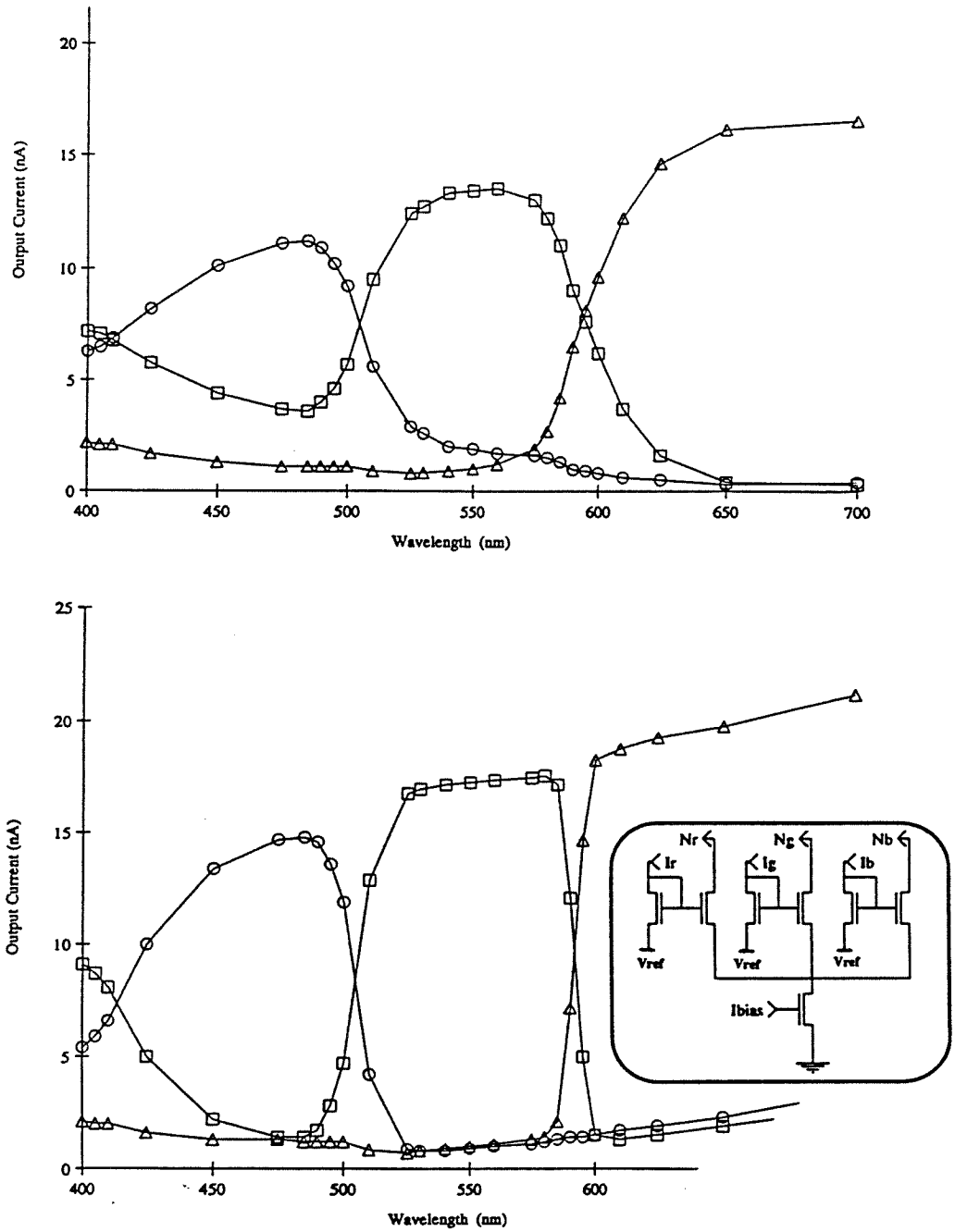


Figure 5.5: Spectral Tuning Curves for Improved Normalized Color Circuit from (a) standard intensity (arbitrarily set) and (b) 10 times standard intensity. The corresponding circuit used for this measurement is shown in the insert. The source of each input transistors is set to some reference voltage (V_{ref} = 0.6 volts) to ensure operation in the saturation regime.

5.4 Hue Sensor

In addition to *multiplicative/scale* invariance, hue has *additive/shift* invariance. To help us construct this extra property, we adopted the opponent cell concept from biological vision processing. Thus, we designed an opponent circuit which computes the following *additive/shift invariant* parameters:

$$\begin{aligned} X^+ &= \min(2R - 2G, 0) \\ X^- &= \min(2G - 2R, 0) \\ Y^+ &= \min(R + G - 2B, 0) \\ Y^- &= \min(2B - R - G, 0) \end{aligned} \tag{5.8}$$

Here R, G and B are the currents from the "red," "green," and "blue" phototransistors. Each of the opponency currents X^+ and X^- are half-wave rectified, such that if either one is positive the other one is zero, that is $X = X^+ - X^-$ (the same applies to the opposing pair Y^+ and Y^-). In an earlier version of this chip, we used a single current X that could take on both negative and positive values. However, imbalances due to inherent mismatched properties of the circuit lead to poor performance. The opponency currents are then processed by the translinear normalization circuit discussed above to give normalized x and y current values that map directly onto a unit diamond on the 4 quadrants of the color plane shown in Fig. 5.6.

These currents are given by

$$x^+ = I_{\text{bias}} \left(\frac{X^+}{X^+ + X^- + Y^+ + Y^-} \right) \tag{5.9}$$

$$x^- = I_{\text{bias}}\left(\frac{X^-}{X^+ + X^- + Y^+ + Y^-}\right)$$

with the corresponding equations for y^+ and y^-. Notice that we slightly warp the hue definition given earlier in Equation 2.7 for expedient hardware implementation[7]. This formulation is reminiscent of (Hurvich, 1981). (R-G is π apart rather than $\frac{2}{3}\pi$). In this case, hue will be defined as

$$Hue = Arctan\left[\frac{R+G-2B}{2R-2G}\right] = Arctan\left[\frac{y^+ - y^-}{x^+ - x^-}\right] \tag{5.10}$$

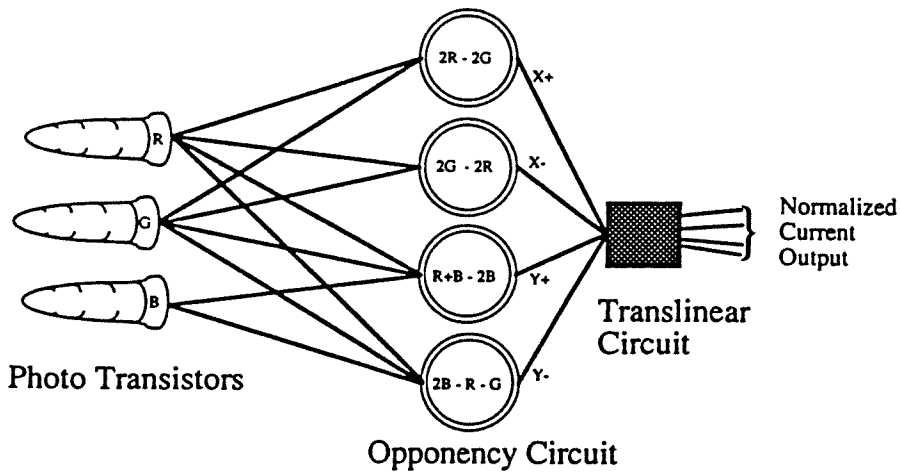
The block diagram circuitry for the hue sensor is shown in Fig. 5.6[8]. The outputs of the chip are the normalized half-wave rectified hue currents x and y, which are shown as a mapping on the unit diamond in the X-Y opponent color space. This mapping is fundamentally different from traditional color space mappings which utilize preprocessed RGB input signals that lie within a fixed region (typically 8 bits) of color space. Whereas traditional color processing operates within the radius of the color circle, the hue sensor element described here operates outside of the color radius as illustrated in Fig. 5.6.

As expected, Fig. 5.7 shows that the hue circuit measured output varies little with changing illumination levels. When the image intensity is increased by a factor of 10 uniformly across the entire spectrum, the x and y currents scale only by at most 25%, in a manner similar to the normalized color circuit.

Finally, we numerically computed off-chip the hue defined by Equation 5.10 on the basis of the measured x and y currents. For comparison, we additionally superimpose

[7]In particular, we avoid the $\sqrt{3}$ scaling required from the use of Equation 2.7.

[8]Each opponency unit contains 6 transistors for addition and subtraction of photocurrents while the translinear section contains 9 transistors to normalize 4 different signals (X^+, X^-, Y^+, Y^-). A basic hue sensor element, therefore, consists of 33 transistors.



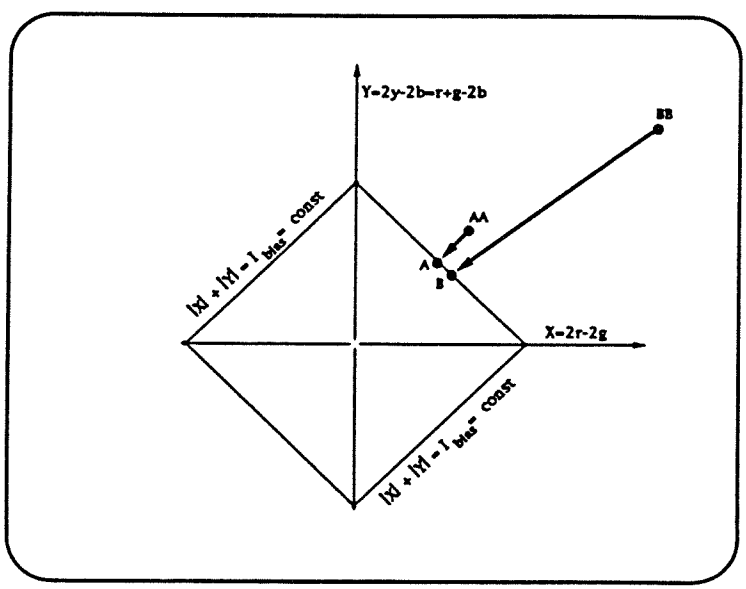


Figure 5.6: Block diagram of Hue sensor element. RGB current inputs are converted into opponency currents which are then converted into x and y components of hue in the translinear normalization circuit. Opponency circuit gives the *additive/shift* invariance property while the translinear part gives *multiplicative/scale* invariance. Opponency X and Y currents in the plane are mapped onto the unit diamond shown in the insert. The unit diamond is defined by $|X|+|Y| = I_{bias} = \text{const}$. In the insert, even though point BB is roughly twice as intense as point AA, their mappings to B and A respectively on the unit diamond indicate their hue similarities.

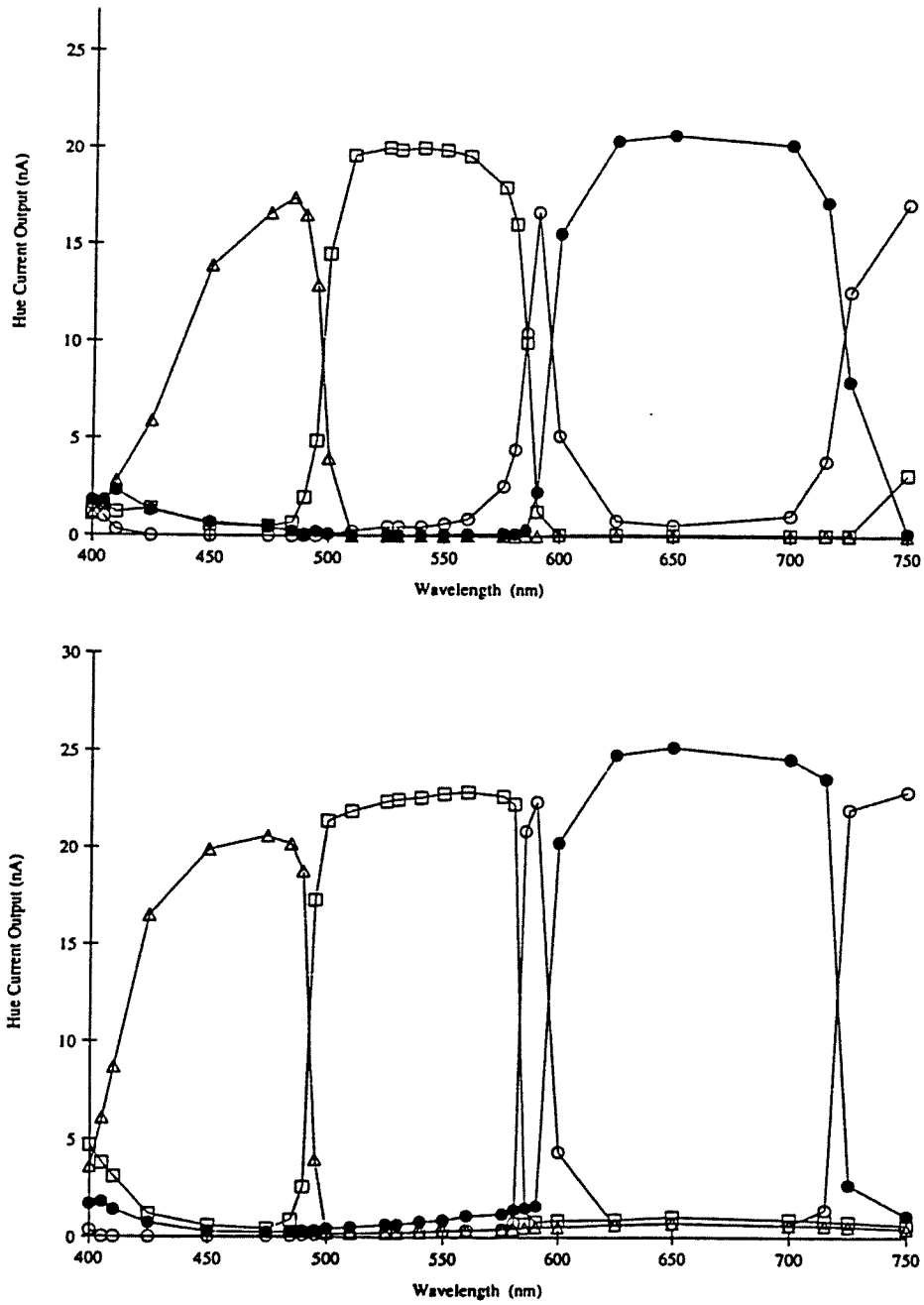


Figure 5.7: Spectral response of the hue circuit at (a) standard intensity (arbitrarily set) and (b) 10 times standard intensity. Similarities of the half-wave rectified hue currents $x^+(\bullet)$, $x^-(\square)$, $y^+(\circ)$, $y^-(\triangle)$ in (a) and (b) indicate that the circuit performs the normalization and thus exhibits *multiplicative/scale* invariance. Here an order of magnitude increase in illumination causes a maximum change of only 25% of current output. Notice the spectral sharpening of the y^+ hue current, which at 10 nm half-bandwidth represents at most 14% of input filter response shown in Fig. 5.1a.

the calculated hue based on photosensor input and human visual input in Fig. 5.8. The former is computed from the measured output photocurrents from RGB color filtered phototransistors shown in Fig. 5.1, while the latter is computed from the equivalent RGB spectral response curves of the human cone system as shown in Fig. 1.7. In both cases, hue is calculated from $Arctan\left[\frac{R+G-2B}{2R-2G}\right]$. In the visible range ($>$ 400 to 700 nm) all curves show a monotonic behavior. The hue chip output agrees within experimental error to the calculated hue based on phototransistor current input. Both have a flattened output spectral response in the 500 to 600 nm band. Because they have different input sensor characteristics, hue calculated from human visual curves does not have this flattened output spectral response.

We select the output of the chip to be the x and y currents, rather than the single variable hue, because we can easily compute them within analog circuitry. In the next section, we show how to utilize these x and y currents in segmentation.

5.5 Segmentation Circuitry

We envision two approaches for designing a hue circuit that locates discontinuities in hue space as discussed earlier. One is to convert the normalized $x^+x^-y^+y^-$ currents into voltages and utilize available circuits for detecting discontinuities (Harris, 1990b; Harris, 1990a)[9].

Another approach is to develop current-mode versions of discontinuity detectors such as resistive fuses. One advantage of the current-mode fuse over a voltage-mode one is that in principle, it can be made into more compact circuitry. Another advantage is that the domain of operation is a current based one so that the signal

[9]One possibility is to treat the x and y variables separately. Two separate networks will be used to calculate Δx and Δy, the differences between neighboring pixel elements. Another possibility is to evaluate the hue distance norm defined by $\Delta x^2 + \Delta y^2$.

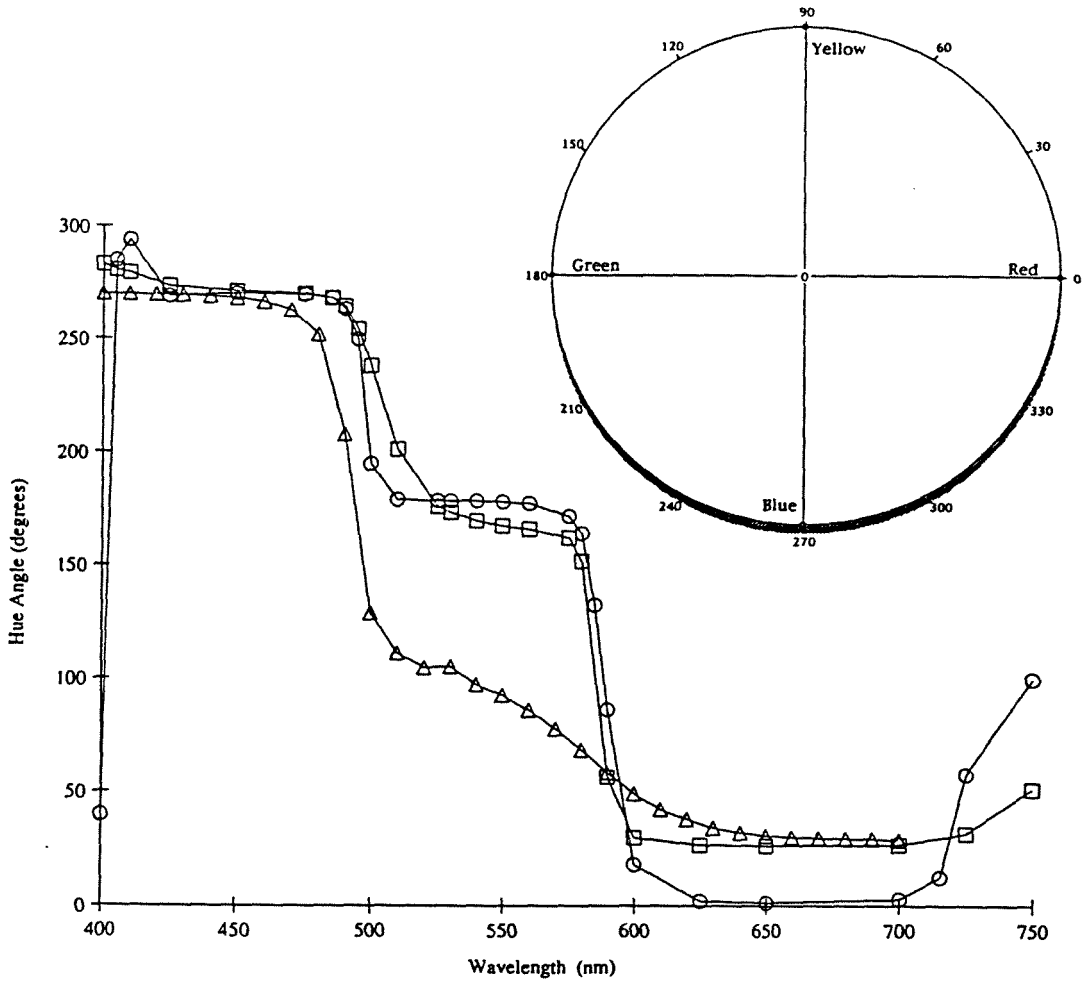


Figure 5.8: Comparison of hue as computed using (1) the x and y output currents of the hue chip (o) from Fig. 5.7, (2) the spectral response curves of the RGB filtered phototransistors (□) from Fig. 5.1, and (3) the spectral response curves (Ingling and Tsou, 1977) from human cones (△) from Fig. 1.7. If we discount the "flat" hue regions and the borders of the spectral range (near 400 nm) where our experimental setup has low signal content, all curves exhibit strictly decreasing hue values in the visible range. The flat range in the band 500nm to 600nm for the hue chip is predicted from the color photosensor input. (This issue and possible improvements are discussed in the next section.) The color wheel in the insert shows the color interpretation of the hue angle. An absence of hue angles in the 270 to 360 degree range is an affirmation that monochromatic light cannot produce "purple" hues.

processing remains self-consistent.

The elemental circuit that achieves one-dimensional current-fusing is shown in Fig. 5.9. When the fuse is not in operation, the fundamental filtering operation of this circuit is achieved by 1-2-1 current mirrors which copy $\frac{1}{4}$ of the current from the left element, $\frac{1}{2}$ from itself, and $\frac{1}{4}$ from the right element (a binomial filter). This filtering operation is reminiscent of current research in *smart* CCD image processing (Seitz et al., 1993; Yang, 1991), but the main difference is that current rather than charge is used as the signal parameter. To complete the current fuse, two high gain amplifiers and several pass transistors control the "fuse" by switching-off the filter when a threshold current is exceeded.

A layout utilizing the current-fuse element is shown in Fig. 5.10. The current-fuse chip layout consists of 7 layers of binomial 1-2-1 filters in the top portion of the die and 3 layers of current-fused binomial 1-2-1 filters at the bottom of the die. Each layer feeds into its succeeding layer resulting in a 1-2-1 convolution at each layer. For example, while the first layer is a 1-2-1 filter, the second layer will result in a 1-4-6-4-1 filter, the third layer will result in a 1-6-15-20-15-6-1 filter, and so on. In the future, elegant designs that utilize feedback to achieve convolution/filtering within one layer should be explored.

However, our goal was to prove the concept, which is validated by experimental data shown in Fig. 5.11. This figure superimposes the output from the second layer of the current-fuse chip when the input current is above and below the fuse threshold current. The results are clear: Filtering occurs below threshold, while the fuse operates above threshold. The filtering and fusing operation of this chip makes it an elementary hardware implementation of the regularization/line-discontinuity algorithm developed previously.

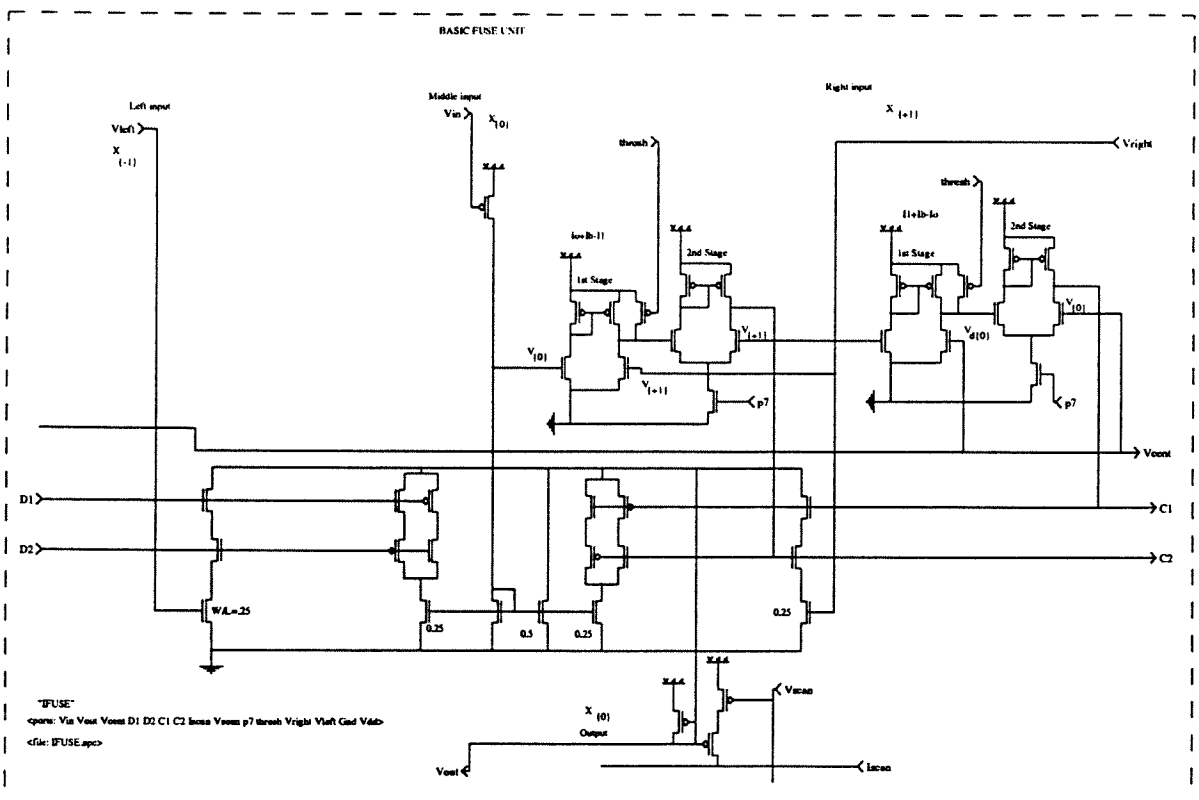


Figure 5.9: One-dimensional current-fuse circuit element which performs a 1-2-1 binomial filtering as long as a threshold current is not exceeded. Two high gain amplifiers control the pass transistors that switch the binomial filter operation.

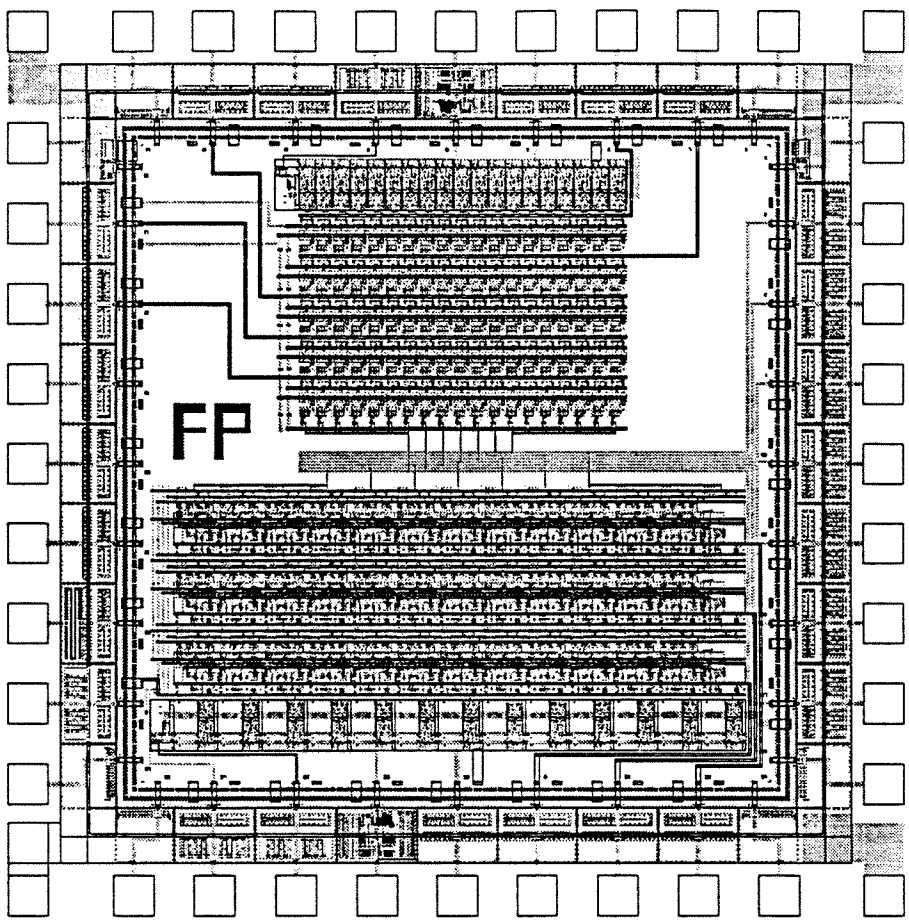


Figure 5.10: The current-fuse chip layout consists of 7 layers of binomial 1-2-1 filters in the top portion of the die and 3 layers of current-fused binomial 1-2-1 filters at the bottom of the die. Each layer feeds into its succeeding layer.

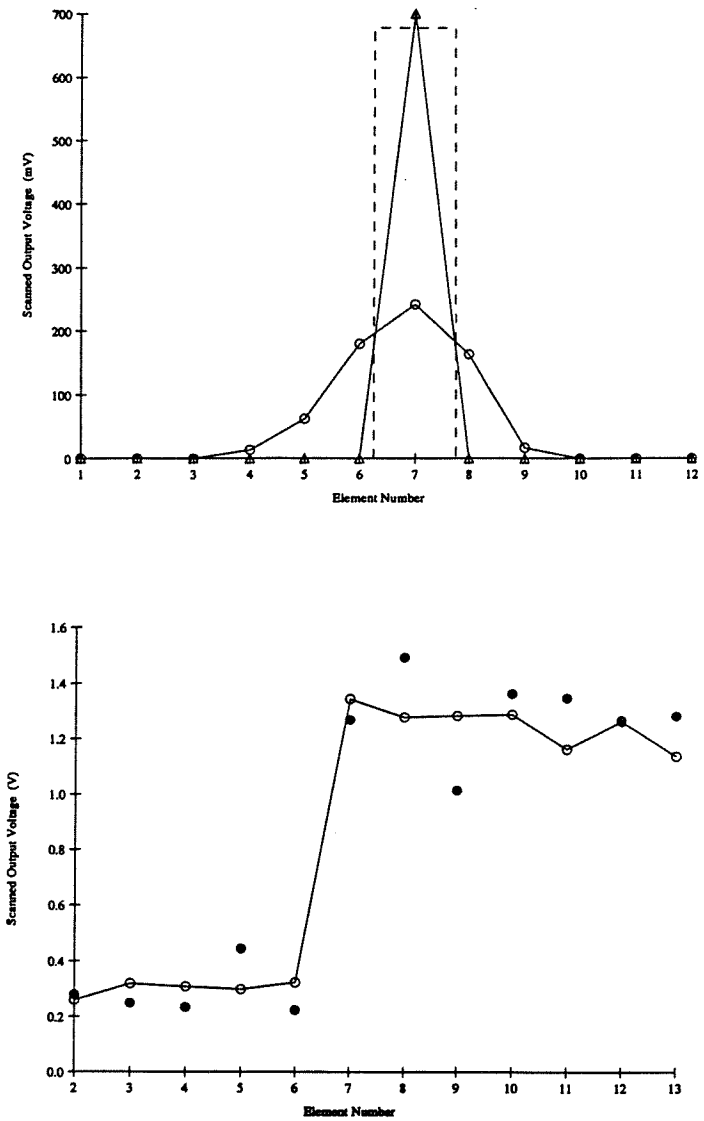


Figure 5.11: (a) Experimental data for current-fuse. The X-axis shows the cell number along a one-dimensional array, while the Y-axis is the scanned sense-amp voltage output which is proportional to current. The input for this experiment is a central current (at element 7) which is perturbed about a threshold current shown in dashed line. When the input current exceeds the threshold current, the output shown as (Δ) is not filtered. As the input current is reduced below the threshold current, the output signal shown as o is filtered according to the binomial expansion—in this case the second layer binomial expansion is a 1-4-6-4-1 filter. (b) Results of current fuse to step edge data. Solid circles show input data. Line through open circles show smoothing and fusing operation.

5.6 Future Hardware Issues

Based on the research presented in this chapter, further hardware development is still required before practical devices can be constructed. To develop color vision systems useful for robotics and autonomous systems, arrays of color sensors must be incorporated and integrated with vision processing units. To develop color sensors useful for inexpensive printing quality control, spectral discrimination and performance accuracy must be improved further.

5.6.1 Color Arrays

In our current unoptimized circuit design, one pixel, including three (RGB) phototransistors as well as normalized RGB, intensity, opponency, and hue circuits, spans 350 μm by 1600 μm on the die, most of which is inactive silicon real estate. This odd geometrical shape was selected to facilitate color filter placement. Without circuit optimization, a 5 by 5 pixel array can be implemented onto the smallest available "Tinychip" (using a 2 μm process), while significantly larger array sizes are possible with circuit optimization, bigger die sizes and smaller design rules. Because phototransistors take up 70% of the active silicon real estate area, their design optimization will have the biggest impact on circuit reduction. Still, manual spectral filter placement will be difficult to achieve so a method to place the filters easily and accurately is desired. One method is to fabricate color array dyes using a lithographic process and bond to the chip (Dillon et al., 1978a). Unfortunately, this process is proprietary to industry and thus unobtainable. A lower resolution, readily available, but cheaper version of this method is to use color slide film as the color filter array. Our evaluation of Kodak Ektachrome 100 HC slide film shows acceptable performance for spectral filtering. The test patterns we created on the slide film have minimum feature size of 30 μm and spectral transmissivity given in Fig. 5.1. The advantage of using

this technique is that one-dimensional and even two-dimensional color arrays can be constructed easily. This fabrication is more robust than the single filter placement method because less manual handling is required and the potential for chip damage is lowered.

Thus, extension to one- or two-dimensional arrays may proceed in one of two ways: either each pixel element contains (1) identical RGB triplet unit-cells or (2) a single R, G, or B sensor element. The first method offers higher spatial resolution, but the latter utilizes a third less hardware. It is the second method that is embraced in both vertebrate biological systems and modern CCD television imagers.

In many ways the reduction in hardware by sampling the image with a tessellated color array with lower resolution than the image, complements the MRF regularization formulation discussed earlier. The missing data points can be interpolated with regularization techniques, at which point the normal computations can proceed as usual. We show an example of the results of this technique in Fig. 5.12. Here, our pepper image is subsampled both spectrally and spatially (to speed up digital computation) and a hue map is constructed and segmented. Although the results seem coarse when compared to the previous chapter, the hue segmentation qualitatively functions similarly.

The traditional two-dimensional tessellation pattern (Bayer, 1976) utilizes a regular rectangular grid of 50% green, 25% red, and 25% blue photosensors. The green detectors are arranged in a checkerboard layout with the red and blue sensors interweaved at a coarser resolution as shown in Fig. 5.13a. The spatial sampling of the green sensors is twice that of either the red or blue ones. However, a problem arises when this array geometry is utilized for CCDs or any imagers using an interlaced readout. In particular, horizontal luminance edges near the Nyquist frequency can exhibit annoying 30 Hz yellow-cyan hue flicker as a result of the green-red interline linear array contrasting with the nearby green-blue array. The solution (Dillon et al.,

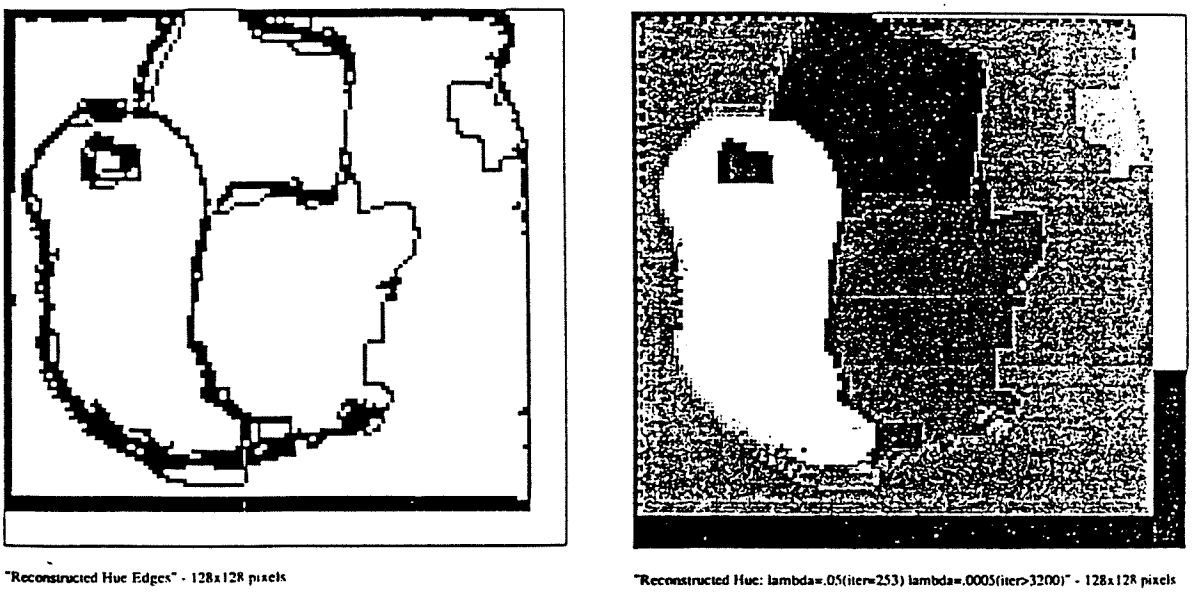


Figure 5.12: Reconstructed (a) hue edges and (b) hue map of subsampled pepper image from digital simulations. The starting configuration was a 128 by 128 pixel array containing a single randomly organized R, G, or B data point. Data reconstruction by the MRF formulation was applied to approximate the missing data, whereupon the hue segmentation algorithms were exercised.

1978b; Watanabe et al., 1984) to eliminate this flicker is to adopt an interline geometry which staggers the subsampled red and blue elements as shown in Fig. 5.13b[10].

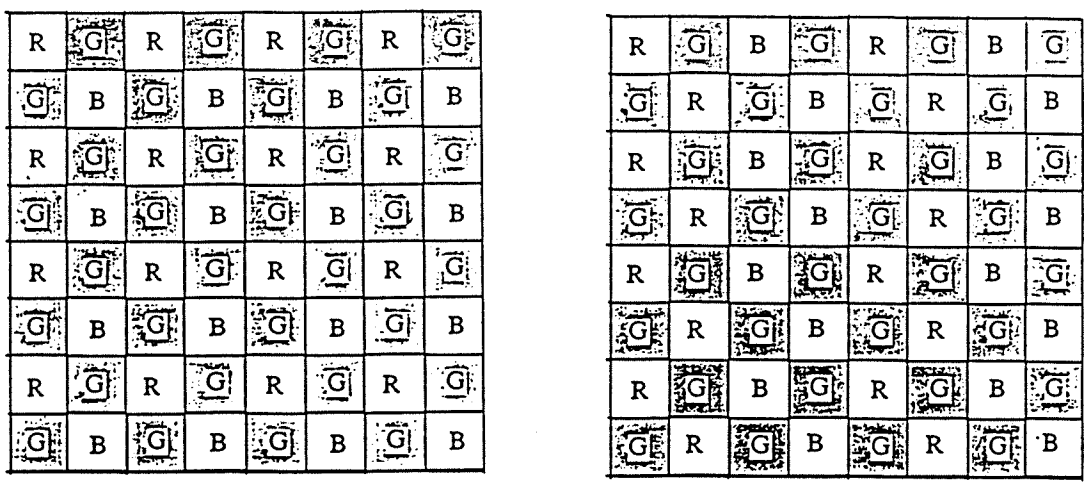


Figure 5.13: Color filter array for the (a) Bayer geometry and the (b) Interline geometry. Both contain 50% green, 25% red, and 25% blue sensors (Dillon et al., 1978). The advantage of (a) is that the regular pattern allows easier implementation in vision circuitry, while (b) removes color flickering for interlaced imaging devices when horizontal luminance edges are near the Nyquist frequency.

Modern CCD imagers further exploit human visual psychophysics by allocating more of the color image array to the luminance channel. In the latest reincarnation, 75% of the elements in the color filter array is allocated to the luminance green photoreceptor while 25% is allocated to the red and blue sensors (Parulski et al., 1992; Khosla, 1992).

[10]As nonsequitur, Welch (1991) shows that the flicker response of an NTSC image can be greatly improved if the saturation and intensity components were spatially filtered. This is achieved at the expense of sharpness response.

To further exploit biological vision, perhaps the array could be arranged into compact triangular tessellation patterns to imitate the human cone mosaic (Williams, 1986). Each element is generally surrounded by a 6 neighbor hexagonal gridwork with the color array forming irregular patterns. (For processing hexagonally sampled data see also (Feinstein, 1988; Mersereau, 1979).) The advantage of an irregular arrangement of color arrays is the elimination of global aliasing that arises with regular layouts. Although it seems impractical to manufacture VLSI circuitry with random patterns, evolutionary biology "thought" otherwise and developed the cones and the connecting neurons anyway.

A two-dimensional color analog VLSI sensors should have some combinations of these features. For example, the Bayer color filter structure is attractive in non-interlaced applications because the regular rectangular patterns allow easier implementation into VLSI circuitry. Triangular tessellation offers efficient packaging of sensors but requires more thought on color array geometries and interfacing.

For spectral extensions beyond trichromacy, the two-dimensional array may not be adequate because of reduced resolution resulting in unacceptable aliasing[11]. If the number of spectral classes is N then the spatial sampling in a two-dimensional array of each spectral class scales as $\sqrt{N}$. One possible solution is to use a scanned one-dimensional spectral array. This will achieve hyper-color vision without loss of spatial resolution. In biology, we see this very solution adopted by the Mantis shrimp.

5.6.2 Improved Hue Sensor

Because of the many-to-one wavelength-to-hue mapping in the spectrum, Fig. 5.8 shows that the hue chip exhibits low discrimination, and thus poor performance,

[11] We assume that only one spectral sample can be obtained per pixel position. This is a limitation of what is essentially a two-dimensional device. If on the other hand, the third dimension could be used for spectral processing, then the spatial resolution would not be compromised.

in the flat bands spanning 500 to 570 nm and 620 to 700 nm. These performance inaccuracies can be traced to two sources: (1) an essential experimental idiosyncrasy and more importantly (2) spectral filter characteristics.

Circuit or Experimental Improvement

The first source of the hue sensor performance inaccuracy can be traced to the non-ideal spectral characteristics of the experimental setup. Fig. 5.2 shows that the intensity increases over 3 orders of magnitude as the spectrum is swept. As shown in Figs. 5.4, 5.5, and 5.7, the increasing intensity skews the translinear circuit spectral tuning curves to flatten or "tanh-out" at wavelengths above 500 nm.

To compensate for this intensity increase, the translinear circuit element CMOS transistor, which is essentially a voltage controlled switch, can be replaced with a bipolar transistor, which is a current controlled device. Simulations shown in Fig. 5.14 using AnaLOG (Gillespie and Lazzaro, 1990), a circuit simulator, indicate that the bipolar translinear circuit performs "linear" normalization at all intensity values[12]. On the other hand, the CMOS translinear circuit performs a "quartic" normalization, and only achieves "linear" normalization at the lower intensities (at the shorter wavelengths in the spectrum). Thus, the use of bipolar transistors will compensate for the experimental idiosyncrasy of widely varying illumination. However, their adoption will not lead to compact circuit designs, since bipolar transistors can be 5 to 10 times larger than their CMOS counterparts.

How do we resolve this dilemma? First we note that real scenes do not have 3 orders of magnitude intensity variation in their spectrum. Typical scenes have reflectance values that span about a factor of 20 only (Moore, 1992); spectral variations span less than that (see for example (Billmeyer and Saltzman, 1981)). Thus,

[12]This is not surprising since Gilbert's analysis (1975) of the translinear circuit was for bipolar transistors.

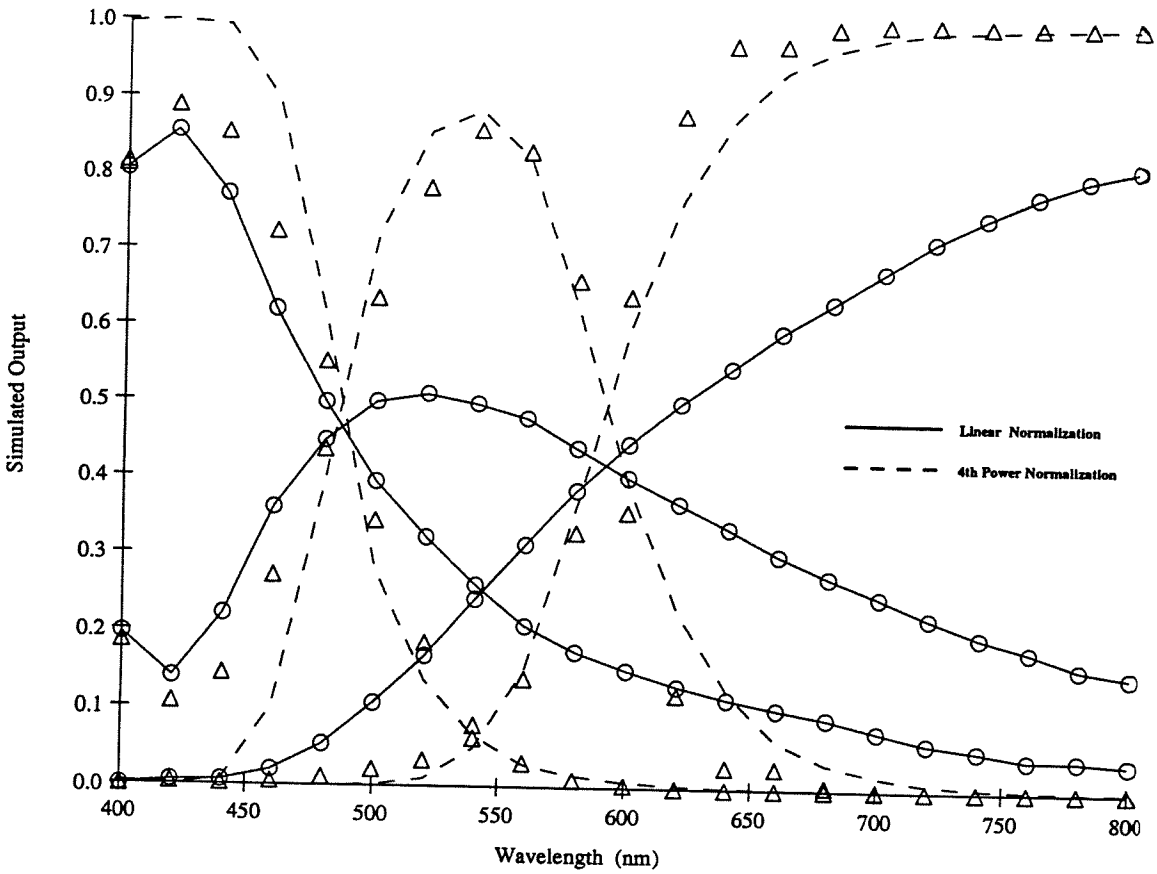


Figure 5.14: Comparison of simulated CMOS (△) and Bipolar (o) Translinear Circuits with inputs from intrinsic diodes consisting of spectral properties shown in Fig. 6.9. Intensity variation over 3 orders of magnitude matched the experimental variation shown in Fig. 5.2. Bipolar translinear circuits perform "linear" normalization over the spectrum (3 orders of magnitude intensity variation), while CMOS translinear circuits approach "quartic" normalization at higher intensities.

substituting bipolar transistors in the translinear circuit will mainly compensate for the intensity variation idiosyncrasy in the experimental setup. One obvious solution is to use a more spectrally uniform light source. The zenon-arc lamp, which closely approximates the natural solar spectrum, comes to mind since its intensity varies by no more than a factor of 2 within the visible range (ACTON, 1990; Hunt, 1987) as opposed to the factor of 500 for the tungsten incandescent light source used in our current experimental setup. The solution we propose, therefore, is to maintain the CMOS translinear circuit system and change the light source.

Yet inaccuracies in the hue discrimination performance are, only to a minor extent, due to non-ideal intensity variation from the experimental setup. For example, computed hue from spectral filter transmissivity indicated as (□) in Fig. 5.8 differs only slightly from the hue evaluated from the measured output of the CMOS translinear circuit (shown as (o) in the figure). In the next section, we show that the main contribution of discrimination inaccuracies comes from the filter characteristics.

Spectral Filter Improvement

The primary colors in slide film, and to some extent other gelatin films, have very little "cross-talk" or spectral overlap. This is reasonable since the photographic process requires that primary spectral intensity components in a visual scene be absorbed sequentially in three emulsion layers. Low spectral overlap in color filters works favorably for color photography and color imagers but is detrimental to the proposed analog processing with our color sensors. Analog processing requires that the filter response of the differential color sensors have *large* spectral overlap. We can see the advantage of spectral overlap in the simulations of ideal Gaussian filters shown in Fig. 5.15. The figure indicates that increasing the spectral half-bandwidth or increasing the number of sensors will improve the hue sensor performance. In a simplified analysis in Appendix C, we show that for a 2 cone system the filters should

have overlapping spectra that crosses over at 32 to 80% of peak. Furthermore, the measurement of the derivative of hue with respect to wavelength, which should be as small as possible, scales as $\delta/((N-1)\sigma^2)$. Here δ is the separation of peak wavelengths between two adjacent spectral filters, σ is the Gaussian width and N is the number of spectral sensors.

In Fig. 5.15b we superimpose the calculated hue output from intrinsic diodes[13] which have wide-band overlapping spectral response curves as shown in Fig. 6.9. On the one hand, Fig. 5.15 shows that intrinsic diodes offer improved hue sensor performance over RGB filtered phototransistors. On the other hand, when we compare the simulated hue x-y photocurrents for intrinsic diodes shown in Fig. 5.16 to the same currents measured for RGB filtered sensors shown in Fig. 5.7, we notice that the spectral half-widths are wider for intrinsic diodes and thus offer less spectral sampling discrimination.

These two performance measures—determining monochromatic stimuli and sampling the spectrum—seem to conflict with one another. They both require color sensors that have overlapping spectral curves. But one requires wide-band curves and the other requires narrow-band curves. This apparent dichotomy will be further discussed in the next chapter, where we also describe improvements to intrinsic diode performance (Section 6.7) by creating many narrow-band overlapping spectral sensors.

[13]We hold off introducing the concept of intrinsic diodes until Chapter 6.

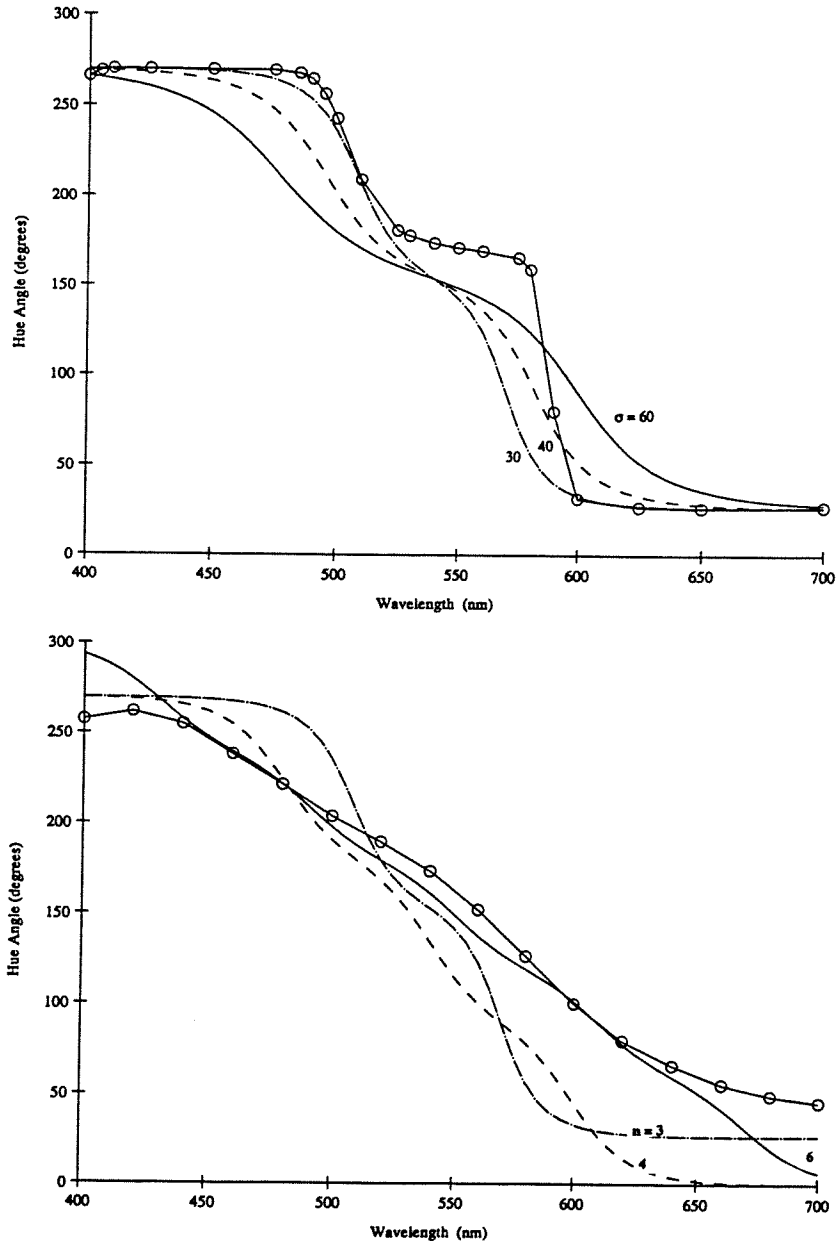


Figure 5.15: Comparison of calculated hue output with ideal Gaussian filter configurations (a) with increasing spectral widths σ, and (b) increasing number of Gaussian spectral sensors. In (a) we superimpose the hue output calculated from the RGB filtered phototransistors (o) as shown in Fig. 5.8. In (b) we superimpose the calculated hue output from intrinsic diodes (o) based on the spectral response curves from Fig. 6.9.

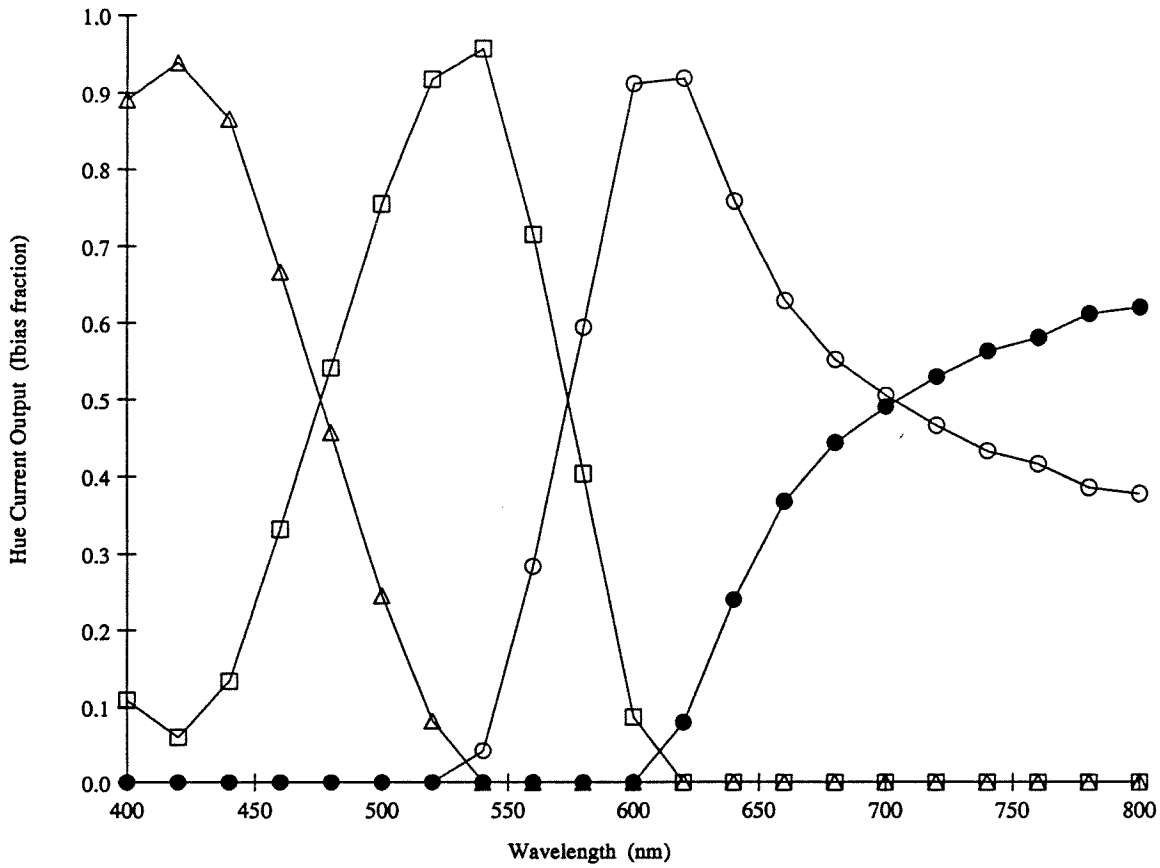


Figure 5.16: Simulated half-wave rectified hue currents $x^+(\bullet)$, $x^-(\square)$, $y^+(\circ)$, $y^-(\triangle)$ for intrinsic diodes based on spectral properties shown in Fig. 6.9. In comparison to Fig. 5.7, the x-y hue currents have wider half-widths and extend beyond the visible range. In particular, the x^+ current is only activated above 700 nm in the infrared region.

Chapter 6

Beyond Trichromacy: The Mantis Shrimp

Perhaps because of our anthropocentric tendencies, the literature places a special emphasis on analyzing and justifying the trichromatic sensor system. The fundamental question asked is "Why does the system have three sensors, why not less, why not more?" Justification for trichromacy is typically carried out *a posteriori* by evaluating the principal component basis functions that approximate natural illuminations and object reflectances. Other times it is carried out through analysis of human psychophysics, which ultimately depends on subjective "perception." In this chapter, we examine incorrectly applied communication theory arguments, and maintain that the trichromatic system is far from optimum in an information theoretic sense. Psychophysical data confirms this; the existence of multi-chromatic systems suggest this. We also examine biological systems that have more than 3 cones, and propose manufacturing an artificial system consisting of six spectral classes. Important concepts that we imitate from biology are signal normalization and the opponency computation. Our model is the Mantis shrimp. But first, we evaluate the arguments justifying

the trichromatic system.

6.1 Trichromacy: Sampling versus Encoding

The signal-processing arguments to justify trichromacy can be traced to Barlow who suggested a novel experiment to test for color sensitivity by using a "comb-filtered" spectrum[1] as stimulus (Barlow, 1982; Bowmaker, 1983). This stimulus is an extension of spatial vision sinusoidal gratings where frequency and amplitude modulation are used to evaluate contrast sensitivities. For color sensitivities frequency and phase are modulated in the "comb-filtered" spectrum.

Barlow's theoretical analysis is conducted in the Fourier Transform space. Tremendous simplification is afforded by the "comb-filtered" spectrum stimulus, since a sinusoidal variation in energy transforms to a single impulse in Fourier space. The modulation of the cone transforms by the "comb-filtered" impulse is a theoretical indication on how the system should respond. We assemble Barlows's analysis in Fig. 6.1, which shows the Fourier transform of the wavelength response of human cones, one of the Mantis shrimp photoreceptors, and the psychophysical response to "comb-filtered" stimulus.

Barlow then invokes the sampling theorem. The sampling theorem establishes the number of independent channels that can be constructed given the available spectral range and sensor spectral bandwidth. The formulation typically used is based on the Shannon-Whittaker Theorem: N, the number of independent samples possible in a segment of waveform of extent E and maximum frequency F is given by

[1]The spectrum is decomposed, for example by a prism, selectively filtered by a sinusoidal grating, and recombined. The term "comb-filtered" connotes placing a comb in between the decomposing and the recombining optics.

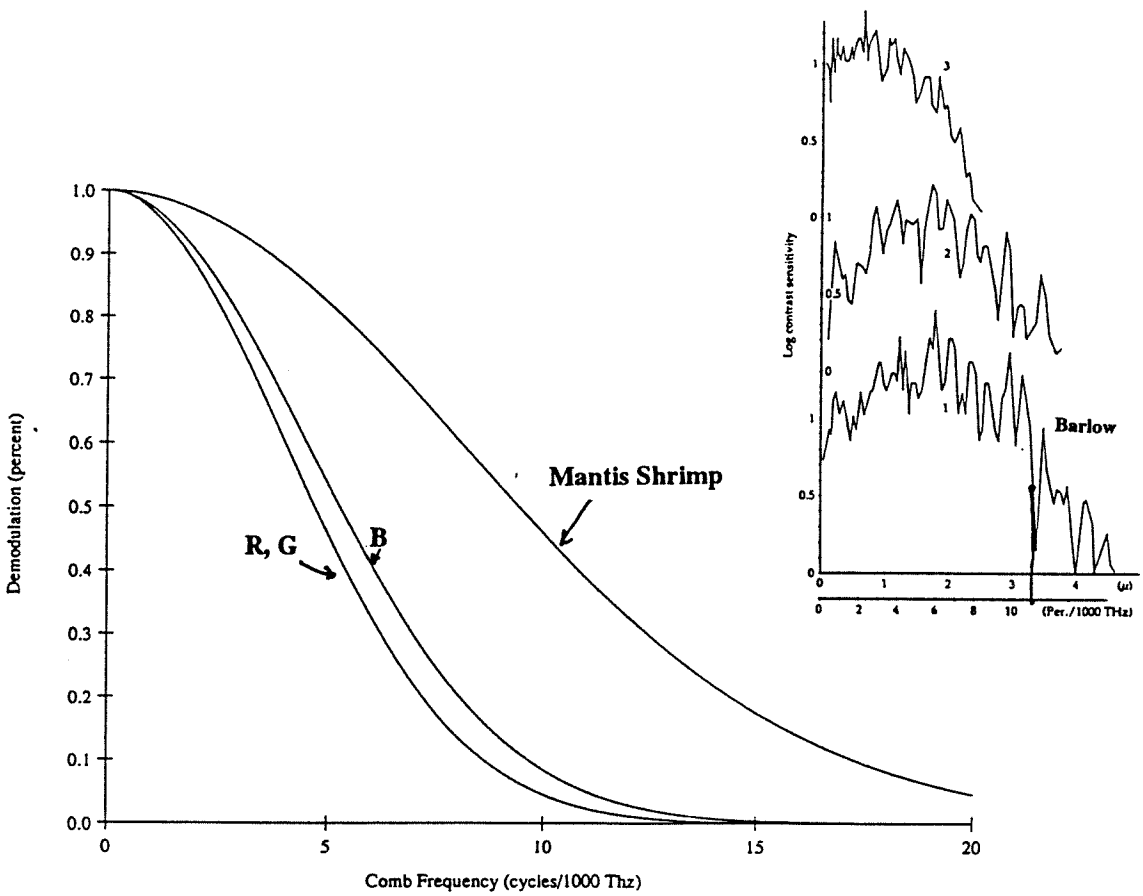


Figure 6.1: Fourier transform of the human receptor system based on a Gaussian approximation of the responses shown in Fig. 1.7. The Mantis shrimp response is a Gaussian approximation of the distal 2 row photoreceptor shown in Fig. 6.5. In the insert we copy the psychophysical response to "comb-filtered" stimuli (Barlow et al., 1983; Gemperlein et al., 1990). Curve 3 is the contrast sensitivity function for a deuteranope, while curves 1 and 2 are for color normal observers. In particular, Curve 1 is the psychophysical performance of Barlow.

$$N = 1 + 2EF \tag{6.1}$$

It should be noted that this statement of the theorem is approximate and only valid when N becomes large. (See also (Brill and Benzschawel, 1985).) The estimate on the error can be as much as (Brown, 1967; Butzer and Stens, 1982; Slepian, 1976).

$$\epsilon = \frac{2}{\pi} \int_{2\pi W}^{\infty} |F(\omega)| d\omega \tag{6.2}$$

where ϵ is the upper bound on the absolute error incurred in sampling a non-bandlimited signal at bandwidth W. $F(\omega)$ is the Fourier transform of the real function $f(t)$ that is to be sampled and reconstructed[2]. With the assumptions of Barlow and Bowmaker (50 % demodulation on the channel bandwidth) the error in reconstruction can be as much as 28 percent[3].

Barlow's analysis (Bowmaker, 1983) essentially works backwards. For trichromacy set $N = 3$. The range of human vision is 435 nm (689 THz) to 650 nm (461 THz), so set $E = 228$ THz. Therefore, the maximum frequency according to Equation 6.1 is F =4.4 cycles/1000 THz. Furthermore, this maximum frequency of 4.4 cycles/1000 THz is justified by an inspection of the Fourier transform of the human spectral sensitivities which shows a 50 to 60% demodulation at that frequency (Fig. 1, Bowmaker, 1983; Fig. 8, Barlow, 1982).

Although this appears as a justification for trichromatic systems, it cannot explain

[2]In our notation, the independent variable "wavelength" is interchangeable with "time."

[3]This estimate is based on reconstructing Gaussian shaped spectra of similar bandwidth to cone spectra. The value is obtained from the Fourier transform pair $(\exp(x^2), \frac{1}{\sqrt{2}} \exp(\frac{\omega^2}{4}))$ and evaluation of Equation 6.2.

the 11 cone Mantis shrimp. When we compare the data for the Mantis shrimp in Fig. 6.5 with human visual data (Fig. 1.7), the maximum half-width reduction is only a factor of 2. With $F = 8.8$ cycles/1000 THz, according to Equation 6.1 the Mantis shrimp should rightfully have only 5 spectral receptors, not 11.

This is a case of a misapplication of anthropocentric thinking. Maloney (1986) argues that for color constancy to work, the band limit of the photoreceptors must be greater or equal to the band limit of illuminant plus the band limit of the surface reflectance. "The appropriate frequency cutoff for each photoreceptor is dictated *not* by the number of photoreceptor classes but by the need to preserve information in the color signal corresponding to the light and surface spectral reflectances." Maloney reevaluates Barlow's data for photoreceptor bandwidths and predicts the number of independent channels that can be formed as 5 to 7.

But, the strongest argument against Barlow's theoretical analysis of trichromacy comes from himself—from psychophysical data (Barlow et al., 1983; Gemperlein et al., 1990) with "comb-filtered" spectra on himself. The insert in Fig. 6.1 shows that Barlow has color contrast sensitivity up to 10.5 cycles/1000 THz which corresponds to $N = 5$. Another color normal subject and a deuteranope (red/green color-blind) measured 11.0 cycles/1000 THz ($N = 6$) and 7.0 cycles/1000 THz ($N = 4$) respectively. (Computer simulations (Benzschawel et al., 1986) predict this bandpass drop-off for deuteranopes.) Obviously, trichromacy is not sacrosanct, and the human visual system has room for an additional two or three more spectral photoreceptors before the Nyquist sampling limit is reached.

Although they do not state it explicitly, the analysis of (Brill and Benzschawel, 1985; Buchsbaum and Gottschalk, 1984) complements the notion that the spectral bandlimit is "more than meets the eye." A figure that we borrow from their work, shown in Fig. 6.2, indicates that the gamut of signals band-limited to 5 cycles/μm (which corresponds to trichromacy in Barlow's analysis) represents a small section of

visual space. Fig. 6.2 hints that we seeing more than sampling theorem allows.

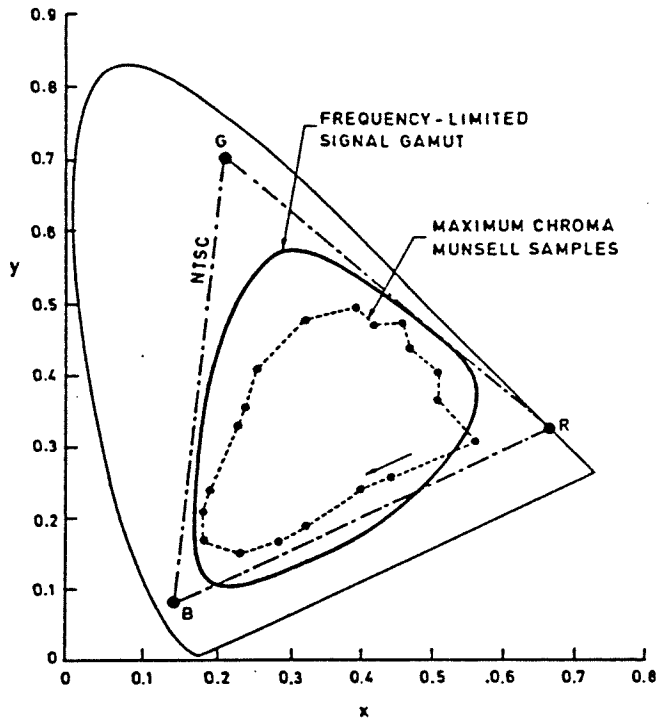


Figure 6.2: Signal gamuts within human visual space, copied from Fig. 1 in (Buchsbaum, 1985). While the frequency limited (5 cycles/μm) signal gamut encompasses all of the Munsell samples, and a significant portion of the color television gamut, it misses a significant portion of the CIE visual space.

Sampling theorem arguments, therefore, appear to directly conflict with psychophysical performance and the necessity for accurate spectral encoding. This dilemma is resolved if we focus on the goal of these seemingly analogous tasks of sampling and encoding as shown in Fig. 6.3. To sample accurately, more narrowly-tuned sensors are required; on the other hand, encoding can be achieved with broadly-tuned overlapping sensors responses.

Furthermore, *hyperacuity* analysis in the spatial domain (Baldi and Heiligenberg, 1988) indicate that for Gaussian type receptor fields the incertitude of spatial measurement, ϵ, scales as

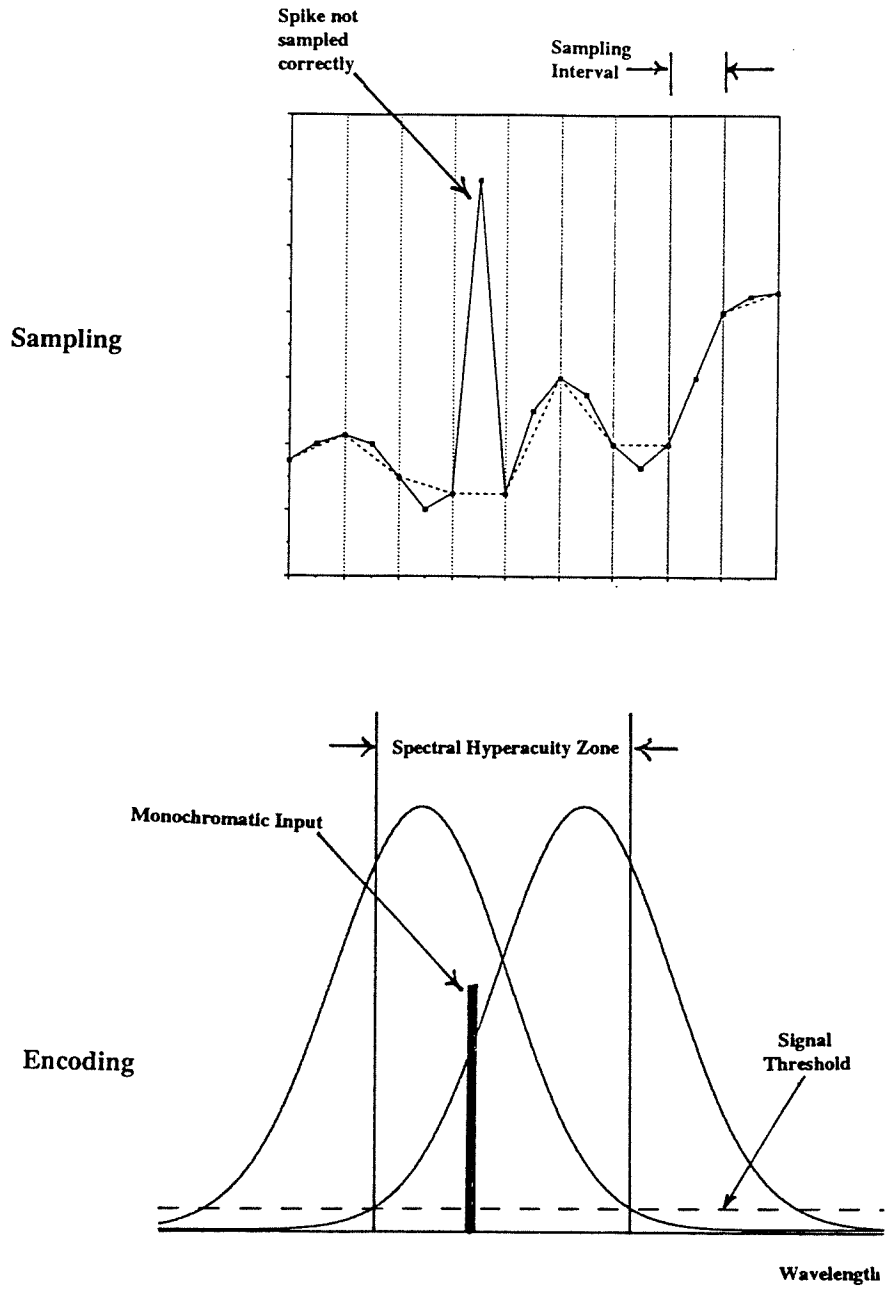


Figure 6.3: Sampling versus encoding. The number of channels needed to accurately sample is determined by the Nyquist frequency of the data. For this operation, the spectral response curves should be as narrow as possible. For accurate spectral encoding, on the other hand, only two overlapping broad-band spectral response curves are needed. The goal of sampling is to reconstruct the total waveform, while the goal of encoding is to determine the wavelength of a monochromatic input.

$$\epsilon = \frac{\eta}{\sqrt{2\pi\sigma}} \tag{6.3}$$

where η is an absolute error in the computation and σ is the Gaussian half-width. This indicates that higher resolutions are achieved with larger σ, i.e., larger curve widths. These results in the spatial domain are directly transferrable to the spectral domain.

In this section we argue that previous analysis (Barlow, 1982; Bowmaker, 1983) on trichromacy may be in error, probably a result of anthropocentric inclinations and incorrect data interpretation. If we extend the arguments of Barlow and Bowmaker to the Mantis shrimp we reach the conclusion that Nature gave the Mantis shrimp 6 more receptors than it could use. This turns out to be untrue. Nevertheless, the problems of sampling and encoding spectral information seem to have conflicting requirements in their solutions.

6.2 Trichromacy: Principal Components

For the most general case, accurately reconstructing a spectrum from the uniformly sampled data of that spectrum requires that the sampling interval be less than half the wavelength of the highest frequency present. This is a consequence of the Nyquist sampling limit. But the information overload of the uniform Nyquist sampling can be reduced if non-uniform sampling, which focuses in the region of interest, is considered. For example, in a "burst" type signal, it is intuitive that uniform sampling would not be the most efficient sampling scheme, over one that adjusted the sampling rate based on estimates of a "local" bandwidth (Clark et al., 1985).

The current fashion in justifying trichromacy is *a posteriori*: evaluate the prin-

cipal component bases functions that approximate natural illuminations and object reflectances, then declare that the visual system must correspond likewise.

An empirical study of surface spectral reflectances shows that Munsell color samples and spectral samples of natural surfaces can be approximated with a reasonable fit by a few (2-6) basis functions (Maloney, 1986; Dannemiller, 1992). The *a priori* use of these few basis functions allows for the efficient coding of object classes.

Even though the visual spectrum is sampled with N data points, these points form an N-dimensional space which is highly correlated. Principal component analysis eliminates these correlations and determines the best eigenvectors that span the N dimensional space. Within multispectral image processing, the determination of the principal components allows tremendous information compaction.

For example, to detect the presence of metal sulfides in soils, only 2 sensors are needed of the 512 spectral sensors operating in the 0.4 to 1.1 μm electromagnetic range. For this particular remote sensing airborne radiometer (Chang and Collins, 1983) task, discriminating the 7 to 10 nm blue shift in the chlorophyll red-edge of coniferous and deciduous trees detects the presence of sulfides. Here, selecting principal components by appropriate filter selection dramatically reduces the information overhead.

6.3 Color Systems Reviewed

Although most primates function quite satisfactorily in a bi-chromatic or tri-chromatic world within the 400 to 750 nm visible bandwidth, for some other biological systems this may constitute perceptual blindness. Some animals push this spectral bandwidth or extend it into other parts of the spectrum such as into the ultraviolet or the infrared. For example, flowers which appear to be yellow or varying shades of yellow to the human visual system can be mapped to blues, purples, reds, or yellows in the

honeybee visual system (Autrum and Thomas, 1973; Dusenbery, 1992). That the same object is perceived far differently for a honeybee than a human is understandable considering the pollinating/nectar-gathering requirements of that ecosystem. It is appropriate that evolution should pick the appropriate spectral range for vision and enhance important details there.

Furthermore, numerous systems sample the spectrum with more than 3 sensors. Insects are known to have 3 or 4 types of color receptor cells, the butterfly (*Papilio xuthus*) is shown to have 5 types of color sensors (Arikawa et al., 1987), and the turtle (*Pseudemys scipta elegans*) is shown to have 6 morphological types of cones (Ohtsuka, 1985). The current champion, as we have hinted in the previous sections, is the Mantis shrimp (*Gonodactylus oerstedii* and *Pseudosquilla ciliata*) with 11 spectral types of photoreceptors (Cronin and Marshall, 1989a)[4].

The compound eye of the Mantis shrimp has approximately 6,000 photosensors arranged in at least 36 ommatidial rows. It is roughly 4 mm in diameter and is divided into three sections: dorsal and ventral hemispheres and a specialized midband (Cronin and Marshall, 1989b; Marshall, 1988). These three sections contain overlapping visual fields, with each section operating independently of the other. The dorsal and ventral hemispheres are essential for occular tracking (Cronin et al., 1991; Cronin et al., 1992) and are thus involved in the visual analysis of form, motion, and position and possibly in rangefinding and controlling ocular scanning movements. Acting like a specialized fovea, the midband is involved in spectral and polarization visual analysis during ocular scanning[5]. Fig. 6.4 illustrates this architecture. The midband of the Mantis shrimp retina consists of 6 rows of ommatidia containing approximately

[4]At first glance, it seems that "lower" life forms need more spectral photoreceptors, suggesting that color is more important for "lower" than "higher" life forms. The resolution of this dilemma is to note that primates "recently" evolved from nocturnal animals. These multispectral sensing "lower" life forms have been in daylight far longer.

[5]Scanning motions typically occur in a 300 msec interval with velocities averaging 40 degrees/sec (Land et al., 1990). Observed peak motion was at 327 degrees/sec within a 10 degree visual field.

160*6=960 sensor elements aligned and stacked in spectrally unique one-dimensional rows. Four of these rows have two tiers of spectral photoreceptors, while the other 2 rows contain single photoreceptor types. Including the peripheral region, the Mantis shrimp contains $4 * 2 + 3 = 11$ types of spectral photoreceptors. With the scanning motion of its retina, the mantis shrimp figuratively paints the world with an 11-color primary system.

6.4 A Model for the Mantis Shrimp

In the Mantis shrimp, we propose that an opponent calculation is performed in the first 4 mid-band rows which contain two photoreceptors per row. Although electrophysiological data is not available for this animal, the opponent computation conjecture seems reasonable in view of the photoreceptor spatial proximity. These two spectral photoreceptors operate within a single ommatidium so they are actually spatially coincident, whereas other spatial dimensions include: intra-row ommatidia-ommatidia spacing of 20 μm, and midband row-row spacing of 60 μm. Fig. 6.5 shows the normalized spectral response of these eight unique spectral photoreceptors[6] in *Pseudosquilla ciliata*) (Cronin and Marshall, 1989a). Note that the half-widths (width of the spectrum where its value is half of its peak value) of the spectral response curves for the Mantis shrimp are slightly smaller than those for the human visual system shown in Fig. 1.7. An analysis of this situation apparently contradicts the conclusions of Barlow (1992).

Performing opponent and normalizing operations on the data of Fig. 6.5 in a manner similar to the hardware implementation of the hue sensor in the last chapter results in Fig. 6.6. The narrow spectral bins in the output response of these normalized

[6]These spectral response curves were measured using microspectrophotometry of frozen sections of dark-adapted retinas .

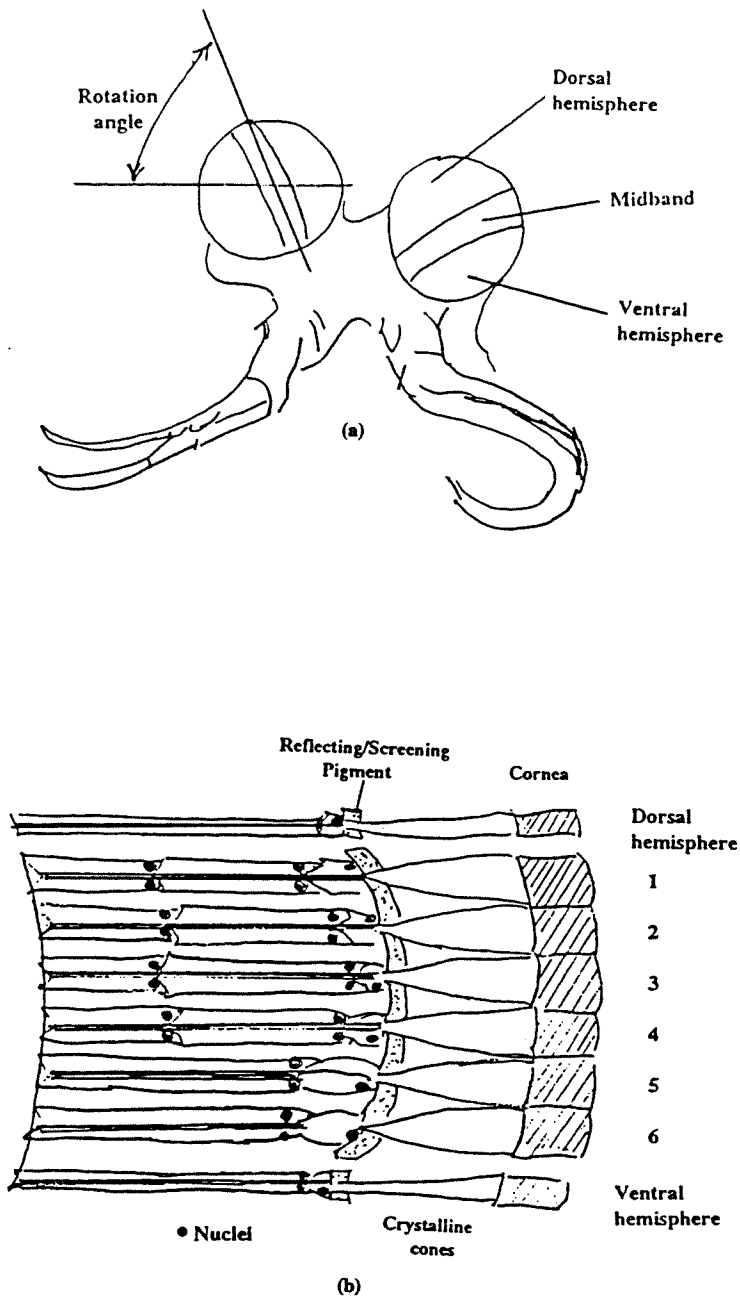


Figure 6.4: Concept of Mantis shrimp visual architecture adapted from (Land et al., 1990; Cronin and Marshall, 1989a) (a) Front view of Mantis shrimp. Eye stalk allows independent movement of each eye, which is roughly 4 mm in diameter. The midband ommatidial rows specialize in color and polarization vision. Triple overlapping of the visual field with the dorsal and ventral hemispheres, allows "monocular stereopsis." (b) Simplified diagram of the optic apparatus of the midband region. The first 4 distal rows have dual spectral tuning function.

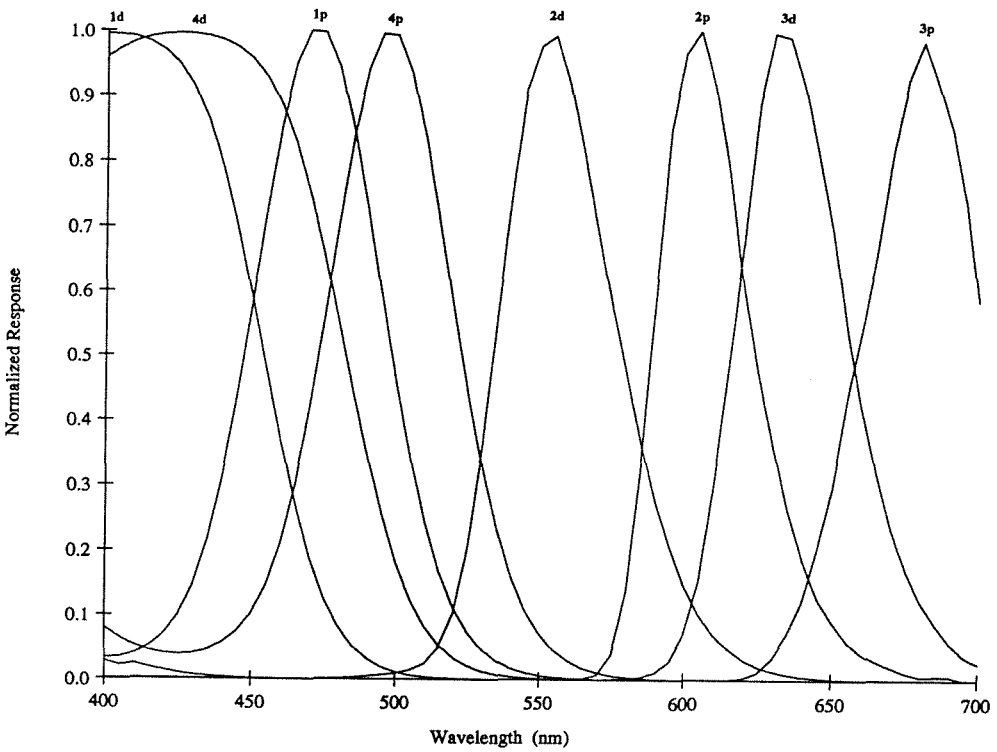


Figure 6.5: Spectral response curves of 8 putative opponent color receptors from the first 4 midband rows in the Mantis shrimp courtesy of Tom Cronin (Cronin and Marshall, 1989a). Labeling on each curve indicates the row and intra-row position of each photoreceptor. For example, "1d" represents distal 1st row photoreceptors while "4p" is the proximal 4th row photoreceptors. Note that the opponent receptors within the 1st and the 4th rows are not spectrally adjacent. Note that the half-widths of the spectral response curves for the Mantis shrimp is slightly less than that for the human visual system shown in Fig. 1.7. For example, the half-width of the proximal 1st row spectral response is 96, 63, and 49 percent of the short-, middle-, and long-wavelength photoreceptors respectively.

opponency cells indicate that the Mantis shrimp could be considered as a biological spectrum analyzer.

6.5 Importance of Opponency

When we review Chapters 2 and 3 of this thesis, we can assemble the advantages afforded by an opponency computation. In the linear-intensity representation, the subtraction operation removes the dc bias, the white light addition, of the spectrum and contributes what we call *additive shift* invariance. In the logarithmic-intensity representation, the subtraction operation compensates for uniform gain changes and contributes *multiplicative scale* invariance[7].

Furthermore, when we look at the numerous spectral sensory channels that convey visual information, the opponency computation offers a natural way to calculate the principal components that allow efficient information transfer (Buchsbaum and Gottschalk, 1983). The main idea here is that since their spectral response curves greatly overlap in the wavelength domain, the original RGB signals are highly correlated. The opponency computation discounts redundancies and enhances signal differences. It is for this reason that the YIQ system for color television works. Additionally the opponency computation in the human visual system has the unexpected benefit (Hurlbert, 1991; Lee, 1990; Lee, 1989) of aligning its chromatic basis vectors (yellow-blue) along the CIE daylight locus—a mechanism proposed to compensate for the changing natural illumination and thus achieve functional color constancy.

Still another advantage of the opponency operation is "spectral sharpening" conceptually illustrated in Fig. 6.7. Although we utilize idealized bandpass spectral

[7]Recall that in our hardware implementation, we use the linear-intensity representation achieving both *additive shift* and *multiplicative scale* invariance by subtraction and normalization operations respectively.

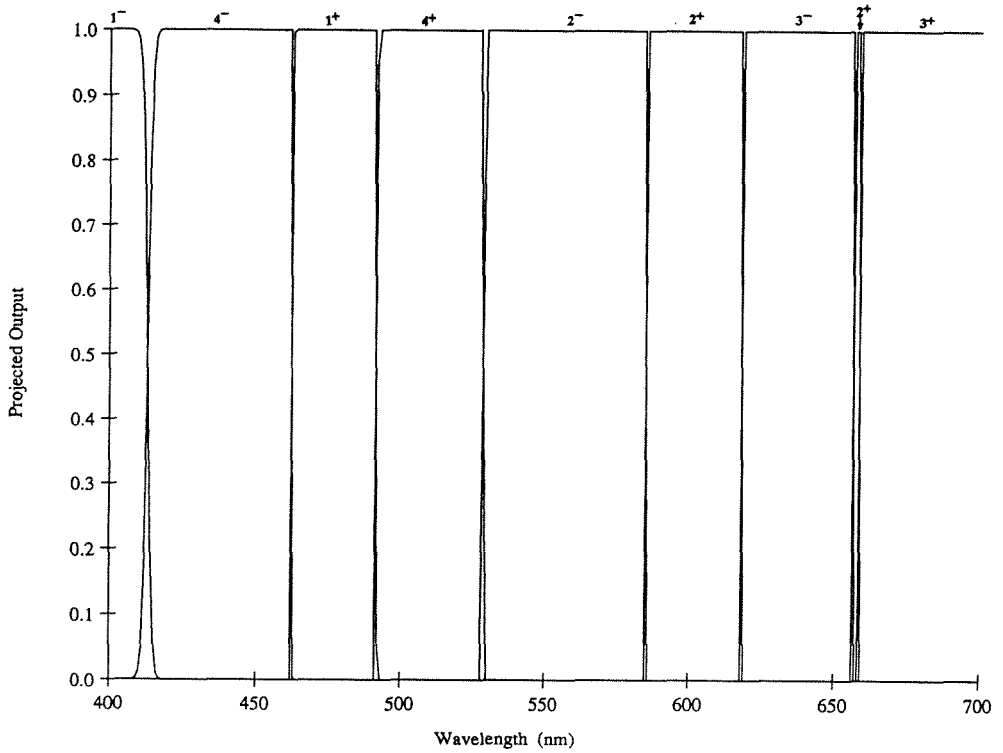


Figure 6.6: Projected spectral response curves for an artificial 8 cone color chip based on the filter response of the Mantis shrimp given in Fig. 6.5. Opponency and normalization operations are performed on this data in a manner similar to the hue chip in Chapter 5. The narrow bands of the spectral output response, which is independent of illumination levels, indicate that this a spectrum analyzer.

response curves for concept clarification, real spectral response curves have typical Gaussian shape but still provide spectral sharpening. In biology for example, although the half bandwidths of the input RGB cones are 50, 80, and 100 nm (Fig. 1.7), a majority of the color selective cells in V4 have mean half-bandwidth of 27 nm (Schein and Desimone, 1990). Also, our hardware example of Fig. 5.7 shows that the y^+ cell has accentuated spectral sharpening to 10 nm from filter inputs that have half-bandwidths of greater than 70 nm.

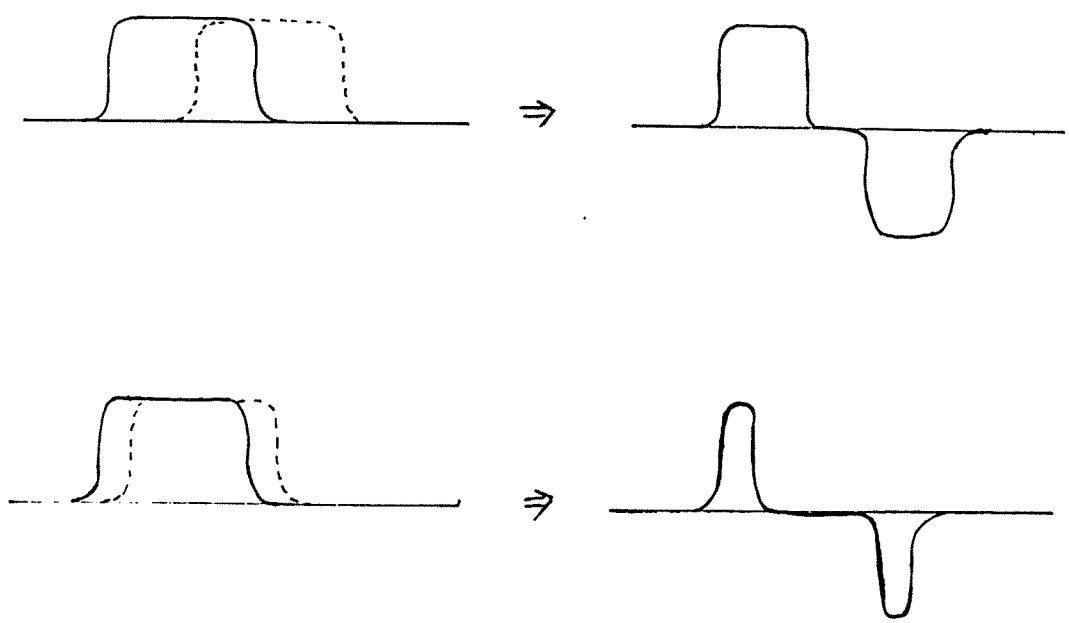


Figure 6.7: Spectral sharpening of spectral response sensors. Subtracting overlapping spectral response function results in a narrower spectral response curve. Increasing spectral overlap further sharpens the output spectral response.

6.6 Optimum Spectral Response Filter Design

The analysis of optimizing the information transfer by constructing other efficient channels starts from specified original channels—the red, green, and blue signals in a trichromatic system. But what happens if we have greater flexibility? If we were designing a color vision system with given spectral filter sensitivities, how should we choose to arrange them? From the previous sections, we learn that the opponent channels should have crossover point at a selected spectral wavelength important to the visual task. What remains is the separation between spectral sensitivities.

In Appendix C we calculate the shift required for two photoreceptor responses to maximize opponent spectral discrimination. For Gaussian type spectral responses of the same width, σ, a shift of 2σ would maximize the derivatives—maximize signal change—at the spectral crosspoint. For this ideal example, the optimum spectral crosspoint occurs at 61% of peak value. The Mantis shrimp appears to follow this rule in its opponency calculations in rows 1 and 4.

In the Japanese yellow swallowtail butterfly (Arikawa et al., 1987), the 5 color receptors appear to also have optimum spectral discrimination. Fig. 6.8 shows the spectral crosspoints occurring at approximately 70%, 40%, 60%, and 60% as the spectrum is swept.

Far from optimum, the human photoreceptor system shown in Fig. 1.7 crosses over at 97% of peak for the red/green and 21% of peak for the blue/green wavelength system. The accepted justification for this substantial overlap of the red/green system is related to the degradation of form vision and spatial acuity due to chromatic aberration effects for uncorrelated spectral curves. (Barlow, 1982; Mollon et al., 1990). For completeness the spectral half-widths are approximately 112 nm, 87 nm, and 57 nm for the red, green, and blue photoreceptor spectra respectively.

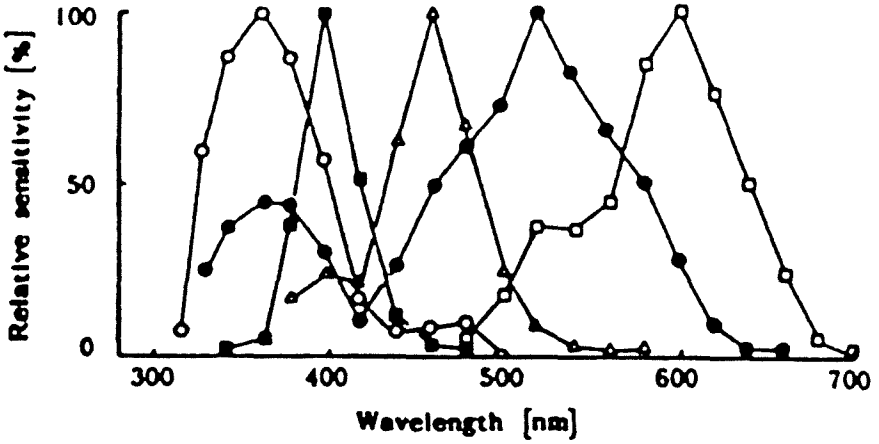


Figure 6.8: Spectral tuning response curves of Japanese yellow swallowtail butterfly (Fig. 1 in Arikawa et al., 1987).

6.7 Improving the Spectral Response of Intrinsic Devices

Intrinsic silicon devices such as buried diodes potentially could be used for spectral sensing because of the differential absorption of light as a function of silicon depth. Utilization of buried silicon diodes has the noteworthy advantage of ease of manufacture since these intrinsic devices are commonly available in standard microelectronics processing. Their use eliminates the extra post-processing manufacturing step that color dye or interference filter deposition requires.

But, based on the maximum spectral change criterion described above, intrinsic photodiodes and phototransistors available in silicon devices appear to be insufficient for high performance color vision. For example, Fig. 6.9 shows the spectral response curves for various silicon devices available from the MOSIS process (Delbrück, 1993). The figure shows that the spectral crossover points are not at the optimum 61% of

peak, but more importantly, the spectral half-widths may be too wide for accurate vision in the visible range.

However, other researchers have measured the color reproduction performance of these silicon devices (Kramer, 1993; Seitz et al., 1993). For example, the CIE color-rendering index of these buried-layer silicon devices is 87% of CCD color cameras. Recent development efforts show their applicability as sensors for determining toast completion for bread (Kramer, 1994). Another circuit under development utilizing these intrinsic silicon devices is a vision sensor that discounts shadow in images (Zucca, 1994) by utilizing the *multiplicative/scale* and *additive/shift* invariance properties discussed in Chapter 3.

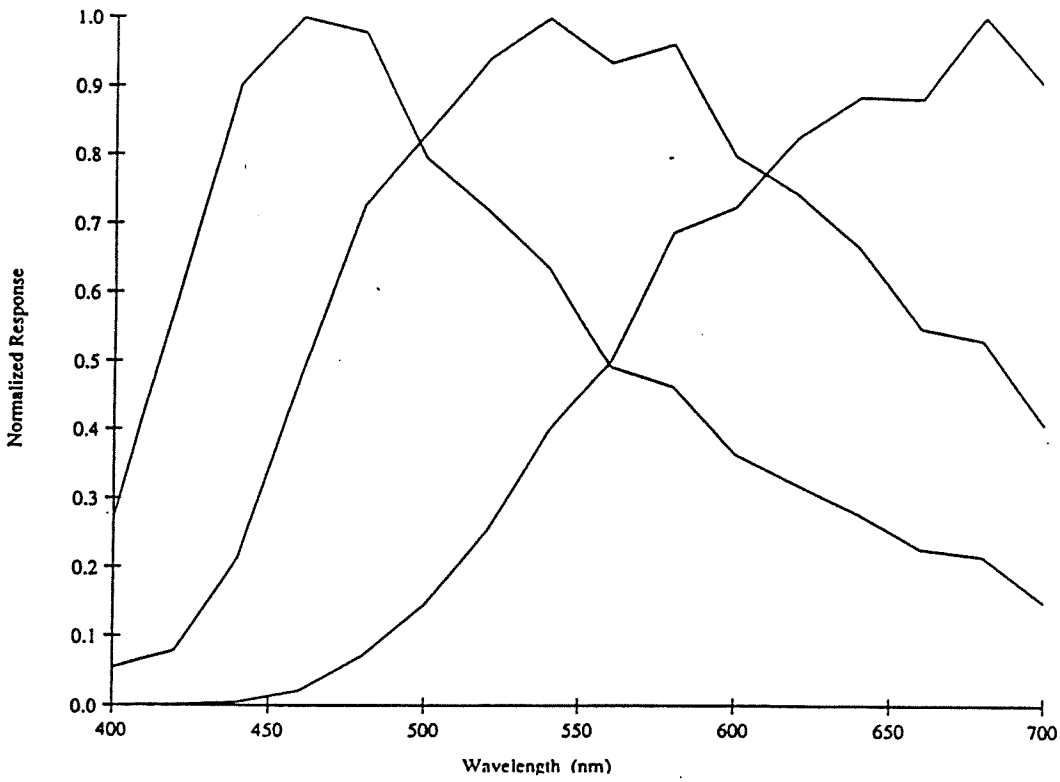


Figure 6.9: Spectral response of various intrinsic silicon devices as measured by Zucca (1994). Properties include (1) crossover points occurring at 75% for the red/green system and 81% for the green/blue system, (2) spectral half-widths of 220 nm for the green channel and 142 nm for the blue channel, and (3) red channel operating in the visible and extending into the infra-red.

Therefore, a re-evaluation of the intrinsic silicon diodes as spectral sensors is in order. On the one hand, the wide-band overlapping spectral response of intrinsic diodes offer improved performance of the hue sensor as shown in Fig. 5.15 in Chapter 5. On the other hand, the very same wide-band spectral response is detrimental in obtaining accurate spectral signatures[8]. This conflict is related to the encoding versus sampling issues summarized in Fig. 6.3. But, with the addition of multichromatic sensors, one can address both issues by using wide-band overlapping filters for accurate encoding of monochromatic signals, and many narrow-band filters for accurate spectral sampling. Filtering intrinsic diodes with color gels produces multichromatic sensors that achieve this goal.

Fig. 6.10 shows that the addition of 2 color gels results in the creation of 6 narrow-tuned overlapping spectral response curves in the blue-green region of the visual spectrum. Long wavelength or "red" gel filters have a sharp spectral drop-off[9] and do not contribute in creating a set of 3 overlapping spectral filters as the green and blue filters do in Fig. 6.10.

To get more overlapping filter responses in the long wavelength region, further research and developement is required for the fabrication of gel or die combinations that have reduced slope at the red filter drop-off wavelength. Another approach is to fabricate multilayer dielectric filters which offer high quality (high cost) custom spectral responses (Morf and Kunz, 1990). Nevertheless, the possibility exists for the creation of at least 10 unique spectral sensors from the addition of slide film over intrinsic silicon diodes: 3 wide-band spectral sensors, 6 narrow-band blue-green sensors, and 1 red spectral "long-pass" sensor.

[8]For example, the half-bandwidths shown in Fig. 5.16 for intrinsic diodes are much larger and more extensive than those of dye filtered phototransistors shown in Fig. 5.7.

[9]Additionally, we measured the spectral transmissivity of numerous red-orange-yellow gel filters from various manufactures (Kodak, Roscolux, Lee) and observed the same undesireable spectral drop-off in all cases.

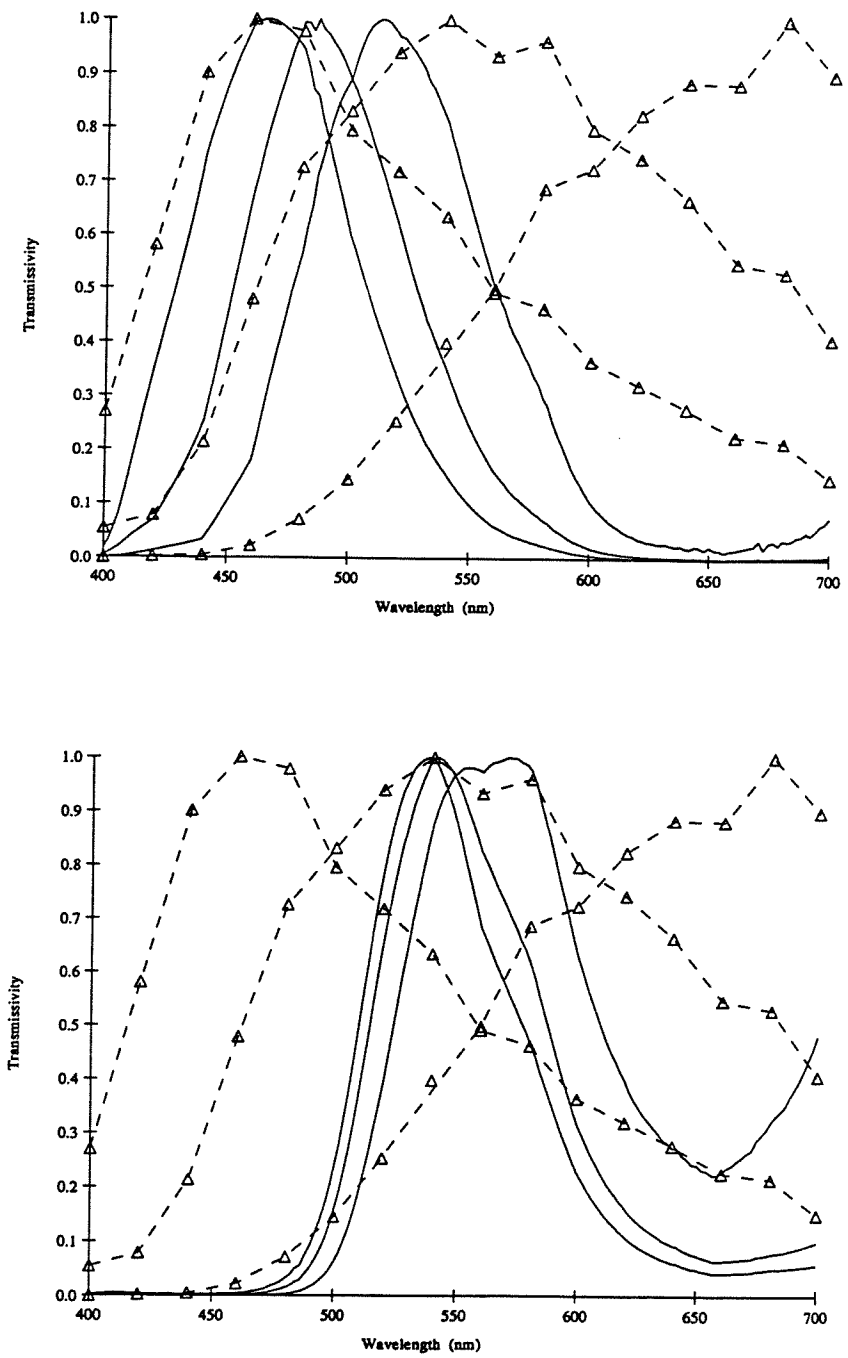


Figure 6.10: Spectral response of various intrinsic silicon devices of Fig. 6.9 sharpened with (a) blue $[0, 128, 255]$ and (b) green $[0, 255, 0]$ slide film. The dashed curves ($\triangle$) indicate the spectral response of the unfiltered silicon diodes while the solid lines show the sharpened spectral response of the diodes filtered by slide film.

6.8 Proposed 6 Spectral Sensor System in Analog VLSI

It is appropriate to design and manufacture multi-spectral systems greater than the trichromatic system. For color reproduction and correction, a 4 or 5 filter system is vastly superior to a 3 filter system (Vrhel and Trussell, 1994).

Multilayer dielectric filters offer the exact filter design for high quality sensors (Glass, 1990; Engelhardt and Seitz, 1993; Morf and Kunz, 1990; Seitz et al., 1993). Unfortunately, these techniques require deposition of multiple layers of optical materials after microelectronics processing with great impact on system cost.

We initially explore inexpensive spectral filtering technologies available in our research. In Fig. 6.11, we show the results of mixing color on readily available slide film. An effective shifting of the peak green response curve results as more red is added to the mixture. Ultimately, we expect the combination of dye deposition combined with intrinsic silicon devices (as shown in the previous section) will give improved performance at a substantial cost advantage.

Based on the above arguments on the Nyquist sampling limit and the available spectral sensitivity curves available from slide film shown in Fig. 6.12a, we design a multispectral image sensor consisting of 6 different spectral photoreceptors.

Utilizing the opponency and normalization operations on the spectral channels in the manner of the hue chip described in Chapter 5 will give an output spectral tuning curve as shown in Fig. 6.12. By manufacturing a one-dimensional array of these 6 photoreceptors, adding opponency and normalization operations, and utilizing scanning techniques, we can create a silicon analog of the Mantis shrimp color vision system. This proposed design offers "hyper-spectral" vision sensing.

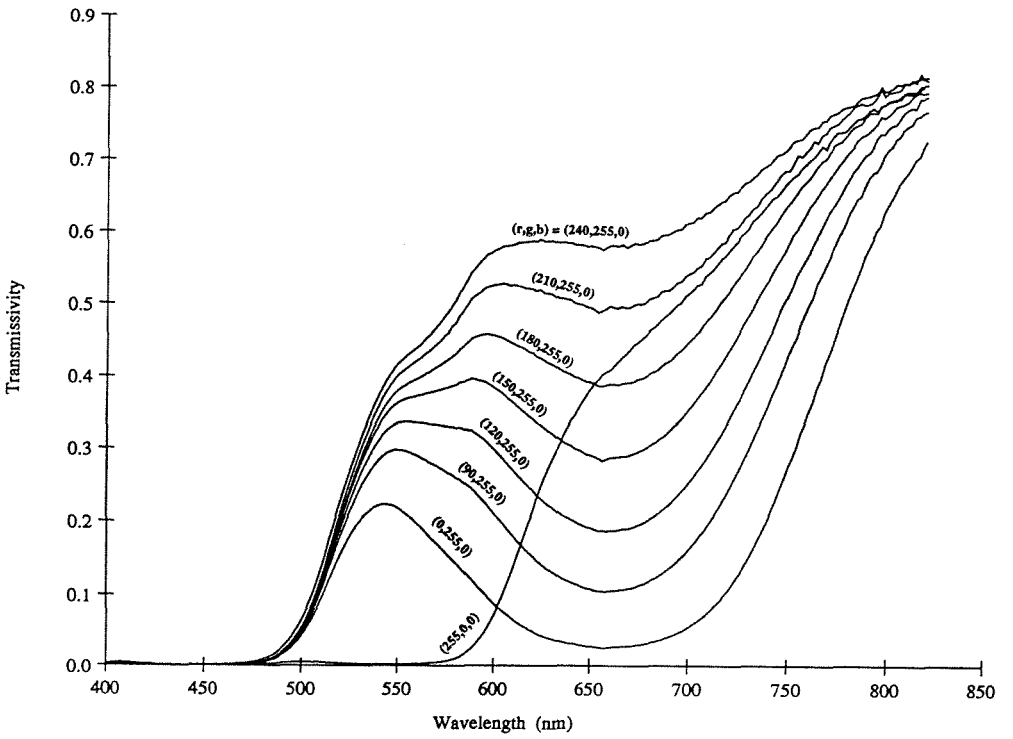


Figure 6.11: Various spectral transmissivity curves of Ektachrome 100 HC slide file with 100% green and various mixtures of red. The peak transmission curves shift to longer wavelengths as more red is added to the mixture. A bifurcation from green dominant to red dominant transmission spectra occurs between 47% (R,G,B=120,255,0) and 59% (150,255,0) red.

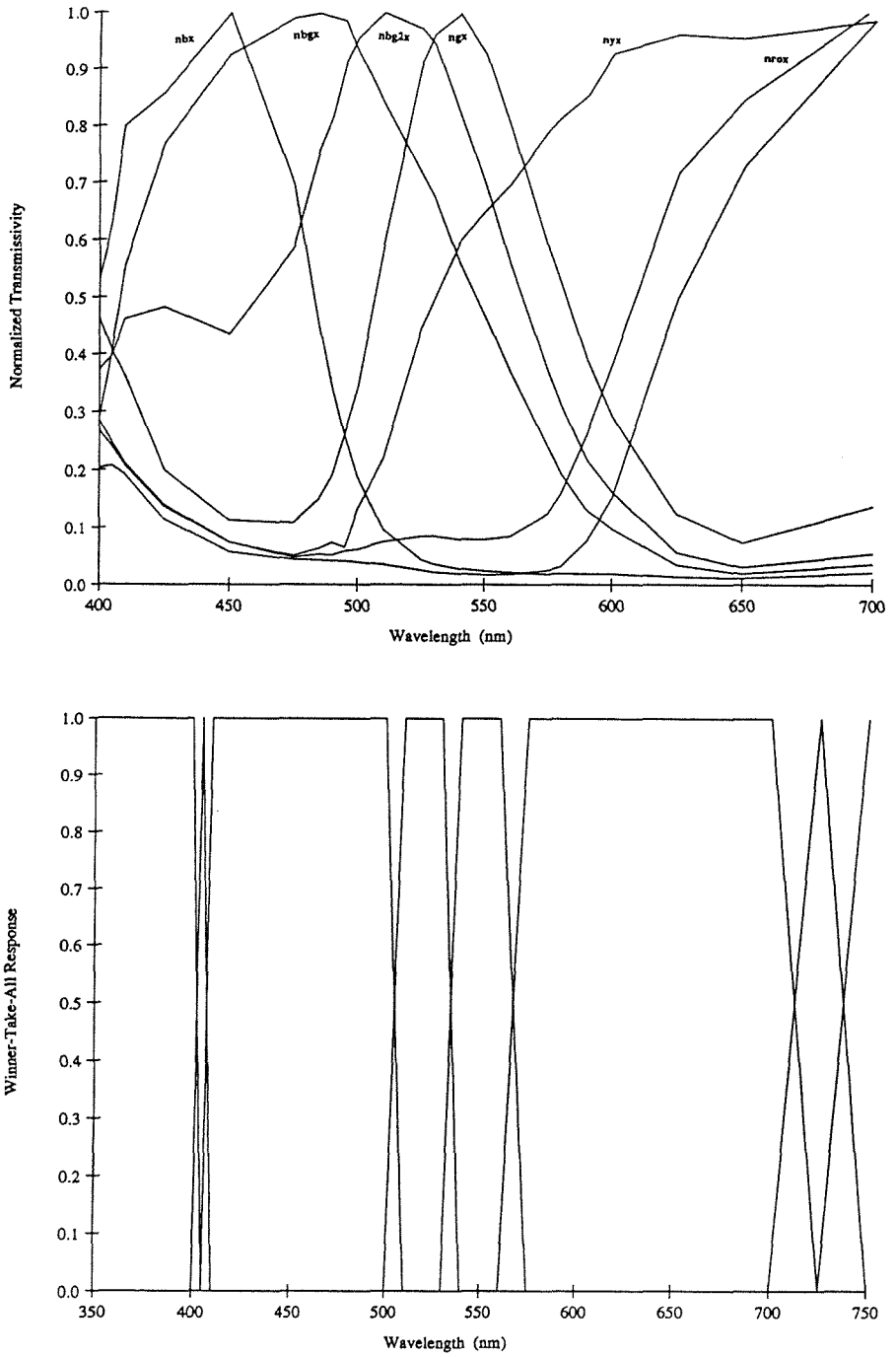


Figure 6.12: Artificial multispectral system (a) filter spectral transmissivity of Kodak Ektachrome 100 HC slide film at various color mixtures and (b) projected spectral response curves for opponent/normalization analog VLSI chip based on that filter response.

Chapter 7

Conclusions

In this study we engineer color vision subsystems based on hue and discontinuities in hue. We argue that edges in hue correlate more directly with material boundaries than edges in intensity, RGB, or Nrgb space. In particular, given the properties of *additive/shift* and *multiplicative/scale* invariance, hue edges are invariant to particular types of shadows, highlights, shading, and transparency. We illustrate this behavior using video-acquired color images. Segmentation in images containing low saturation and intensity image values can be improved with the help of a smoothing operation. Using an intensity edge map as a starting point and applying a first-order smoothness operator to the hue map results in an edge map which discounts confounding cues. In comparison to physically based models with their associated color clustering and histogramming, the method developed here is comparable in performance yet simpler in concept and execution. The main advantage of utilizing this operation is the straightforward application to analog VLSI hardware.

We also developed modulo operators to enhance traditional image processing algorithms for segmentation of variables defined on a circular space. For shadowed images, hue edges corresponded more closely to material boundaries than edge types in all

other color space measures. While modulo operators are elegant they are not feasible for analog VLSI implementation. However, they may have utilization in custom digital hardware.

We greatly favor the simple definition of hue employed in the HSI space over the definition of hue in the CIE spaces, since the former is relatively straightforward to implement in analog electronic integrated circuits. On that basis we have proceeded to build electronic test circuits to facilitate color processing. The first electronic integrated structure to be manufactured and tested was the Nrgb color sensor. The next electronic structure was the Hue color sensor. Tests on these sensors show correct performance to varying illumination. We also manufacture and test segmentation circuits useful in determining hue boundaries. With this hardware effort, we have the fundamental circuit building blocks necessary for a *smart* color vision sensor.

In designing artificial vision systems, there is some benefit to carefully selecting spectral filters such that the *integrated white* condition is satisfied. In doing so, the input spectral sensors will ultimately construct a hue space that is invariant to transparencies, highlights, shadows, and surface orientation.

Also, there is some advantage to controlling the spectral design of filter response curves. Increasing the filter spectral width so that different spectral filters have a greater degree of overlap in the spectrum increases the resolution for spectral encoding. The hue measurement performance, for example, was found to be enhanced by broadly overlapping spectral filters that crossover at 32% to 80% of peak value. But the performance enhancement that comes from utilizing broadly overlapping spectral filters comes at the price of penalizing spectral sampling reconstruction. Extension to the multispectral domain by increasing the number of spectral sensors will alleviate this concern.

For Gaussian shaped spectral response filters, the relative placement that maximizes signal change at the spectral crosspoint occurs when the spectral response

curves intersect at 61% of the peak value. This same filter design optimization appears to be adopted in four photoreceptors within the Mantis shrimp retina.

7.1 Future Directions

Even though the achromatic signal contains a majority of the signal energy, chromatic signals can enhance a visual system's performance by facilitating object segmentation and discounting illumination changes. This can be achieved without too much increase in hardware and complexity. In terms of information content in the visual environment and in terms of biological sensors, color information and color sensing utilizes few resources. In primates, based roughly on the cone-to-rod ratio, only 9% of the "front-end" visual sensors (16% in the Mantis shrimp) is for spectral sampling. Thus, to create a multichromatic (N) system, there is no need to increase the achromatic hardware by a factor of N. For example, one could use the subsampling tricks that biology and some television systems utilize for two-dimensional arrays or use the scanning of a one-dimensional array that the 11-cone Mantis shrimp exploits for higher spectral resolution. The manufacture of these devices in analog VLSI will have some speed, cost, power, and efficiency advantages over digital color CCD sensor systems for vision tasks.

To make useful devices, much work still needs to be pursued. For example, test performance on our preliminary sensors was based on inputs from monochromatic light sources. For commercial applications spanning inexpensive color printing controllers to color vision chips, the visual input should be more realistic than monochromatic lights. Inputs from polychromatic light sources and even illuminated natural objects should be examined. Altogether, this future work requires calibrated environments and thus more sophisticated equipment.

Conflicting issues of spectral discriminability versus spectral signature reconstruc-

tion must also be quantified experimentally. Perhaps custom spectral filter design or the addition of more spectral sensors will resolve these issues. In our push to multichromatic systems we implicitly assume that all spectral sensors have equal contribution to the color computation. Variants to this scheme including pyramid strategies that evaluate coarse to fine spectral hierarchies and other dynamic search strategies should be explored.

Also, the use of intrinsic diodes should be explored, especially when combined with color filters which sharpen the overall spectral response. Some of these concerns were addressed in Chapter 6 and in the "Future Hardware Issues" Section in Chapter 5. With the adoption of appropriate equipment and completion of characterization tests, the next task that remains is the integration of one- and two-dimensional color sensors and their inclusion into a complete vision system.

Further down the road, our color sensors can be integrated with color constancy hardware to compensate not only for varying illumination but for varying spectral illumination. The use of adaptation in color constancy algorithms may offer other benefits, especially in simplifying the filter/circuit design by automatically compensating for the non-normalized filter transmission characteristics. This use of adaptation for filter-balancing represents unexplored research territory. Ultimately, the combination of color segmentation and color constancy hardware with on-chip spectral sensors will produce a *smart color sensor* that maximally utilizes the spectral content of the visual environment.

Appendix A

Other Models for Discounting highlights

This appendix is a continuation of Section 3.4.2 and describes other computer vision models which consider highlights in images.

A.1 Cook-Torrance Model

The Cook-Torrance model is a more advanced model than the Phong model in that it incorporates a physically based model rather than empirical results to account for specularity or highlights. It is distinguished from the Phong model by the following factors (Watt, 1989):

- It is based on a consideration of incident energy rather than intensity
- The specular term is based on a physical microfacet model.
- Color change within the highlight is based on Fresnel's law and measured characteristics of the material.

These differences are enhanced when the light source is at a low angle of incidence to a material and the contribution from the Fresnel interaction becomes important. It is this effect that accounts for color change in a specular highlight and is not predicted by the Phong shading model.

The governing equation for the Cook-Torrance model for a single source illumination (Cook and Torrance, 1981) is given by

$$I_r = I_a R_a f + I_p \cos(\theta) d\omega_p R_{bd} \tag{A.1}$$

where I_r is the reflected intensity, I_a is the ambient intensity, R_a is the ambient reflectance, f is the fraction of the hemisphere not blocked, I_p is the light source intensity, θ is the angle between the normal and the light source (can also be defined by $\vec{N} \cdot \vec{L}$), $d\omega_p$ is the solid angle defined by *light source area*$/r^2$, and R_{bd} is the bidirectional reflectance which is segregated into specular and diffuse components and defined by

$$R_{bd} = f_s R_s + f_d R_d$$

$$f_s + f_d = 1$$

Of significance is the specular term R_s defined by

$$R_s = \frac{FDG}{\pi(\vec{N} \cdot \vec{V})(\vec{N} \cdot \vec{L})}$$

where G is a geometrical attenuation factor which accounts for shadowing and masking of one micro-facet by another. D is a facet slope distribution function that represents the fraction of the facets oriented in a particular direction. Cook and Torrance (Cook and Torrance, 1981) utilize the Beckman distribution function which is a function of geometry and surface roughness. Finally, F is the Fresnel term which

describes how light is reflected from each micro-facet. It is a complicated function of geometry, material properties, and wavelength.

Therefore, Equation A.1 can be written as

$$\overline{I_r} = \overline{I_a R_a} f + \overline{I_p} \cos(\theta) d\omega_p [f_d \hat{R}_d + f_s \hat{R}_s] \tag{A.2}$$

where $\overline{x}$ signifies that x is a function of wavelength λ, and $\hat{y}$ that y is a function of both wavelength and geometry. This equation can be used to give the tristimulus values. Namely,

$$K = \int_\lambda \overline{I_a S_K R_a} f d\lambda + \cos(\theta) d\omega_p f_d \int_\lambda \overline{I_p S_K} \hat{R}_d d\lambda + \cos(\theta) d\omega_p f_s \int_\lambda \overline{I_p S_K} \hat{R}_s d\lambda \quad \text{for } K = (R, G, B)$$

In the non-highlighted region $\hat{R}_s = 0$, therefore

$$K_{NH} = \int_\lambda \overline{I_a S_K R_a} f d\lambda + \cos(\theta) d\omega_p f_d \int_\lambda \overline{I_p S_K} \hat{R}_d d\lambda \quad \text{for } K = (R, G, B) \tag{A.3}$$

For the highlighted regions, the tristimulus equations become

$$K_H = K_{NH} + \cos(\theta) d\omega_p f_s \int_\lambda \overline{I_p S_K} \hat{R}_s d\lambda \quad \text{for } K = (R, G, B) \tag{A.4}$$

Thus, the equivalent *integrated white* condition for hue invariance due to highlights for the Cook-Torrance model becomes

$$\int_\lambda \overline{I_p S_R} \hat{R}_s d\lambda = \int_\lambda \overline{I_p S_G} \hat{R}_s d\lambda = \int_\lambda \overline{I_p S_B} \hat{R}_s d\lambda \tag{A.5}$$

If the Fresnel component is negligible (i.e., R_s is not a function of wavelength) the *integrated white* condition of Equation A.5 reduces to the generalized Phong shading

model given by Equation 3.10, that is,

$$\int_\lambda \overline{I_p S_R}\lambda = \int_\lambda \overline{I_p S_G}\lambda = \int_\lambda \overline{I_p S_B}\lambda$$

A.2 Dichromatic Model

The Dichromatic Reflection model was developed by (Shafer, 1985) to be used for extracting the specular and diffuse components of an image through its span in the color space. The model states that

$$I_r = I_s(\lambda, i, e, g) + I_d(\lambda, i, e, g) \tag{A.6}$$

where I_r is the reflected radiance which is composed of a specular component I_s and a diffuse component I_d (Shafer calls these interface and body reflections respectively), and i, e, g are geometric angles as shown in Figure 1.2. Furthermore, the Dichromatic Reflection model assumes that the spectral and geometric variations can be decomposed from each other. Namely that

$$I_r = m_s(i, e, g)C_s(\lambda) + m_d(i, e, g)C_d(\lambda) \tag{A.7}$$

This assumption of spectral and geometric decomposition will not be valid when the Fresnel contribution is not negligible such as when the illumination is at a low angle of incidence—a condition which Shafer states is not common. Otherwise, the Dichromatic Reflection model is a generalization of the Cook-Torrance model.

For the Dichromatic Reflection model the tristimulus values in the non-highlighted region become

$$K_{NH} = m_d(i, e, g) \int_\lambda C_d(\lambda)S_K(\lambda)d\lambda \quad \text{for } K = (R, G, B) \tag{A.8}$$

For the highlighted region the tristimulus values become

$$K_H = K_{NH} + m_s(i,e,g)\int_\lambda C_s(\lambda)S_K(\lambda)d\lambda \quad \text{for } K = (R,G,B) \tag{A.9}$$

Similar to previous findings, we see that hue invariance with confounding highlights will work exactly if and only if the *integrated white* condition holds, namely if

$$\int_\lambda C_s(\lambda)S_R(\lambda)d\lambda = \int_\lambda C_s(\lambda)S_G(\lambda)d\lambda = \int_\lambda C_s(\lambda)S_B(\lambda)d\lambda \tag{A.10}$$

This condition is similar to the Generalized Phong and the Cook-Torrance *integrated white* conditions. As a result, the analysis that ensued in their examination applies.

The result of this study shows that hue will discount highlights exactly for Phong shading models and approximately for more complicated physically based models. Still, hue is preferred over normalized color for discounting highlights.

Appendix B

Modulo Algorithms for Digital Hardware

Here a digital implementation for calculating modulo distances is described. The basic operation that calculates absolute distance in 2π modulo space is the following equation:

$$AbsoluteDist(x, y) = \min \begin{pmatrix} |y - x| \\ 2\pi - |y - x| \end{pmatrix} \tag{B.1}$$

A convenient implementation of this equation in 8-bit digital space would utilize the following kernel

$$\min(A, B) = (A - B)[(B - A) >> 7] + A \tag{B.2}$$

where $[x >> 7]$ represents a 7-bit shift and results in 0 for $x \geq 0$ and -1 for $x < 0$ in two's-complement digital architecture. To show that this relationship is true, define

$$\nabla\{A \geq B\} \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } A \geq B \\ -1 & \text{otherwise} \end{cases} \tag{B.3}$$

For 8-bit two's-complement digital representations Equation B.3 is equivalent to $(B - A) >> 7$ by a trivial inspection. That is

$$\nabla\{A \geq B\} \stackrel{\text{def}}{=} (B - A) >> 7 \tag{B.4}$$

Furthermore, we define in the usual manner $[A == B]$

$$[A == B] \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } A = B \\ 0 & \text{otherwise} \end{cases}$$

This relationship can be equivalently written as

$$[A == B] \stackrel{\text{def}}{=} -B\nabla\{A \geq B\} + -A\nabla\{B \geq A\} + A \tag{B.5}$$

By inspection,

$$\min(A, B) = -A\nabla\{A \geq B\} + -B\nabla\{B \geq A\} + A[A == B] \tag{B.6}$$

Substituting Equation B.4 and Equation B.5 into Equation B.6 results in the representation for calculating modulo distances as given by Equation B.2.

Thus, the equation that calculates relative modulo distance in 2π modulo space is given by

$$RelativeDist(x,y) = y - x + \begin{cases} +2\pi & \text{if } y - x < -\pi \\ -2\pi & \text{if } y - x > \pi \\ 0 & \text{otherwise} \end{cases} \tag{B.7}$$

The relative modulo distance is a measure of how much needs to be added to x to get to y. (The term "relative" is used to imply that the measure has a notion of direction which is contained in the sign of the variable.) Its equivalent digital implementation in 8-bit modulo space would give

$$\begin{aligned} RelativeDist(x,y) = y - x \quad & +256\{1 + [(x - y - 128) >> 7]\} \\ & -256\{1 + [(y - x - 128) >> 7]\} \end{aligned} \tag{B.8}$$

Equation B.2 and Equation B.8 have the desirable property that they do not rely on conditional operators. Their feedforward computational nature allows its use in nonlinear mask convolutions in custom-designed digital hardware.

Appendix C

Spectral Filter Design

In this appendix we approach spectral filter design through analysis of Gaussian shaped spectral response curves. We do this for mathematical tractability and ease of analysis. In the first section, we examine how filter design and filter quantity affects the hue computation. In the second section, we address optimum filter placement in the spectrum.

C.1 Filter Parameters that Affect Hue

This section presents a simplified analysis for Section 5.6.2. To streamline the problem, let us construct a hue space from a two cone system. Position the spectral response curves of the two cones at a set separation distance in the spectrum, and evaluate the effect of varying the widths of these spectral responses. Without loss of generality, position one cone, $Y(\lambda)$, at the origin and the other cone, $X(\lambda)$, at δ as shown in Fig. C.1. The plan is to examine the effect of modifying σ. That is

$$X(\lambda) = \exp(-\frac{(\lambda - \delta)^2}{2\sigma^2}) \tag{C.1}$$

$$Y(\lambda) = \exp(-\frac{(\lambda)^2}{2\sigma^2})$$

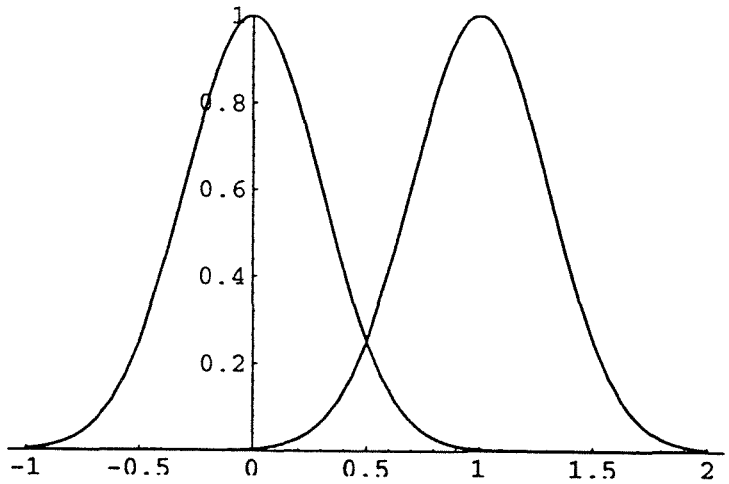


Figure C.1: Idealized two cone system, $X(\lambda)$ and $Y(\lambda)$, for the evaluation of hue. The spectrum of interest spans the range of $\lambda = [0, \delta]$. Here, the x-axis, λ, is in δ units.

If we limit the spectrum to $\lambda = [0, \delta]$ and define hue as in Equation C.2, it is clear that we are limiting the hue angle to the range of $[0, \pi/2]$.

$$Hue(\lambda) = Arctan\left[\frac{Y(\lambda)}{X(\lambda)}\right] \tag{C.2}$$

Examining this measure of hue at varying spectral sensor widths as shown in Fig. C.2 indicates that increasing the overlap (increasing σ) favorably reduces the slope of the hue output at the crossover point, but unfavorably decreases the effective span of the hue angle. At σ less than 0.333δ (or conversely δ greater than 3σ) the hue measure abruptly changes at the spectral crosspoint with large slope and flattens out at the spectrum boundaries contributing low discrimination and thus

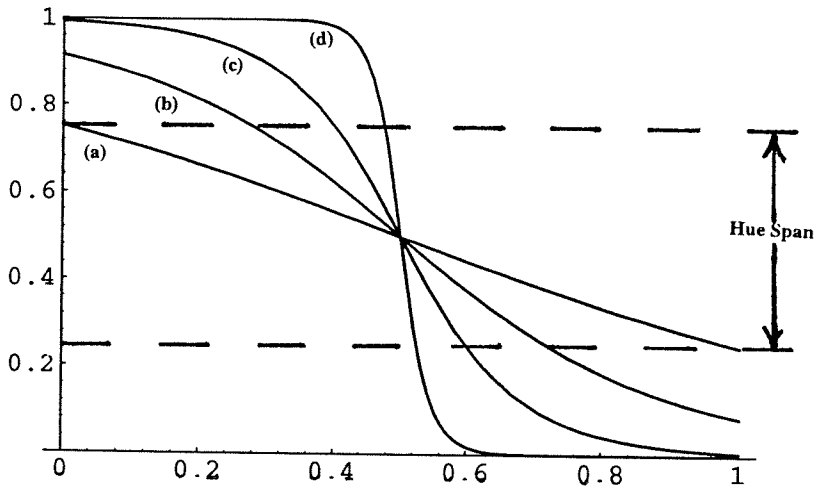


Figure C.2: Hues computed from a 2 cone system of varying spectral widths on the basis of Equation C.2. The curve shows the results when the Gaussian width σ is progressively reduced from (a) 1δ, (b) 0.5δ, (c) 0.33333δ, and (d) 0.166667δ. Note this is equivalent to $\delta =$ (a) 1σ, (b) 2σ, (c) 3σ, and (d) 6σ.

poor performance. This condition corresponds to having the spectral crosspoint at 32% or less of peak value for the Gaussian response curves.

From straightforward calculus, the slope of hue angle can be defined as

$$\frac{\partial}{\partial\lambda}\left[Hue(\lambda)\right] = -\frac{\delta}{\sigma^2}\left[\frac{\exp(\frac{\delta^2}{2\sigma^2} - \frac{\lambda\delta}{\sigma^2})}{1+\exp(\frac{\delta^2}{\sigma^2} - \frac{2\lambda\delta}{\sigma^2})}\right] \tag{C.3}$$

Arguing from symmetry, the maximum slope for hue occurs at the midpoint $\delta/2$. An evaluation gives the maximum hue slope, $MHS(\sigma)$

$$MHS(\sigma) = -\frac{\delta}{2\sigma^2} \tag{C.4}$$

Increasing σ too much, however, contributes to the offsetting hue span reduction of the hue measure. Therefore, it appears that there is some "optimum" range for σ.

Mathematically, the hue span reduction, $SR(\sigma) \in [0,1]$ can be defined as

$$SR(\sigma) = \frac{2}{\pi}\left[Arctan\left(\exp(\frac{1}{2\sigma^2})\right) - Arctan\left(\exp(-\frac{1}{2\sigma^2})\right)\right] \tag{C.5}$$

The behavior of $SR(\sigma)$ in Equation C.5 is shown plotted in Fig. C.3.

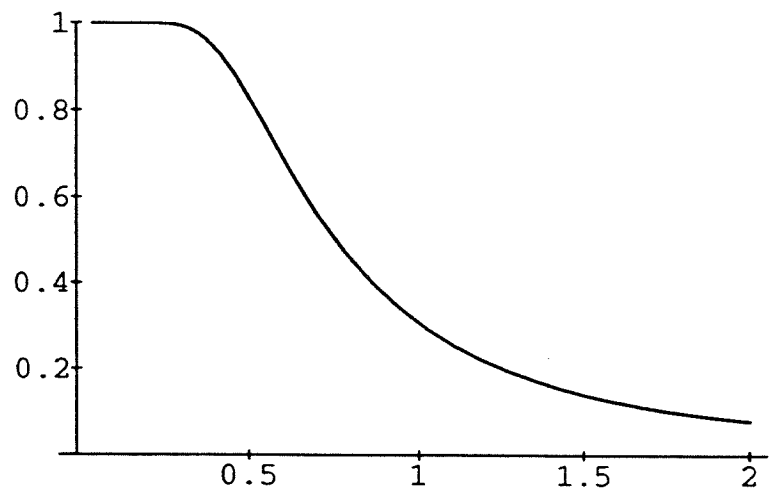


Figure C.3: Hues span reduction as a function of Gaussian width, σ in δ units. Reduction starts at about $\sigma = 0.333\delta$ (or conversely $\delta = 3\sigma$).

Both the maximum hue slope, $MHS(\sigma)$, and the hue span reduction, $SR(\sigma)$, places offseting bounds on the value of σ. However, we can construct a figure of merit, $FOM(\sigma)$ that incorporates these ideas together.

$$FOM(\sigma) = \frac{SR(\sigma)^P}{MHS(\sigma)} \tag{C.6}$$

In this function we chose to raise $SR(\sigma)$ to some power P greater than 1.0, since in the limit, $FOM(\sigma)$ will reach an asymptote and not an extrema. The reason for

this is that as σ gets large the Taylor expansion of $SR(\sigma)$ approaches $1/\sigma^2$ which is equivalent to $MHS(\sigma)$. The effect of setting P greater than 1.0 is equivalent to placing more emphasis on hue span reduction than on minimizing abrupt hue change. Fig. C.4 plots Equation C.6 for different values of P.

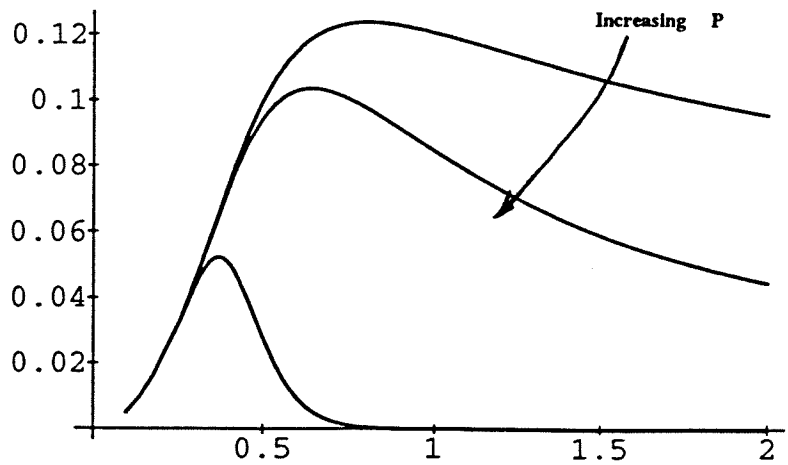


Figure C.4: Figure of merit as defined by Equation C.6 which maintains suitable hue span usage while minimizing abrupt hue changes. Increasing emphasis is placed on the hue span reduction by changing the parameter P in the figure.

From examination of Fig. C.2 and inspection of Fig. C.4, the appropriate spectral width appears to be in the range of $\sigma = 0.33$ to $0.75\ \delta$ (or $\delta = 1.33$ to $3\ \sigma$). This translates to spectral crossover points between 32% and 80% of peak values.

If the number of cones were increased to say N cones (N greater than 2) the cones can be distributed over the hue space, thus reducing the "workload" (hue span) for the original 2 cones system. For each 2 cone system the hue span is essentially reduced by the factor of $1/(N-1)$. For example, we can immediately modify Equation C.4 as

$$MHS(\sigma) = -\frac{\delta}{2(N-1)\sigma^2} \tag{C.7}$$

Therefore, for satisfactory hue measure the filter response of a 2 cone system should have overlapping spectral response with cross over point at somewhere between 32 to 80 percent of peak value. For conditions below that range, the consequences are abrupt hue angle transistions and flattened boundary regions that contribute to low hue discrimation. Above that range, the consequence is a reduction in the hue span and thus an ineffective use of hue space. Increasing the number of spectral sensors will decrease the span workload over the 2 cone system and improve hue discrimination and performance.

C.2 Optimum Separation of Gaussian Filters

A derivation of optimal spectral filter selection as mentioned in Section 6.6 is shown below. This simplified analysis is partially motivated by the observation that the opponency computation in multispectral biological systems such as the Mantis shrimp is not performed between cone types with the nearest spectral response. Also, from discussions on material changes from Chapter 3, it is clear that the objective function that should be maximized is the slope of the spectral opponent curve at the crossover point (Rubin and Richards, 1984).

First, assume that the cone spectral response, C_1, is a unit magnitude Gaussian centered about wavelength $\lambda = \lambda_0$. That is

$$C_1(\lambda) = \exp(-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}) \tag{C.8}$$

If another cone spectral response C_2, with the same σ, is shifted δ away from C_1 then the opponency operation of the *Difference of Offset Gaussians*, ($DofG = C_1 - C_2$), becomes

$$DofG(\lambda, \delta) = \exp(-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}) - \exp(-\frac{(\lambda - \lambda_0 - \delta)^2}{2\sigma^2}) \tag{C.9}$$

To find the required offset such that the slope of the $DofG$ is maximized entails evaluating the following equation for $\delta_{\max}$

$$\left(\frac{\partial}{\partial \delta}\left[\frac{\partial DofG(\lambda, \delta)}{\partial \lambda}\right]\right)_{\delta=\delta_{\max}} = 0 \tag{C.10}$$

To clarify these concepts we illustrate the $DofG(\lambda)$ and its derivative in Fig. C.5.

Evaluating Equation C.10 at the crossover point $\lambda = \lambda_0 + \frac{\delta}{2}$ gives

$$-\frac{1}{\sigma^2}\exp(\frac{-[\frac{\delta_{\max}}{2}]^2}{2\sigma^2}) + \frac{1}{\sigma^4}(\frac{\delta_{\max}}{2})^2\exp(\frac{-[\frac{\delta_{\max}}{2}]^2}{2\sigma^2}) = 0$$

Which further simplifies to $\delta_{\max} = 2\sigma$.

This result indicates that for Gaussian spectral response curves, the optimum curve should be shifted by 2σ from the baseline Gaussian spectral curve. The value of the spectral response at the crossover point, $\lambda = \lambda_0 + \frac{\delta_{\max}}{2}$, becomes $\exp(-\frac{\sigma^2}{2\sigma^2}) = \exp(-\frac{1}{2})$ or 61% of the peak value. This criterion is validated in the Mantis shrimp, where upon examination of Figure 6.5, the crossover points for the spectral response curves in rows 1 and 4 occur at 58% and 62% respectively of peak response.

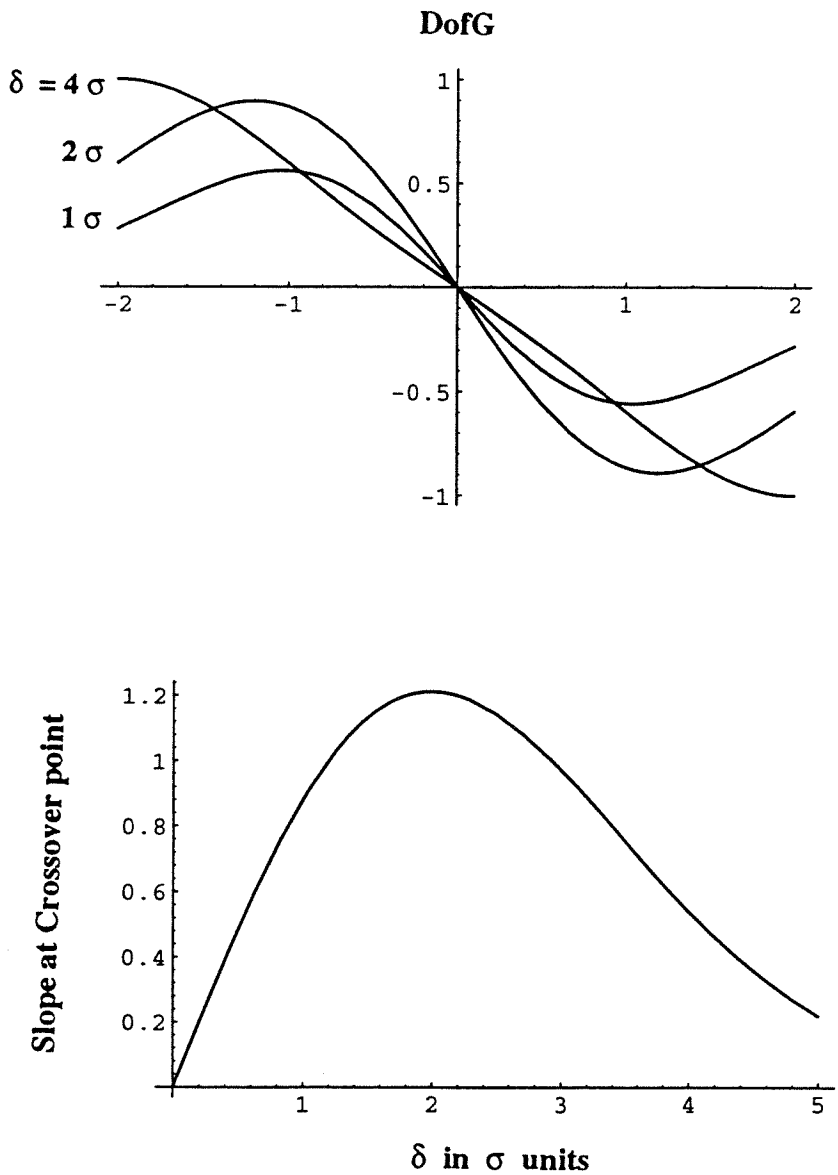


Figure C.5: (a) Difference of Offset Gaussians ($DofG$) for a given offset δ. As δ is increased, the slope at the crossover point (at zero) reaches a maximum, as shown in (b).

References

Abdou, I. and Pratt, W. (1979). Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proc. IEEE*, 67(5):753–763.

ACTON (1990). *Light Sources.* Acton Research Corporation, 525 Main Street, Acton MA, 01720.

Arikawa, K., Inokuma, K., and Eguchi, E. (1987). Pentachromatic visual system in a butterfly. *Naturwissenschaften*, 74:297–298.

Autrum, H. and Thomas, I. (1973). *Comparative Physiology of Colour Vision in Animals*, volume VII/3 of *Handbook of Sensory Physiology*, pages 661–692. Springer-Verlag.

Bajcsy, R., Lee, S., and Leonardis, A. (1990). Color image segmentation with detection of highlights and local illumination induced by inter-reflections. In *10th Intl. Conf. on Pattern Recognition B*, pages 785–790.

Baldi, P. and Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics*, 59:313–318.

Barlow, H. (1982). What causes trichromacy? a theoretical analysis using comb-filtered spectra. *Vision Research*, 22:635–643.

Barlow, H., Gemperlein, R., Paul, R., and Steiner, A. (1983). Human contrast sensitivity for comb-filtered spectra. *Journal Physiology*, 340:50P.

Barth, M., Parthasarathy, S., Wang, E., Hu, S., Hackwood, S., and Beni, G. (1986). A color vision system for microelectronics: Application to oxide thickness measurement. *IEEE International Conference on Robotics and Automation*, pages 1242–1247.

Bayer, B. (1976). Color imaging array. *U.S Patent (No. 3971065), Eastman Kodak Company.*

Benzschawel, T., Brill, M., and Cohn, T. (1986). Analysis of human color mechanisms using sinusoidal spectral power distributions. *Journal Optical Society of America, A*, 3(10):1713–1725.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution.* University of California.

Billmeyer, F. and Saltzman, M. (1981). *Principles of Color Technology.* John Wiley and Sons, 2nd edition.

Bliss, J. (1993). Theory and characteristics of phototransistors. In *Optoelectronics Device Data*, pages 11.13–11.23. Motorola.

Boahen, K. and Andreou, G. (1991). A contrast sensitive silicon retina with reciprocal synapses. In Moody, J., editor, *Advances in Neural Information Processing 4*, pages 764–772, San Mateo, CA. Morgan Kauffman.

Bowmaker, J. (1983). Trichromatic colour vision: Why only three receptor channels? *Trends is Neuro Sciences*, pages 41–43.

Brill, M. and Benzschawel, T. (1985). Remarks on signal-processing explanations of the trichromacy of vision. *Journal Optical Society of America, A*, 2(10):1794–1796.

Brown, J. (1967). On the error in reconstructing a non-bandlimited function by means of the bandpass sampling theorem. *Journal of Mathematical Analysis and Applications*, 18:75–84.

Buchsbaum, J. (1985). Chromaticity coordinates of frequency-limited functions: erratum. *Journal Optical Society of America, A*, 2:95.

Buchsbaum, J. (1987). Color signal coding: color vision and color television. *COLOR research and application*, 12(5):266–269.

Buchsbaum, J. and Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. Royal Society London, B*, 220:89–113.

Buchsbaum, J. and Gottschalk, A. (1984). Chromaticity coordinates of frequency-limited functions. *Journal Optical Society of America, A*, 1:885–887.

Butzer, P. and Stens, R. (1982). Sampling theory for not necessary band-limited functions: A historical overview. *SIAM Review*, 54(1):40–53.

Canny, J. (1983). Finding edges and lines in images. AI Memo 720, Massachusetts Institute of Technology.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.

Celenk, M. (1990). A color clustering technique for image segmentation. *Computer Vision, Graphics, and Image Processing*, 52:145–170.

Chang, S. and Collins, W. (1983). Confirmation of the airborne biogeophysical mineral exploration technique using laboratory methods. *Econ. Geol.*, 78:723–736.

Clark, J., Palmer, M., and Lawrence, P. (1985). A transformation method for the reconstruction of functions from nonuniformly spaced samples. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33(4):1151–1165.

Cook, R. and Torrance, K. (1981). A reflectance model for computer graphics. *Computer Graphics*, 15(3):307–316.

Cronin, T., Hong, Y., and Bidle, K. (1992). Regional specialization for control of ocular movements in the compound eyes of a Stomatopod Crustacean. *J. exp. Biology*, 171:373–393.

Cronin, T. and Marshall, N. (1989a). Multiple spectral classes of photoreceptors in the retinas of Gonodactyloid Stomatopod Crustaceans. *J. Comparative Physiology A*, 166:261–275.

Cronin, T. and Marshall, N. (1989b). A retina with at least ten spectral types of photoreceptors in a Mantis shrimp. *Nature*, 339(6220):137–140.

Cronin, T., Marshall, N., and Land, M. (1991). Optokinesis in gonodactyloid mantis shrimps (Crustacea;Stomatopoda;Gonodactylidae). *J. Comparative Physiology A*, 168:233–240.

Dahlquist, G. and Bjorck, A. (1974). *Numerical Methods.* Prentice-Hall.

Daily, M. (1989). Color image segmentation using Markov random fields. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 304–312.

Dannemiller, J. (1992). Spectral reflectance of natural objects: How many basis functions are necessary? *Journal Optical Society of America, A*, 9(4):507–515.

De Valois, R. and De Valois, K. (1975). *Neural Coding of Color*, volume V of *Handbook of Perception*, pages 117–166. Academic Press.

Delbrück, T. (1993). *Investigations of Analog VLSI Visual Transduction and Motion Processing.* Ph.D. thesis, California Institute of Technology.

Desimone, R., Schein, S., Moran, J., and Ungerleider, L. (1985). Contour, color, and shape analysis beyond the striate cortex. *Vision Res.*, 25:441.

DeWeerth, S. (1991). *Analog VLSI Circuits for Sensorimotor Feedback.* Ph.D. thesis, California Institute of Technology.

Dillon, P., Brault, A., Horak, J., Garcia, E., Martin, T., and Light, W. (1978a). Fabrication and performance of color filter arrays for solid-state imagers. *IEEE Trans. Electron Devices*, ED-25(2):102–107.

Dillon, P., Lewis, D., and Kaspar, F. (1978b). Color imaging system using a single CCD area array. *IEEE Trans. Electron Devices*, ED-25(2):97–101.

D'Luna, L. and Parulski, K. (1991). A system approach to custom VLSI for a digital color imaging system. *IEEE Journal of Solid-State Circuits*, 26(5):727–737.

Dusenbery, D. (1992). *Sensory Ecology.* W.H. Freeman and Company.

D'Zmura, D. and Lennie, P. (1986). Mechanisms of color constancy. *Journal Optical Society of America, A*, 3(10):1662–1672.

Engelhardt, K. and Seitz, P. (1993). Optimum color filters for CCD digital cameras. *Applied Optics*, 32(16):3015–3023.

Feinstein, D. (1988). The hexagonal resistive network and the circular approximation. Technical report, Caltech-CS-TR-88-7, California Institute of Technology.

Foley, J., Van Dam, A., Feiner, S., and Hughes, J. (1990). *Computer Graphics: Principles and Practice.* Addison-Wesley.

Geiger, D. and Girosi, F. (1990). Parallel and deterministic algorithms from MRF's: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):401–412.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gemperlein, R., Paul, R., and Steiner, A. (1990). Fourier interferometric stimulation (FIS): the method and its applications. In Blakemore, C., editor, *Vision: Coding and Efficiency*, chapter 11, pages 142–149. Cambridge University Press.

Genz, S. (1990). Real time chip set simplifies color image processing. In *Image Processing Handbook*, pages 56–59. Data Translation.

Gershon, R. (1985). Aspects of perception and computation in color vision. *Computer Vision, Graphics, and Image Processing*, 32:245–278.

Gershon, R., Jepson, A., and Tsotsos, J. (1986). Ambient illumination and the determination of material changes. *Journal Optical Society of America, A*, 3(10):1700–1707.

Gilbert, B. (1975). Translinear circuits: A proposed classification. *Electronic Letters*, 11(1):14–16.

Gilbert, B. (1984). A 16-channel array normalizer. *IEEE Journal of Solid-State Circuits*, 19:956–963.

Gillespie, D. and Lazzaro, D. (1990). *An Introduction to AnaLOG.* California Institute of Technology. Online manual associated with AnaLOG, a CAD VLSI public domain software.

Glass, A. (1990). Optimizing the performance of spectrally selective photodiodes by simulated annealing techniques. *Sensors and Actuators*, A21-A23:564–569.

Harris, J. (1990a). Analog hardware for detecting discontinuities in early vision. *International Journal of Computer Vision*, 4:211–233.

Harris, J. (1990b). A two-dimensional analog VLSI circuit for detecting discontinuities in early vision. *Science*, 248:1209–1211.

Harris, J. (1991). *Analog Models for Early Vision.* Ph.D. thesis, California Institute of Technology.

Healey, G. (1989). Using color for geometry-insensitive segmentation. *Journal Optical Society of America, A*, 6(6):920–937.

Healey, G. and Binford, T. (1987). The role and use of color in a general vision system. In *DARPA-Image Understanding Workshop*, pages 599–613, Los Angeles, California.

Heider, E. and Olivier, D. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 3:337–354.

Horiuchi, T., Bishofberger, B., and Koch, C. (1994). An analog VLSI saccadic eye movement system. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances*

in Neural Information Processing Systems 6, pages 582–589. Morgan Kaufmann Publishers.

Horn, B. (1986). *Robot Vision.* The MIT Press.

Hunt, R. (1987). *The Reproduction of Colour in Photography, Printing, and Television.* Fountain Press, 4th edition.

Hurlbert, A. (1989). *The Computation of Color.* Ph.D. thesis, Massachusetts Institute of Technology.

Hurlbert, A. (1991). Deciphering the colour code. *Nature*, 349:191–193.

Hurlbert, A. and Poggio, T. (1989). A network for image segmentation using color. In *Advances in Neural Information Processing Systems I*, pages 297–304. Morgan Kaufmann.

Hurvich, L. (1981). *Color Vision.* Sinauer Associates.

Hurvich, L. and Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64(6):384–404.

Ingling, C. and Tsou, B. (1977). Orthogonal combination on the three visual channels. *Vision Research*, 17:1075–1082.

Jain, A. (1989). *Fundamentals of Digital Image Processing.* Prentice-Hall.

Joblove, G. and Green, D. (1978). Color spaces for computer graphics. *Computer Graphics*, 12(3):20–25.

Kauzmann, W. (1957). *Quantum Chemistry.* Academic Press. pgs. 582, 669-671.

Kender, J. (1976). Saturation, hue, and normalized color: Calculation, digitization, and use. Technical report, Carnegie-Mellon University.

Khosla, R. (1992). From photons to bits. *Physics Today*, pages 42–49.

Klinker, G., Shafer, S., and Kanade, T. (1988). The measurement of highlights in color images. *International Journal of Computer Vision*, 2:7–32.

Klinker, G., Shafer, S., and Kanade, T. (1990). A physical approach to color image understanding. *International Journal of Computer Vision*, 4:7–38.

Koch, C. (1989). Seeing chips: Analog VLSI circuits for computer vision. *Neural Computation*, 1:184–200.

Koch, C., Moore, A., Bair, W., Horiuchi, T., Bishofberger, B., and Lazzaro, J. (1991). Computing motion using analog VLSI vision chips: An experimental comparison among four approaches. In *IEEE Workshop on Visual Motion*, pages 312–324.

KODAK (1981). Kodak Filters for Scientific and Technical Uses. Technical Report B-3, Eastman Kodak Company. ISBN 0-87985-282-2.

Kramer, J. (1993). *Photo-ASICs: Integrated Optical Metrology Systems with Industrial CMOS Technology.* Ph.D. thesis, Swiss Federal Institute of Technology, Zurich.

Kramer, J. (1994). Personal Communication.

Land, M., Marshall, N., Brownless, D., and Cronin, T. (1990). The eye-movements of the mantis shrimp Odontodactylus Scyllarus (Crustacea: stomatopoda). *J. Comparative Physiology A*, 167:155–166.

Lazzaro, J. (1990). *Silicon Models of Early Audition.* Ph.D. thesis, California Institute of Technology.

Lazzaro, J., Ryckebusch, S., Mahowald, M., and Mead, C. (1988). Winner-take-all networks of O(N) complexity. In Tourestzky, D., editor, *Advances in Neural Information Processing Systems 1*, pages 703–711. Morgan Kaufmann Publishers.

Lee, H. (1989). A computational model of human color encoding. Unpublished.

Lee, H. (1990). A computational model for opponent color encoding. In *SPSE's 43 Annual Conference*, pages 178–181. Society for Imaging Science and Technology.

Lee, S. (1991). *Understanding of Surface Reflections in Computer Vision by Color and Multiple Views.* Ph.D. thesis, University of Pennsylvania.

Lenny, P. and D'Zmura, M. (1988). Mechanisms of color vision. *CRC Critical Reviews*, 3(4):333–400.

MacKay, D. (1992). *Bayesian Methods for Adaptive Models.* Ph.D. thesis, California Institute of Technology.

Mahowald, M. (1992). *VLSI Analogs of Neuronal Visual Processing: A Synthesis of Form and Function.* Ph.D. thesis, California Institute of Technology.

Mahowald, M. and Douglas, R. (1991). A silicon neuron. *Nature*, 354:515–517.

Maloney, L. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal Optical Society of America, A*, 3(10):1673–1683.

Maloney, L. and Wandell, B. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal Optical Society of America, A*, 3(1):29–33.

Marr, D. (1982). *Vision.* Freeman.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proc. of the Royal Society of London B*, 207:187–217.

Marshall, N. (1988). A unique colour and polarization vision system in mantis shrimps. *Nature*, 333:557–560.

Mead, C. (1989). *Analog VLSI and Neural Systems.* Addison-Wesley.

Mersereau, R. (1979). The processing of hexagonally sampled two-dimensional signals. *Proceedings of the IEEE*, 67(6):930–949.

Mollon, J., Estevez, O., and Cavonius, C. (1990). The two subsystems of colour vision and their roles in wavelength discrimination. In Blakemore, C., editor, *Vision: coding and cfficiency*, chapter 11, pages 119–131. Cambridge University Press.

Moore, A. (1992). *Spatial Filtering in Tone Reproduction and Vision.* Ph.D. thesis, California Institute of Technology.

Moore, A., Allman, J., and Goodman, R. (1991). A real-time neural system for color constancy. *IEEE Trans. on Neural Networks*, 2(2):237–247.

Morf, R. and Kunz, R. (1990). Dielectric filter optimization by simulated thermal annealing: a simulated zone-melting approach. In Herrmann, R., editor, *Optical Thin Films and Applications*, pages 11–17. SPIE Vol. 1270.

Nassau, K. (1983). *The Physics and Chemistry of Color: The Fifteen Causes of Color.* John Wiley and Sons.

Nathans, J., Thomas, D., and Hogness, D. (1986). Molecular genetics of human color vision: the genes encoding blue, green and red pigments. *Science*, 232:192–202.

Nevatia, R. (1977). A color edge detector and its use in scene segmentation. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(11):820–826.

Ohlander, R. (1976). *Analysis of Natural Scenes.* Ph.D. thesis, Carnegie Mellon University.

Ohta, Y., Kanade, T., and Sakai, T. (1980). Color information for region segmentation. *Computer Graphics and Image Processing*, 13:222–241.

Ohtsuka, T. (1985). Relation of spectral types of oil droplets in cones of turtle retina. *Science*, 229:874–877.

Parulski, K., D'Luna, L., Benamati, B., and Shelley, P. (1992). High-performance digital color video camera. *Journal of Electronic Imaging*, 1(1):35–45.

Perez, F. and Koch, C. (1992). Toward color image segmentation in analog VLSI. In *Rockwell 14th Annual Technical Conf. on Control and Signal Processing*, pages 246–263.

Perez, F. and Koch, C. (1994). Toward color image segmentation in analog VLSI: Algorithm and hardware. *International Journal of Computer Vision*, 12(1):17–42.

Phong, B. (1975). Illumination for computer generated pictures. *Comm. of the ACM*, 18(6):311–317.

Poggio, T., Gamble, E., and Little, J. (1988). Parallel integration of vision modules. *Science*, 242:436–440.

Poggio, T. and Koch, C. (1985). Ill-posed problems in early vision: From computational theory to analog networks. *Proc. of the Royal Society of London B*, 226:303–323.

Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317(6035):314–319.

Pratt, W. (1978). *Digital Image Processing.* John Wiley and Sons.

Rubin, J. and Richards, W. (1982). Color vision and image intensities: When are changes material? *Biological Cybernetics*, 45:215–226.

Rubin, J. and Richards, W. (1984). Color vision: Representing material categories. AI Memo 764, Massachusetts Institute of Technology.

Schein, S. and Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Neuroscience*, 10(10):3369–3389.

Schwarz, M., Cowan, W., and Beatty, J. (1987). Experimental comparison of RGB, YIQ, LAB, and Opponent color models. *ACM Transactions on Graphics*, 6(2):123–158.

Seevink, E. (1988). *Analysis and Synthesis of Translinear Integrated Circuits.* Elsevier.

Seitz, P., Leipold, D., Kramer, J., and Raynor, J. (1993). Smart optical and image sensors fabricated with industrial CMOS/CCD semiconductor processes. *SPIE Proceedings*, 1900:30–39.

Shafer, S. (1985). Using color to separate reflection components. *COLOR research and application*, 10(4):210–218.

Siegel, R. and Howell, J. (1981). *Thermal Radiation Heat Transfer.* McGraw-Hill.

Sivilotti, M., Mahowald, M., and Mead, C. (1987). Real-time visual computation using analog CMOS processing arrays. In *1987 Stanford Conference on VLSI*, pages 295–312.

Slepian, D. (1976). On bandwidth. *Proceedings of IEEE*, 64(3):292–300.

Smith, A. (1985). Color gamut transform pairs. *Computer Graphics*, 10(4):210–218.

Suzuki, H. (1967). *Electronic Absorption Spectra and Geometry of Organic Molecules: An Application of Molecular Orbital Theory.* Academic Press. pgs. 6, 59, 79.

Terzopoulos, D. (1985). Computing visible-surface representations. AI Memo 800, Massachusetts Institute of Technology.

Terzopoulos, D. (1986). Regularization of inverse visual problems involving discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8:413–424.

Tominaga, S. (1987). Expansion of color images using three perceptual attribute. *Pattern Recognition Letters*, 6:77–85.

Tominaga, S. (1990). A color classification method for color images using a uniform color space. In *10th Intl. Conf. on Pattern Recognition*, pages 803–807.

Toumazou, C., Lidgey, F., and Haigh, D. (1990). *Analogue IC Design: The Current Mode Approach.* Short Run Press.

Vrhel, M. and Trussell, H. (1994). Filter considerations in color correction. *IEEE Transactions on Image Processing*, 3(2):147–161.

Watanabe, T., Hashiguchi, K., Yamano, T., Nakai, J., Miyatake, S., Matsui, O., and Awane, K. (1984). A CCD color signal separation IC for single-chip color imagers. *IEEE Journal of Solid-State Circuits*, SC-19(1):49–84.

Watt, A. (1989). *Fundamentals of Three-Dimensional Computer Graphics.* Addison Wesley.

Welch, E. (1991). *Image Processing in the HSI Color Space.* Ph.D. thesis, Mississippi State University.

Williams, D. (1986). Seeing through the photoreceptor mosaic. *Trends is Neuro Sciences*, pages 193–198.

Winfree, A. (1990). *The Geometry of Biological Time.* Springer-Verlag.

Wolfe, W. and Zissis, G. (1985). *The Infrared Handbook.* Environmental Research Institute of Michigan.

Wright, W. (1989). A Markov Random Field approach to data fusion and colour segmentation. *Image and Vision Computing*, 7(2):144–150.

Wyszecki, G. and Stiles, W. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley and Sons.

Yang, W. (1991). Analog CCD processors for image filtering. In *From Neurons to Chips*, pages 114–127. SPIE Proc. 1473.

Yoon, K. (1991). *Aspects of Image Segmentation and Clustering Based on HSI Color Model.* Ph.D. thesis, University of Connecticut.

Zeki, S. (1983). Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, 9(4):741–765.

Zucca, J. (1994). Personal Communication.