

ACTIVE BAYESIAN COLOR CONSTANCY WITH NON-UNIFORM SENSORS

Sandra Skaff

Department of Electrical and Computer Engineering
McGill University, Montréal

September 2002

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfilment of the requirements of the degree of
Master of Engineering

© SANDRA SKAFF, 2002



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-85899-5

Our file Notre référence

ISBN: 0-612-85899-5

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

To My Parents

Abstract

The human impression of the color of an object is the same whether it is viewed foveally or peripherally, despite the non-uniformity of the spatial distribution of the photoreceptors on the retina. It is proposed that it is precisely the changing retinal signal resulting from moving the eye that leads to the perception of *color constancy*. Adhering to a bilinear model, which provides a parametrization of surface patch and illuminant spectra in a scene, the problem addressed in this thesis is defined to be that of estimating these spectral parameters, given only the spectral response of the sensor. The inherent non-uniqueness and thus ill-posedness of the problem lies in that different spectra of lights and surfaces can lead to the same photosensor response.

In this thesis, a multi-sensor regularization technique that solves for the surface reflectance and lighting parameters is developed. This technique sequentially acquires measurements from independent sensors over the scene. Evidence for the parameters is spatially accumulated through a Bayesian probability formulation, which is a statistical approach to designing solutions to perceptual problems. Two cases are presented: (i) one with two sets of sensors, each with different absorption characteristics, and (ii) one in which there is a Gaussian variation in the absorption characteristics across the photoreceptor array. Simulations indicate that there is improvement in the parameter estimation with the additional information provided by the extra sensors.

ABSTRACT

Improvement is also observed in the two-sensors case when the camera is moved to acquire multiple images of the same scene.

Résumé

Malgré la variabilité de la distribution spatiale des capteurs optiques de la rétine, la perception par l'humain de la couleur d'un objet est la même qu'il soit vu au niveau de la fovéa ou en périphérie. Il est proposé que la perception de *constance des couleurs* soit due au signal rétinien variable qui résulte des mouvements de l'oeil. Le problème adressé par cette thèse est celui d'estimer les paramètres spectraux d'un modèle bilinéaire permettant la paramétrisation d'une section de surface et d'un spectre illuminant dans une scène, à partir seulement de la réponse spectrale du senseur. La non-unicité, et donc la nature mal-posée inhérente au problème vient du fait que plusieurs spectres différents de lumière et de surfaces peuvent mener à la même réponse des photocapteurs.

Dans cette thèse, une technique de régularisation à senseurs multiples qui résout l'équation en fonction du facteur de réflexion et des paramètres d'éclairage, est développée. Cette technique acquiert séquentiellement des mesures de senseurs indépendamment distribuées dans la scène. L'information nécessaire au raffinement des paramètres est accumulée spatialement par une formulation probabiliste Bayésienne, une approche statistique pour l'élaboration de solutions aux problèmes perceptuels. Deux cas sont présentés: le premier avec deux ensembles de senseurs ayant chacun des caractéristiques d'absorption différentes, et le second où les caractéristiques d'absorption varient selon une distribution gaussienne dans la matrice de capteurs optiques. Les

RÉSUMÉ

simulations indiquent une amélioration de l'estimation des paramètres grâce à l'information fournie par les senseurs additionnels. Une amélioration est également observée dans le cas à deux senseurs, lorsque la caméra est déplacée pour capter plusieurs images d'une même scène.

Acknowledgements

First, I would like to thank my supervisor, Professor James J. Clark, for his assistance, on an academic and financial level, as well as his support on a moral level. He has been very understanding and supportive throughout my period of research and graduate studies. I would also like to thank my co-supervisor, Professor Tal Arbel, for her moral support and technical guidance as well, especially in developing the probability formulations, which were an essential part of this thesis.

Next, I wish to acknowledge Professor Kathy Mullen, who was kind enough to provide us with the Munsell patches that were used in the experiments.

Furthermore, I would like to thank all the research group members, past and present, for providing such a pleasant atmosphere. Of those group members, special thanks go to my colleague Ziad Hafed for his valuable discussions and his assistance throughout the period of my research. His patience and moral support are highly appreciated. Also, I would like to thank Pierre-Olivier Laprise for his translation of the thesis abstract into French.

I wish to thank all faculty, staff, and students of the McGill Centre for Intelligent Machines (CIM) for providing such a friendly working place. Of those CIM members, I would like to acknowledge Harkirat Sahambi for all his Latex tips, which were very useful for writing this thesis. I would also like to thank Miguel Torres-Torriti for his tips on optimization algorithms, as well as his Latex hints.

ACKNOWLEDGEMENTS

Finally, I would like to thank all of my friends for their moral support. I also wish to express my greatest thanks and appreciation to my family, especially my parents, for everything they have provided me with so far, the moral and financial support, the encouragement...

TABLE OF CONTENTS

Abstract	i
Résumé	iii
Acknowledgements	v
LIST OF FIGURES	ix
LIST OF TABLES	xiii
CHAPTER 1. Introduction	1
1.1. Motivation	2
1.2. Thesis Overview	5
CHAPTER 2. Previous Approaches	6
2.1. The Maloney-Wandell Bilinear Model	10
2.1.1. Principal Components Analysis	12
2.2. The Brainard and Freeman Bayesian Model	16
CHAPTER 3. The Sequential Multiple Sensor Model	22
3.1. The Binary Formulation	23
3.1.1. A Single Patch	23
3.1.2. Multiple Patches	24
3.2. The Non-Uniform Formulation	32

TABLE OF CONTENTS

3.2.1. A Single Patch	32
3.2.2. Multiple Patches	33
CHAPTER 4. Simulation Results	36
4.1. The Algorithm	37
4.2. Simulation Results	40
4.2.1. Single versus Multiple Patches	40
4.2.2. The Different Sensors: A Comparison	44
4.2.3. Moving Camera	46
4.3. Summary	49
CHAPTER 5. Experiments and Analysis	50
5.1. The Experimental Setup	50
5.2. The Algorithm	51
5.3. Experimental Results	55
5.3.1. A Single Patch	55
5.3.2. Multiple Patches	57
CHAPTER 6. Conclusions and Future Work	66
REFERENCES	69
APPENDIX A. Munsell Patches	74
A.1. The Notation System	74
A.2. Correspondence: Munsell Label to Orava's Label	76

LIST OF FIGURES

2.1	The estimated and actual spectra of the (a) illuminant, (b) surface patch (Munsell patch 300) for the Brainard-Freeman Bayesian model using the MAP estimate when there are multiple patches in the scene.	20
3.1	A single patch in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part).	24
3.2	Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part).	25
3.3	Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part): The camera is moved once.	28
3.4	Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part	

	and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part): The camera is moved twice.	30
4.1	(a) Surface basis functions and (b) Daylight and Skylight basis functions. "bf" stands for "basis function".	39
4.2	The estimated and actual spectra of the illuminant (number 9 from Parkkinen's set of spectra) for the 'no filter' case: (a) one patch in the scene, (b) multiple patches in the scene.	41
4.3	The estimated and actual spectra of a surface patch (Munsell patch 875) for the 'no filter' case: (a) one patch in the scene, (b) multiple patches in the scene.	42
4.4	The estimated and actual spectra of the illuminant (number 9 from Parkkinen's set of spectra) for the binary filter case: (a) one patch in the scene, (b) multiple patches in the scene.	43
4.5	The estimated and actual spectra of a surface patch (Munsell patch 875) for the binary filter case: (a) one patch in the scene, (b) multiple patches in the scene.	44
4.6	The estimated and actual spectra of the illuminant (number 2 from Parkkinen's set of spectra) when there are multiple surface patches in the scene in the cases of (a) 'no filter', (b) binary filter, (c) Gaussian filter.	45
4.7	The estimated and actual spectra of a surface patch (Munsell patch 300) when there are multiple surface patches in the scene in the cases of (a) 'no filter', (b) binary filter, (c) Gaussian filter.	45
4.8	The estimated and actual spectra of the illuminant (number 30 from Parkkinen's set of spectra) when there are multiple patches	

LIST OF FIGURES

	in the scene in the (a) 'no filter' case and in the binary filter case when the camera is: (b) not moved, (c) moved once, (d) moved twice.	47
4.9	The estimated and actual spectra of a surface patch (Munsell patch 400) when there are multiple patches in the scene in the (a) 'no filter' case and in the binary filter case when the camera is: (b) not moved, (c) moved once, (d) moved twice.	48
5.1	CMY (Cyan, Magenta, Yellow) Camera Color Filter Spectral Response	52
5.2	The IR cutoff filter	53
5.3	The incandescent light source spectral power distribution. . . .	54
5.4	The estimated and actual spectra of a green Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.	56
5.5	The estimated and actual spectra of a red Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.	57
5.6	The estimated and actual spectra of a blue Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.	58
5.7	The estimated and actual spectra of a purple Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.	59

LIST OF FIGURES

5.8	The estimated and actual spectra of the illuminant when there is a single surface patch in the scene in the cases of (a) 'no filter', (b) binary filter.	60
5.9	The estimated and actual spectra of the Munsell patches (green, blue, red) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.	61
5.10	The estimated and actual spectra of the Munsell patches (purple, green, yellow) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.	62
5.11	The estimated and actual spectra of the Munsell patches (purple, blue, red) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.	63
5.12	The estimated and actual spectra of the illuminant when there are multiple patches in the scene in the 'no filter' case.	64
A.1	The value scale.	75

LIST OF TABLES

4.1	RMS error for the illuminant (number 9 from Parkkinen's set of spectra) and surface patch (Munsell patch 875) spectra for the 'no filter' case when there is either a single patch or multiple patches in the scene.	41
4.2	RMS error for the illuminant (number 9 from Parkkinen's set of spectra) and surface patch (Munsell patch 875) spectra for the binary case when there is either a single patch or multiple patches in the scene.	43
4.3	RMS error for the illuminant and surface patch spectra for the 'no filter', binary filter and Gaussian filter cases when there are multiple surface patches in the scene.	46
4.4	RMS error for the illuminant (number 30 from Parkkinen's set of spectra) and surface patch spectra (Munsell patch 400) when there are multiple surface patches in the scene for the 'no filter', and binary filter case when the camera is: not moved, moved once, and then moved twice.	46
4.5	Average RMS error for Munsell Patch 400 and the illuminant spectra when there are multiple patches in the scene for the 'no	

LIST OF TABLES

	filter' case and the binary filter case when the camera is: not moved, moved once, and moved twice.	49
5.1	RMS error for each of the Munsell surface patch spectra when there is a single patch in the scene for the 'no filter' and the binary filter cases.	56
5.2	RMS error for each of the Munsell surface patch spectra (green, blue, red) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.	58
5.3	RMS error for each of the Munsell surface patch spectra (purple, green, yellow) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.	59
5.4	RMS error for each of the Munsell surface patch spectra (purple, blue, red) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.	60
A.1	Table of correspondence between the Munsell patch labels and Orava's labels for the glossy Munsell patches used in our experiments.	80

CHAPTER 1

Introduction

Color perception has been a topic of great interest for researchers for quite a significant amount of time. As a start, we shall define some recurring terms, which we believe will be keys in understanding the context of this thesis. *Color* is the quality of any visual experience except its spatial attribute of extent or bulk, as defined by Helson [26]. Judd [28] defines *surface color* as the color experienced as a property of a surface. He also defines *lightness* as the attribute of any surface color that permits it to be classified as equivalent to some member of the series of gray level values ranging between black and white. *Reflectance*, according to Helson [26], is defined as the fraction of the incident light which a hypothetical perfectly diffusing sample would need to reflect in order to yield the same brightness as the actual sample under the same conditions of illuminating and viewing.

Color vision and color perception deal with the eye, the incentive for computer vision research, and its primary stimulus, light. The human retina is clearly non-homogeneous, but the human perception of color is remarkably invariant to eye position [38] and also invariant to illuminant changes. In other words, the human visual system can differentiate between illuminant changes and changes in surface colors. If we, humans, look at a green rectangle under daylight and then under incandescent

light, we would see it as having two different colors, but we would still *perceive* it as having the same color. However, can a robot perceive that rectangle as having the same green color under those two different lighting conditions?

Clark and O'Regan [8] proposed that the human visual system uses a stable-world assumption in developing position invariant color perception. In their approach, it is precisely the changing retinal signal resulting from moving the eye when viewing a stable colored surface patch that determines the perceived color. We propose that *color constancy*, the perception of a constant color across variations in illumination, can be explained by a process similar to Clark and O'Regan's explanation of color stability across eye movements.

Brainard and Freeman [7] used a Bayesian technique to regularize the problem of estimating surface and light source colors. The main focus of this thesis is to estimate surface and light source colors using Brainard and Freeman's Bayesian approach as it makes clearer the connection from earlier computational approaches to color constancy (to be discussed in Chapter 2), on one hand; yet it still retains the sensorimotor nature of the Clark-O'Regan learning approach, on the other hand. The main contributions of our work are in extending the Brainard-Freeman technique by using a large set of independent sensors, which will be described in a later section. Following the next section, which describes the motivation for this research, Section 1.2 concludes this chapter with a brief overview of this thesis and its organization.

1.1. Motivation

Why Color and why color constancy, in particular? Color vision is a very interesting aspect of computer vision, as it helps in the process of perception or understanding of the visual scene. Clearly, the colors recorded by a device should be constant across

a change in the scene illumination in order to improve object recognition, in particular, and the perception process, in general. In other words, if the color of an object were to be a cue in the recognition task, then this color should be constant despite variations in illumination. In addition to perception, other applications, such as digital photography, require color constancy across a scene in order to capture consistent images. Furthermore, if a photograph is meant to be an accurate representation of what the photographer saw, then the scene illumination in the photograph must be corrected in the same way as the human visual system corrects partially for the prevailing scene illumination [21]. Other industrial applications require stable color constancy algorithms in order to be carried out successfully. In health inspection applications, variation in illumination should not affect the perception of the color of products in classifying them as good or bad. Also, in dentistry applications, matching false teeth to the original ones requires the perception of constant color despite variations in illumination. All of these problems are very likely to occur, as lighting conditions tend to vary easily from one location to another.

Despite the importance of color as described above, there is no reliable algorithm for recognizing the color of an object under different lighting conditions. This problem is ill-posed like many problems in computer vision, which have to recover 3D properties of surfaces from 2D images [5]. A problem is *well-posed* in the sense of Hadamard [24] if it has a solution, this solution is unique, and the solution varies continuously with a set of data or the solution is stable. If a problem does not satisfy all three of these conditions, then it is said to be *ill-posed*. Many inverse problems are ill-posed. Vision problems are often inverse problems, and they also tend to be ill-posed. A regularization technique [40] can be used to make the problem well posed and thus solvable. The regularization technique used in this thesis is the Bayesian

approach which is a statistical approach [4, 11]. This method is utilized in this work to design solutions to our perceptual problem. It is described in detail in Section 2.2.

The human retina is non-uniform and non-homogeneous, as the spatial distribution of its photoreceptors, the rods and cones, is irregular and the density of the receptors decreases steadily away from the central region, or fovea. Less well-known is the fact that the retina has a marked non-homogeneity in the spectral sensitivity of its photoreceptors. The retina has a small, yellowish central region, which is called the macula, and it is the area providing the clearest, most distinct vision. In the central, or macular, region there exists three wavelength selective photoreceptor cone classes, of which the long and medium wavelength sensitive are most numerous. In the periphery, short wavelength cones are more prevalent than the other two cone classes. An additional non-homogeneity arises from the fact that the macular pigment, which is a yellowish jelly that covers the macula, absorbs up to 50% of the light in the short wavelength range [6], causing a significant shift in the color sensitivity of foveal receptors. For small pupil sizes the absorption of the lens material, which is supposed to focus light onto the retina, also provides a wavelength dependent attenuation, similar in effect to the macular pigment absorption. In spite of these non-homogeneities, the human perception of color is remarkably invariant to eye position. We have the subjective impression that the color of an object is the same when viewed peripherally as when viewed foveally.

In this thesis we propose that *color constancy*, the perception of a constant color across variations in illumination, can be explained by a process similar to the Clark-O'Regan explanation of color stability referred to previously. Under this new view, color constancy arises from the learning of the pattern of sensory variation corresponding to a stable physical aspect (the color of a surface patch) as the eye moves, taken across a variety of different illumination conditions. In other words, perception

of a canonical color, e.g. 'red', will occur whenever the appropriate pattern of sensory activity with respect to eye movements is obtained.

The approach that we will take in this thesis is somewhat different from that taken by Clark and O'Regan [8], and will follow the Bayesian approach of Brainard and Freeman [7]. Our method, however, involves an extension of the Brainard-Freeman technique to use a large set of independent sensors (corresponding to the photoreceptors at different retinal locations) from which measurements are obtained sequentially (corresponding to moving the gaze across a surface patch). Two cases are studied, one in which there are only two sets of sensors, each having a different spectral sensitivity characteristic, the binary filter Bayesian formulation, and one in which there is a smooth variation in the spectral sensitivity across the sensor array, the Gaussian filter Bayesian formulation.

1.2. Thesis Overview

This thesis proposes a method for achieving color constancy by extending the Brainard-Freeman approach, in which only one sensor was modeled. Chapter 2 describes the latter in addition to other approaches necessary to clarify it. In our approach, measurements are acquired from a large set of independent sensors, which are described in Chapter 3, which also derives the posterior of the Bayesian probability formulation in both cases. Chapter 4 details the simulations; in particular it explains how the scenes were constructed and how the individual terms of the Bayesian probability formulation were arrived at. Chapter 5 shows the experimental results, which were carried out in our laboratory and compared to the simulations done in Chapter 4. Last but not least, Chapter 6 concludes this thesis and provides some insight into future work, which may be done to improve on the results.

CHAPTER 2

Previous Approaches

Color constancy, which is the perception of a constant color across variations in illumination, has been a topic of chief interest in the psychology of color since the beginning of the twentieth century. This problem has been addressed by Ives in 1912 [27], for example, who stated, “The ... problem ... of constant color appearance is met by just one condition, namely restriction to one light source of constant spectral character.” This is obvious, as we humans see an object under lights of different spectra as having different colors, but we perceive them as having the same color. Later on, Finlayson *et al.* [19] define color constancy as the ability to discount the effect of the illuminant. A color camera, like the human eye, has three color sensors; therefore, in a color image each pixel is a three dimensional vector, RGB , having one component per sensor channel (Red, Green, Blue). Reviewing the research that has been done in the field of color constancy brings us upon several approaches devised to explain human color constancy, which served as a key later on to formulate models for machine color constancy. In fact, many authors have tried to solve the under-constrained problem by making additional assumptions about the world.

The Retinex theory of Land [29] was one of the first computational theories which attempted to explain human color constancy. This theory is based on the observation

that the light incident on a white patch is, after reflection, spectrally unchanged, which means that it has the same spectrum. Also, a white reflectance must induce maximal *RGB* camera responses; therefore, the *R*, *G*, and *B* responses can be used as an estimate of the color of the scene illuminant.

The chromaticity constancy problem has proven to be much easier to solve than the color constancy problem. A chromaticity vector can be calculated from a color vector as follows:

$$R' = \frac{R}{B}, \quad G' = \frac{G}{B}, \quad 1 = \frac{B}{B}. \quad (2.1)$$

Since the third component of the chromaticity vector is always 1, it can be represented as a 2D vector (R', G') . This implies that the 3D space in color coordinates is narrowed down to a 2D space in chromaticity coordinates. Finlayson [18] made two important observations. First, the gamut of all possible image chromaticities depends on the illuminant color. Furthermore, the illuminant color is limited. According to Finlayson, an image chromaticity is said to be consistent with a particular light if it is within the gamut of all possible chromaticities under that light. A single chromaticity will be consistent with many illuminants or an illuminant set, and the set of feasible illuminants is the set resulting from the intersection of all the illuminant sets. Then selecting the median or mean illuminant from the feasible set would lead to good color constancy, with a few exceptions.

D’Zmura and Lennie [16] propose a model that finds the colors of objects, and this model depends on light adaptation mechanisms as well as eye movements. They suggest that after performing principal components analyses [11, 32] on daylight and surface reflectances much of the variance of these domains may be accounted for by only a small number of components. They show in [16] how mechanisms of light

adaptation can find and discount the illuminant and then proceed to show how a stable hue descriptor can be obtained.

Finlayson, Funt, and Barnard describe the varying illumination algorithm in [20]. Their approach differs from other approaches in that they vary the lighting conditions in order to obtain three *RGB* measurements for each surface, two under unknown illumination conditions and one under a canonical illumination condition. According to Finlayson *et al.* [20], the function that maps the *RGB* measurements obtained under the unknown illumination to those obtained under the canonical illumination solves the color constancy problem. Finlayson, Hordley, and Hubel [21] describe a technique to recover the ambient light in the scene hoping that after the light is known, the color of the surface can be easily recovered. They build a correlation matrix to correlate possible image colors with each of the set of possible illuminants. For each illuminant, they characterize the range of possible colors or chromaticities that can be observed under that light, and they build a probability distribution that gives the likelihood of observing an image color under a given light. Therefore, their algorithm requires some knowledge about the range and distribution of image colors that can be recorded by a camera under a set of possible lights. An interesting and unique approach to the problem of color constancy in the literature is the learning algorithm developed by Funt *et al.* [23]. They hypothesize that a back-propagation neural network can learn the relationship between the image of a scene and the chromaticity of scene illumination. Then the trained network would be able to determine the scene illumination from its image; and therefore, the image colors could be corrected relative to the illuminant. In fact, the neural network approach can be defined as a method of dealing with the inherent uncertainty of the problem.

The varying illumination, correlation, and learning approaches show how to compute the illuminant only. The authors in [20, 21, 23] do not show how to compute the

surface colors given the illuminants. This might be a trivial computation; however, there may be some difficulties involved due to a particular computation.

Maloney and Wandell [30] described a bilinear model to attain color constancy, which is detailed in Section 2.1. In their approach, the authors use the quantum catch data first to obtain a description of an illuminant spectral power distribution and then, in the second stage, recover descriptions of surface reflectance functions. On the other hand, D’Zmura in [12] first recovers the reflectances and then the illuminants. For two-stage linear recovery to be feasible, it is necessary that the number of photoreceptor types equals or exceeds either the dimension of the illumination model or the number of the reflectance model. D’Zmura and Iverson approach the problem of color constancy in [13] by using a bilinear model similar to that developed by Maloney and Wandell [30]. They analyze a two-stage recovery algorithm for recovering reflectance and illuminant descriptors from quantum catch data and determine necessary and sufficient conditions for the recovery to work. Then they proceed to present the model check algorithm, which is a way to test whether a given bilinear model provides unique recovery. Further research done by D’Zmura and Iverson on this work is described in [14, 15].

In recent work, the ill-posed nature of the color constancy problem has been tackled using the Bayesian probability theory [37, 7, 21]. Given some knowledge of typical scenes, it is possible to calculate the probability of observing the chromaticity under a certain illuminant. In this chapter, some work found in the literature was briefly summarized and then the two fundamental algorithms to our model are explained in detail in Sections 2.1 and 2.2, which includes a probabilistic model.

Our color constancy algorithm builds on the Bayesian technique of Brainard and Freeman [7], which in turn uses the bilinear model of Maloney and Wandell [30] to

provide a parametrization of the surface and illuminant spectra. We shall describe these two models in detail in the coming sections.

2.1. The Maloney-Wandell Bilinear Model

Maloney and Wandell [30] developed one of the most influential approaches to color constancy, which has served as the inspiration for many subsequent techniques. Their approach is based on providing parametric models for the spectra of the illuminant and light sources, and then they proposed an algorithm for computing the model parameters from measurements of the light reflected from a set of surface patches.

They start by describing the light arriving at location x on an array of sensors by the function

$$E(\lambda)S^x(\lambda). \quad (2.2)$$

$E(\lambda)$ is the spectral power distribution of the ambient light in the scene, and $S^x(\lambda)$ is the surface spectral reflectance, while λ denotes the wavelength. Since there are p classes of sensors at each location x , and the relative wavelength sensitivity of the k^{th} is $R_k(\lambda)$, the response recorded at each location is given by:

$$\rho_k^x = \int E(\lambda)S^x(\lambda)R_k(\lambda)d\lambda, \quad k = 1, 2, \dots, p. \quad (2.3)$$

S and E can be represented as linear models:

$$S^x = \sum_{j=1}^n \sigma_j^x S_j(\lambda), \quad (2.4)$$

$$E(\lambda) = \sum_{i=1}^m \epsilon_i E_i(\lambda). \quad (2.5)$$

2.1 THE MALONEY-WANDELL BILINEAR MODEL

In both cases, the basis functions are fixed and they are assumed known. The basis functions, $S_j(\lambda)$, which correspond to the surface reflectances, are computed using principal components analysis (see section 2.1.1) on a set of 150 Munsell color chips (see Appendix A for details), which includes a wide range of colors. The basis lights $E_i(\lambda)$ are computed using principal components analysis on a set of 622 functions describing the spectral distribution of natural daylight measured over a range of weather conditions and times of day. Therefore, finding the surface reflectance and ambient light involves recovering the basis function weights σ_j^x and ϵ_i . The sensor responses are seen to be bilinear functions of the unknown basis function weights. This bilinearity implies that the problem of finding the weights is ill-posed, as different choices of ϵ and σ_x can produce the same sensor measurements.

Substituting Equations 2.4 and 2.5 in Equation 2.3 yields the following equation:

$$\rho_k^x = \sum_{j=1}^n \sigma_j^x S_j(\lambda) \sum_{i=1}^m \epsilon_i E_i(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, \dots, p. \quad (2.6)$$

They rewrite Equation 2.6 as:

$$\rho^x = \Delta_\epsilon \sigma^x, \quad (2.7)$$

where Δ_ϵ contains information about the light source in the scene. If the ambient light is unknown, the equation above is unsolvable. Therefore, they concluded that additional information is needed to solve the equation. This led them to suppose that there are $p = n + 1$ linearly independent sensors. Then they determine the plane spanning the sensor quantum catches, which would help them recover the ambient light vector ϵ . After recovering ϵ , Δ_ϵ is recovered and then the weights of the surface reflectance function by inverting the transformation. Therefore, the sensors in this case contain enough information to find the surface reflectance function.

Assuming that a surface can be modeled by only two basis functions is a very harsh assumption made by Maloney and Wandell. They need such an assumption because their algorithm requires more sensors than surface basis functions, and there are only three types of sensors, the red, green, and blue. This is a serious limitation of the algorithm in the sense that it will not produce accurate results when presented with surfaces that cannot be represented accurately with only two basis functions. This observation was made by Parkkinen *et al.* in [36]. However, since the first few principal components contain the most information, this assumption might work out well. On the other hand, the number of sensors could be increased with the number of surfaces, such that $p = n + 1$, to get better results. However, having more sensors means decreasing the spatial sampling density within individual sensor classes and thus results in less resolution. Therefore, there is a trade-off between the two approaches to solving the problem. Moreover, if the number of sensors were to be increased, how would the correct number of sensors and the corresponding spatial resolution of each be determined? This problem which results from Maloney and Wandell's attempt to solve the ill-posedness problem resulting from the bilinearity of their formulation led us to consider the Bayesian approach, which builds upon their model.

2.1.1. Principal Components Analysis. The approach of Maloney and Wandell [30] described in Section 2.1 uses basis function expansions of the illuminant and surface spectra as the parametric models. They determine the basis functions through a principal components analysis, which is explained in this section.

Principal components analysis (PCA), which is also known as the *Karhunen-Loève transformation*, is a standard technique commonly used for data reduction in statistical pattern recognition and signal processing [11, 32]. Feature selection refers to a process whereby a data space is transformed into a feature space, which has the

2.1 THE MALONEY-WANDELL BILINEAR MODEL

same dimension as the original data space. This transformation is designed in such a way that the number of dimensions can be reduced, which means that the data set may be represented by a reduced number of “effective” features and still retain most of the intrinsic information content of the data.

Let \mathbf{X} denote an m -dimensional vector representing a certain environment. Assume that \mathbf{X} has zero mean:

$$E[\mathbf{X}] = 0.$$

where E is the statistical expectation operator. Let q be a unit vector, also of dimension m , onto which the vector \mathbf{X} is to be projected:

$$A = \mathbf{X}^T q = q^T \mathbf{X}. \quad (2.8)$$

The projection A is a random variable and since the random vector X has zero mean, the mean value of the projection A is zero too:

$$E[A] = q^T E[\mathbf{X}] = 0. \quad (2.9)$$

The variance of A is:

$$\begin{aligned} \sigma^2 &= E[A^2] \\ &= E[(q^T \mathbf{X})(\mathbf{X}^T q)] \\ &= q^T E[\mathbf{X}\mathbf{X}^T] q \\ &= q^T \mathbf{R} q. \end{aligned} \quad (2.10)$$

The m -by- m matrix R is the correlation matrix of the random vector \mathbf{X} :

$$\mathbf{R} = E[\mathbf{X}\mathbf{X}^T]. \quad (2.11)$$

2.1 THE MALONEY-WANDELL BILINEAR MODEL

The principal components can be found by solving the following equation:

$$\mathbf{R} \mathbf{q} = \lambda \mathbf{q}. \quad (2.12)$$

Equation 2.12 is recognized as the *eigenvalue problem*. The problem has nontrivial solutions (i.e. $q \neq 0$) only for special values of λ that are called eigenvalues of the correlation matrix \mathbf{R} . These are denoted by $\lambda_1, \lambda_2, \dots, \lambda_m$. The associated values of q (q_1, q_2, \dots, q_m) are called eigenvectors, and they are unique assuming that the eigenvalues are distinct. Then:

$$\mathbf{R} \mathbf{q}_j = \lambda_j \mathbf{q}_j, \quad j = 1, 2, \dots, m. \quad (2.13)$$

Let the eigenvalues be arranged in decreasing order:

$$\lambda_1 > \lambda_2 > \dots > \lambda_j > \dots > \lambda_m. \quad (2.14)$$

so that $\lambda_1 = \lambda_{max}$. Let the associated eigenvectors be used to construct an m -by- m matrix:

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j, \dots, \mathbf{q}_m]. \quad (2.15)$$

Therefore, combining the set of m equations represented in Equation 2.13, the following equation is obtained:

$$\mathbf{R} \mathbf{Q} = \mathbf{Q} \Lambda. \quad (2.16)$$

where Λ is a diagonal matrix containing the eigenvalues of \mathbf{R} :

$$\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_m]. \quad (2.17)$$

2.1 THE MALONEY-WANDELL BILINEAR MODEL

The matrix \mathbf{Q} is an orthogonal, unitary matrix, which means that its column vectors (i.e. the eigenvectors of \mathbf{R}) satisfy the conditions of orthonormality:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (2.18)$$

Equation 2.18 can also be written as:

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}. \quad (2.19)$$

From equations 2.16 and 2.18, the eigendecomposition problem of \mathbf{R} can be formulated by the following equation:

$$\mathbf{Q}^T \mathbf{R} \mathbf{Q} = \Lambda. \quad (2.20)$$

After computing the eigenvectors of the correlation matrix, which are the same as the principal components of the data, any m -dimensional vector x in the space can be represented with l dimensions such that $l \leq m$. This is attained by discarding those linear combinations of eigenvalues and eigenvectors that have small variance (corresponding to small values of \mathbf{R}) and retaining only those that have large variance (corresponding to the largest eigenvalues of \mathbf{R}). Therefore, the approximate vector $\hat{\mathbf{x}}$ can be represented as follows:

$$\begin{aligned} \hat{\mathbf{x}} &= \sum_{i=1}^l a_i \mathbf{q}_i \\ &= [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_l] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix}, \quad l \leq m. \end{aligned} \quad (2.21)$$

In conclusion, if l is much smaller than m , and the corresponding eigenvalues $\lambda_l, \lambda_{l+1}, \dots, \lambda_m$ are negligible, then significant dimensionality reduction can be achieved. In this thesis, for example, the 1250 Munsell patches [2] are represented with only 8 basis functions, which are obtained by principal components analysis, as we shall see in Chapter 4.

2.2. The Brainard and Freeman Bayesian Model

A “higher level” approach to the solution of color constancy is Brainard and Freeman’s algorithm [7], which builds on Maloney and Wandell’s bilinear model [30] described in Section 2.1. The bilinearity of the latter model implies that the problem of recovering the basis function weights of the ambient light ϵ_i and the surface reflectances σ_j^x is ill-posed, as different choices of σ_x and ϵ can produce the same sensor measurements. Brainard and Freeman used a regularization technique to solve the problem of computing the values of the illuminant and surface parameters. The regularization technique used is the Bayesian decision theory [4, 11], which is a fundamental statistical approach. This method is utilized in this work to design a solution to our perceptual problem. This approach is based on the assumption that the theory is posed in probabilistic terms, and that all of the relevant probability values are known. The Bayesian approach combines information contained in an image with information given a priori about likely physical configurations of the scene. This prior information can resolve ambiguities in the image data.

Three probabilities are the basic terms of the Bayesian probability theory: the prior, posterior, and likelihood. The prior probability describes what is known before observing the data, while the posterior probability describes what is known after observing the data. The likelihood can be thought of as a rendering equation and it expresses the relation between the data and the parameters. Let the vector of the

2.2 THE BRAINARD AND FREEMAN BAYESIAN MODEL

surface and illumination model weights be denoted as x , and the sensor responses by y . We can obtain a statistical model for x by the conditional *posterior* density function, $p(x|y)$, of x given the measurement:

$$p(x|y) = \frac{p(y|x) p(x)}{p(y)} \approx p(y|x) p(x). \quad (2.22)$$

$p(y|x)$ is the *likelihood* which models the relation between the illuminant spectrum model, the surface spectra models and the sensor responses. $p(x)$ represents the *prior* information on the model parameters. $p(y)$ represents the probability of the sensor responses or the measurements. We will drop it in the equations yet to follow because of the fact that we are searching for minima, and having a normalization term will not affect the locations of these minima. In the Brainard-Freeman formulation [7], the prior is estimated through principal component analysis (PCA) techniques where the principal components are built from the surface reflectances of a fixed set of Munsell color patches (see Appendix A) [2, 36]. The likelihood $p(y|x)$ is also represented by a normal distribution. Given the posterior $p(x|y)$ they compute a loss function, which they call the Bayesian expected loss:

$$\bar{L}(\tilde{x}|y) = \int_x \bar{L}(\tilde{x}|x) p(x|y) dx. \quad (2.23)$$

This function computes the penalty for choosing a single estimate \tilde{x} when the actual parameters are x [4]. Brainard and Freeman choose an estimate for \tilde{x} such that the loss is minimum. The loss function is shift invariant, such that it depends on the difference $\tilde{x} - x$. In this case, the expected loss is simply the posterior convolved by the loss function with its argument negated.

Why Bayesian estimation? Bayesian estimation provides a way to choose an optimal estimate that uses all of the information contained in the data. What are the drawbacks? It may be difficult to specify a prior distribution that adequately captures

2.2 THE BRAINARD AND FREEMAN BAYESIAN MODEL

what is known about the structure of the parameters. In the case of color constancy, the prior must specify how likely it is that a particular illuminant spectral power distribution and surface reflectances will occur. Also, it may be difficult to determine a loss function that captures how costly errors of different types are. Finally, it may be computationally expensive to minimize the expected loss. Three types of rules, tackled by Brainard and Freeman [7], corresponding to minimizing the expected loss are discussed below.

The maximum a posteriori (MAP) rule corresponds to minimizing the expected loss with respect to the minus delta loss function:

$$L(\tilde{x}, x) = -\delta(\tilde{x} - x). \quad (2.24)$$

Convolving the posterior with the minus delta loss function does not alter the relative shape of the posterior; therefore the estimate that maximizes the posterior also minimizes the corresponding expected loss.

The minimum mean squared error (MMSE) chooses the posterior mean as the best estimate for the scene parameters. This estimator is widely used; for example, it is used in Kalman filtering. The MMSE rule corresponds to minimizing the expected loss with respect to the squared-error loss function [34]:

$$L(\tilde{x}, x) = |\tilde{x} - x|^2. \quad (2.25)$$

The maximum local mass (MLM) estimator corresponds to minimizing the expected loss for the local mass loss function:

$$L(\tilde{x}, x) = -\exp[-|K_L^{-1/2}(\tilde{x} - x)|^2], \quad (2.26)$$

such that

$$|K^{-1/2}|^2 = x^T K^{-1} x. \quad (2.27)$$

For matrices K_L of sufficiently small eigenvalues, this loss function rewards approximately correct estimates and penalizes all grossly incorrect estimates equally.

In real computational vision problems, non-Gaussian or multimodal posterior probability distributions occur. In these cases, the MAP and the MMSE can disagree and both can be unsatisfactory unless the actual loss function for the estimation problem matches that implied by the estimator. However, in more simple cases, both the MAP and the MMSE rules provide intuitively plausible estimates that tend to agree with each other, and the posterior probability mass would be well localized in the scene parameter space.

According to Brainard and Freeman [7], the MAP estimator ignores relevant information in the posterior distribution, since it takes into account the point of maximum probability mass only; therefore it results in a systematic bias of its estimates. In other words, the MAP estimator is sensitive to the height of the posterior distribution only and ignores information in its structure, which might be important. The MMSE chooses the scene parameter values that minimize the average squared distance from the true scene parameter values. However, the MMSE estimate may require a computationally intensive integration over the entire scene parameter space, on one hand; on the other hand, it may select scene parameters which do not explain the visual data at all or which are nonsensical from a perceptual standpoint, and it may penalize large errors too much compared to the MLM. In short, the MLM function takes into account both the shape and the structure of the posterior and is not biased by the prior, so it produces the best results. Its results are also consistent with

2.2 THE BRAINARD AND FREEMAN BAYESIAN MODEL

the observed data because the local mass function is set to be local in the parameter space.

Although the MLM would give the best results compared to the other two estimators as discussed above, the MAP estimator was the one implemented in this thesis because the MLM is computationally expensive. Brainard and Freeman in [7] show the results of the MAP, MMSE, and MLM estimators in estimating the illuminant power. They also show results for other algorithms: Gray World, Realizability 1, Realizability 2, and Subspace. Brainard and Freeman's model was first implemented

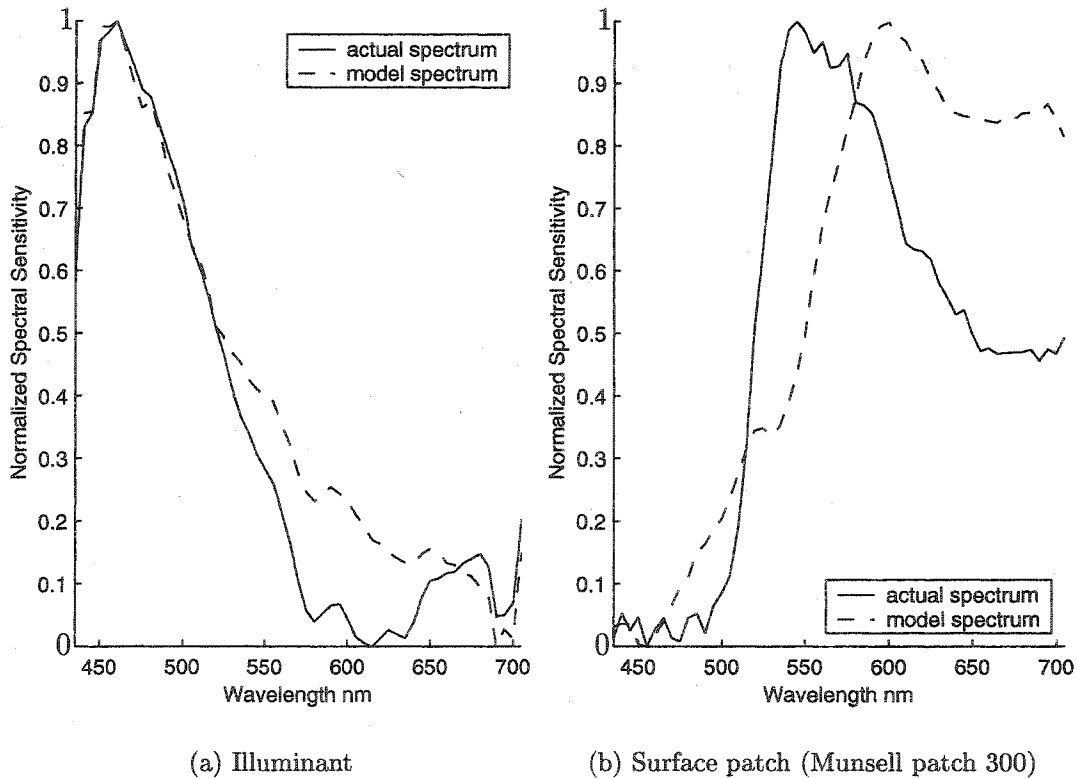


FIGURE 2.1. The estimated and actual spectra of the (a) illuminant, (b) surface patch (Munsell patch 300) for the Brainard-Freeman Bayesian model using the MAP estimate when there are multiple patches in the scene.

as mentioned in Chapter 1. The surface patch in this case is viewed with two other patches in the scene under a certain light source. The estimated and actual spectra

for both the illuminant and one of the surface patches in the scene are plotted in Figure 2.1. However, there are some cases for which their model does not work. This is expected for all models, as they tend to pose assumptions, which in turn pose limitations on the algorithm. The details of the simulations and an extensive discussion of the results will be discussed in Chapter 4 in relation to the contribution made in this thesis, which is an extension to the Brainard-Freeman Bayesian model as mentioned in Chapter 1.

CHAPTER 3

The Sequential Multiple Sensor Model

In this thesis, we introduce a strategy that is an extension to the Brainard-Freeman Bayesian approach described in Section 2.2, in which only one sensor was modeled. In our approach, measurements are acquired from a large set of independent sensors, each with its own spectral sensitivities. The inherent ill-posedness of the problem is therefore addressed through the introduction of more sources of information. Measurements are acquired sequentially from each sensor, much as when a person's gaze moves across a surface in the scene. *Color stability*, as described by Clark and O'Regan [8] is achieved through the accumulation of evidence from the various sensors through a sequential Bayesian estimation process.

As in the Brainard and Freeman approach, we represent evidence for the lighting and surface color parameters of either one or many surface patches in a scene by a conditional probability density function, given the sensor measurements. This probabilistic evidence is then accumulated sequentially over sensors with different spectral sensitivities through a Bayesian chaining approach. We model two cases: (i) one in which there are only two different types of sensors, and (ii) one in which there is a smooth variation in spectral sensitivity curves across the sensor array. In practice, this multi-sensor formulation can be modeled through the placement of a filter with

the appropriate absorption characteristics onto a single camera lens. In this case, evidence can be accumulated spatially over the pixels of the image of the scene (see Chapter 4).

3.1. The Binary Formulation

In this type of formulation, the measurements of a scene are acquired by two sensor types, each with a different spectral sensitivity. The measurements are assumed to be conditionally statistically independent. It is also assumed that each individual sensor actually produces three independent measurements, which are referred to as R , G , and B .

3.1.1. A Single Patch. In this case there is only one surface patch in the scene, illuminated with a single light source. Its RGB measurements from sensors of type X will be denoted by RGB_X , and those from sensors of type Y will be denoted as RGB_Y . The surface spectral model weights vector is denoted by a and the illuminant spectral model weights vector by b . Suppose that the patch is visible to both sensors as shown in Figure 3.1. Let $\{RGB\}$ denote the total set of measurements of the scene:

$$\{RGB\} \triangleq RGB_X, RGB_Y.$$

To accumulate probabilistic evidence of the scene over different sensors, we derive the conditional posterior density function for the parameters, a and b , given the set of measurements of the entire scene, $\{RGB\}$:

$$\begin{aligned} p(a, b | \{RGB\}) \\ \approx p(\{RGB\} | a, b) p(a, b) \end{aligned}$$

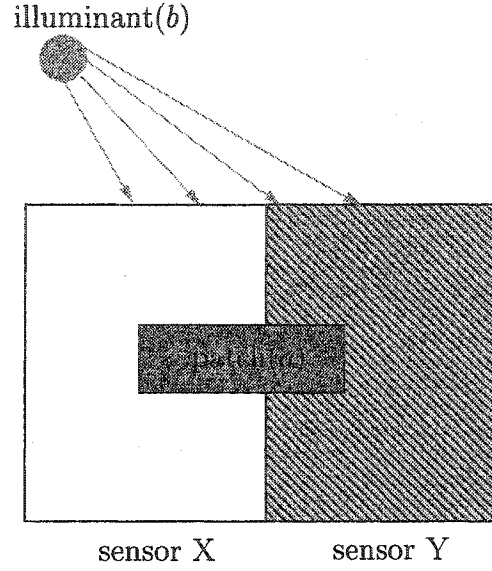


FIGURE 3.1. A single patch in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part).

$$\begin{aligned}
 &\approx p(RGB_X, RGB_Y | a, b) p(a, b) \\
 &\approx p(RGB_X | a, b) p(RGB_Y | a, b) p(a, b) \\
 &\approx p(a, b | RGB_X) p(RGB_Y | a, b).
 \end{aligned} \tag{3.1}$$

In obtaining this formulation, it is assumed that the surface reflectance weight a is statistically independent of the spectral function weight of the illuminant, b . From Equation 3.1, we can conclude that the posterior for the entire scene is a function of the posterior for one sensor (sensor X), which is $p(a, b | RGB_X)$ and the likelihood of the RGB measurement from the other sensor (sensor Y), which is $p(RGB_Y | a, b)$.

3.1.2. Multiple Patches. In this case it is assumed that there are three surface patches in the scene, illuminated with a single light source. RGB measurements of a particular surface obtained from sensors of type X will be denoted RGB_{Xn} , and

those from sensors of type Y as RGB_{Yn} , where n refers to the surface patch being viewed. The surface spectral model weight vector is denoted by a_n and the illuminant spectral model weights vector by b . Two cases are considered in this case. The first one is when the image is acquired once, that is with a static camera, while the second one is when the image of the scene is acquired several times, from different positions; that is the camera is moved.

3.1.2.1. Static Camera. Suppose that surfaces 1 and 3 are visible to both sensor types, while surface 2 is visible to sensor type Y alone, as shown in Figure 3.2. Let $\{RGB\}$ denote the total set of measurements of the scene:

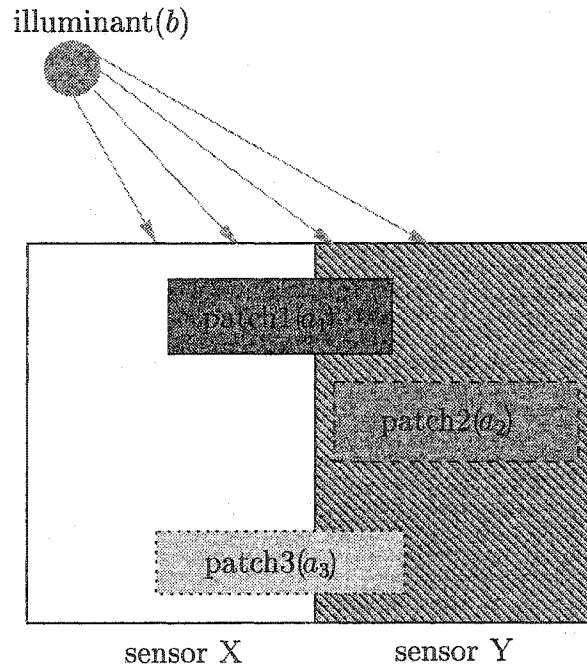


FIGURE 3.2. Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part).

$$\{RGB\} \triangleq RGB_{X1}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3}.$$

The conditional posterior density function for the parameters, a_1, a_2, a_3 and b , given the set of measurements of the scene, $\{RGB\}$, is given by:

$$\begin{aligned}
 & p(a_1, a_2, a_3, b | \{RGB\}) \\
 & \approx p(\{RGB\} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
 & \approx p(RGB_{X1}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
 & \approx p(RGB_{X1}, RGB_{X3} | a_1, a_3, b) p(RGB_{Y1}, RGB_{Y2}, RGB_{Y3} | a_1, a_2, a_3, b) p(a_1, a_3, b) p(a_2) \\
 & \approx p(a_1, a_3, b | RGB_{X1}, RGB_{X3}) p(RGB_{Y1} | a_1, b) p(RGB_{Y2} | a_2, b) p(RGB_{Y3} | a_3, b) p(a_2).
 \end{aligned} \tag{3.2}$$

The assumptions made in obtaining this formulation, in addition to the statistical independence of the measurements obtained, are stated here. First, it is assumed that the prior probabilities for each surface reflectance weight a_n are statistically independent of each other and of the spectral function weights of the illuminant, b . It is also assumed that there is no interreflection between surfaces, and therefore the spectral reflectance for a surface and the lighting vectors are sufficient statistics for the measurement of that surface. This leads to a simplification:

$$p(RGB_{Y2} | a_1, a_2, a_3, b) = p(RGB_{Y2} | a_2, b).$$

From Equation 3.2, we can conclude that the posterior for the entire scene is a function of the posterior for one sensor (sensor X), the likelihood of the RGB measurement of the other sensor (sensor Y) and the prior of its spectral function weights. Given the assumptions above, the posterior for sensor X can be obtained by:

$$\begin{aligned}
 & p(a_1, a_3, b | RGB_{X1}, RGB_{X3}) \\
 & \approx p(RGB_{X1}, RGB_{X3} | a_1, a_3, b) p(a_1, a_3, b)
 \end{aligned}$$

$$\begin{aligned}
&\approx p(RGB_{X1}|a_1, b) p(RGB_{X3}|a_3, b) p(a_1) p(a_3) p(b) \\
&\approx p(a_1, b|RGB_{X1}) p(RGB_{X3}|a_3, b) p(a_3).
\end{aligned} \tag{3.3}$$

Note that this posterior is a function of the likelihood and prior for surface 3, and the posterior for surface 1.

3.1.2.2. *Moving Camera.* Passive vision may be very problematic due to ill-posedness. Active vision techniques have been proposed to overcome some of these problems [3]. The idea of active vision occurred to scientists from the way humans perceive the world around them; that is humans always move their eyes, thus constantly updating the image of the scene in front of them. This allows them to acquire better knowledge of the world around them. Therefore, in our model, the camera is moved to acquire multiple images of the same scene, similar to when a person moves his/her gaze across the scene, and thus more information is gathered about the surface patches.

Two cases are examined here. The first one is when the camera is moved once (*One Move*), and the second is when the camera is moved twice (*Two Moves*).

One Move. In this case, the camera is moved once such that two images of the scene are taken from two different positions. This makes the surface patches in different locations with respect to the filter as shown in Figure 3.3. Two posteriors are derived; the first one is P_1 , which is that obtained from the first image after gathering the first set of data and P_2 , which is that obtained from the second image after the camera is moved and thus more data is gathered. P_1 and P_2 are derived as shown below.

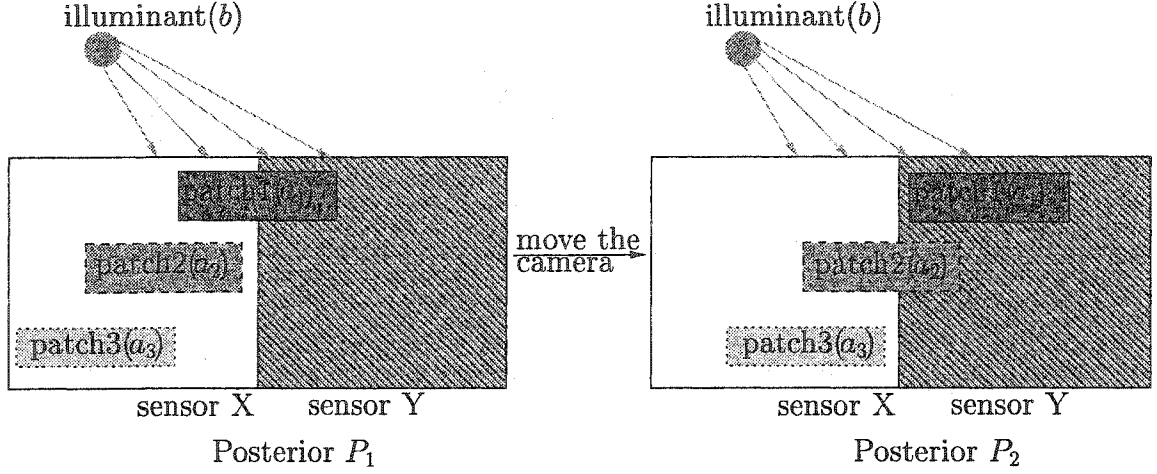


FIGURE 3.3. Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part): The camera is moved once.

At the initial stage, that is before the camera is moved, the total set of measurements denoted by $\{RGB\}$ is given by:

$$\{RGB\} \triangleq RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1}.$$

The posterior can then be computed given the set of measurements $\{RGB\}$ from the scene:

$$\begin{aligned}
 P_1 &= p(a_1, a_2, a_3, b | \{RGB\}) \\
 &\approx p(\{RGB\} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
 &\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
 &\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3} | a_1, a_2, a_3, b) p(RGB_{Y1} | a_1, b) p(a_1, a_2, a_3, b) \\
 &\approx p(a_1, a_2, a_3, b | RGB_{X1}, RGB_{X2}, RGB_{X3}) p(RGB_{Y1} | a_1, b). \tag{3.4}
 \end{aligned}$$

The next step involves the movement of the camera and gathering a new set of *RGB* measurements, which includes the old set and is given by:

$$\{RGB\} \triangleq RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}.$$

P_1 acts as a prior for P_2 , which can now be computed given the new *RGB* measurements:

$$\begin{aligned}
P_2 &= p(a_1, a_2, a_3, b | \{RGB\}) \\
&\approx p(\{RGB\} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
&\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1}, RGB_{Y2} | a_1, a_2, a_3, b) \\
&\quad p(a_1, a_2, a_3, b) \\
&\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3} | a_1, a_2, a_3, b) p(RGB_{Y1}, RGB_{Y2} | a_1, a_2, a_3, b) \\
&\quad p(a_1, a_2, a_3, b) \\
&\approx p(a_1, a_2, a_3, b | RGB_{X1}, RGB_{X2}, RGB_{X3}) p(RGB_{Y1} | a_1, b) p(RGB_{Y2} | a_2, b) \\
&\approx P_1 p(RGB_{Y2} | a_2, b). \tag{3.5}
\end{aligned}$$

Two Moves. In this case, the camera is moved twice such that three images of the scene are taken from three different positions. This makes the surface patches in different locations with respect to the filter as shown in Figure 3.4. In this case, three posteriors are obtained: one from the set of data of each image. P_1 and P_2 can be derived in a similar way as that shown in Equations 3.4 and 3.5. As explained in the previous section, P_1 acts as a prior for P_2 . In this section, the camera is moved a second time, and a new set of data is gathered, and thus a new set of *RGB* measurements is obtained in addition to the old one that was obtained from the two

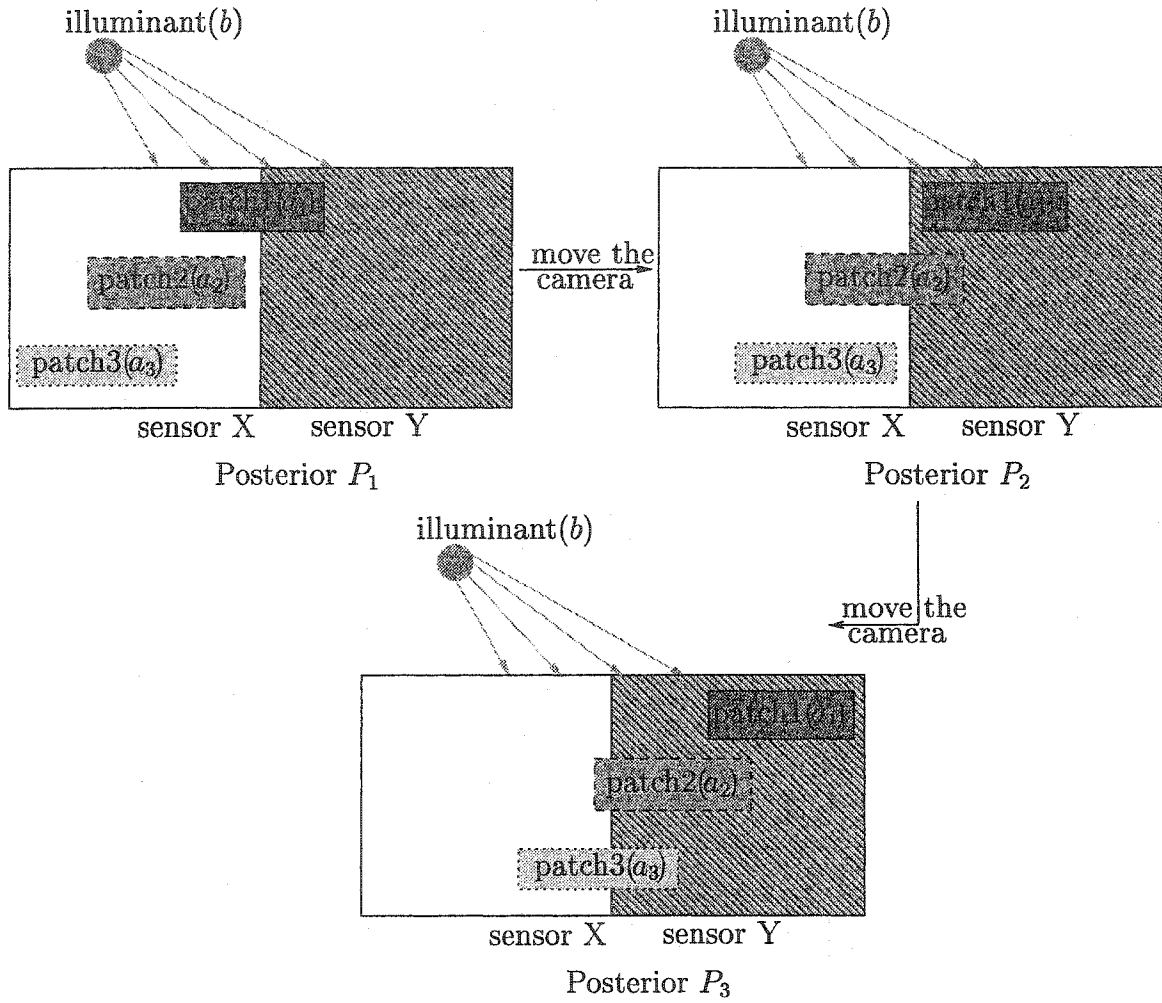


FIGURE 3.4. Three patches in the scene, illuminated with a single light source viewed with two sensors or a binary filter, which has one clear part and one part tinted yellow, for example, as discussed in Chapter 4 (represented with the hashed part): The camera is moved twice.

initial movements. This new set of RGB measurements is given by the following equation:

$$\{RGB\} \triangleq RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3}.$$

P_2 acts as a prior for P_3 :

$$\begin{aligned}
P_3 &= p(a_1, a_2, a_3, b | \{RGB\}) \\
&\approx p(\{RGB\} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
&\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3}, RGB_{Y1}, RGB_{Y2}, RGB_{Y3} | a_1, a_2, a_3, b) p(a_1, a_2, a_3, b) \\
&\approx p(RGB_{X1}, RGB_{X2}, RGB_{X3} | a_1, a_2, a_3, b) p(RGB_{Y1} | a_3, b) p(RGB_{Y2} | a_3, b) \\
&\quad p(RGB_{Y3} | a_3, b) p(a_1, a_2, a_3, b) \\
&\approx p(a_1, a_2, a_3, b | RGB_{X1}, RGB_{X2}, RGB_{X3}) p(RGB_{Y1} | a_3, b) \\
&\quad p(RGB_{Y2} | a_3, b) p(RGB_{Y3} | a_3, b) \\
&\approx P_2 p(RGB_{Y3} | a_3, b). \tag{3.6}
\end{aligned}$$

The final posterior is P_3 , which is obtained after the camera is moved twice and the most data is gathered. In the same way as information is accumulated over different sensors, information is accumulated over scenes, such that the posterior for each image acts as a prior for the set of data of the next image. The posteriors obtained from images due to additional moves can be derived in a similar way.

In short, the active Bayesian formulation is, in fact, *recursive*, where the posterior for each sensor acts as a prior for the next sensor. This is similar to the recursive nature of the strategy in the case of the surface patches, where the posterior of each surface patch acts as a prior for the next surface patch. The addition of more surface patches in the scene provides more information regarding the illumination and thus, in turn, about the surface colors themselves. On the other hand, moving the camera provides more information about the scene as the surface patches are viewed from different locations, and thus many images of the same scene are obtained.

3.2. The Non-Uniform Formulation

The non-uniform formulation is used for the case of a smooth variation in the spectral sensitivities across the sensor array. The thickness of the covering of the macula, which is the small, yellowish central portion of the retina, drops off with eccentricity. Also, there is a drop-off in the path length of light through the lens, which focuses light onto the retina. A Gaussian falloff curve is used to model these effects. In this section, as opposed to the previous section, the indices for the RGB 's are reversed. This means that the RGB measurements of a particular surface obtained from sensors of type X will be denoted RGB_{nX} , where n refers to the surface patch being viewed.

3.2.1. A Single Patch. In this case there is only one surface patch in the scene, illuminated with a single light source. The weight vector for the surface patch spectrum model shall be denoted by a_1 , while that for the light source spectrum model shall be denoted by b . Suppose that the surface has N sensor responses. Thus the sensor responses can be denoted as $\{RGB\}_{11\dots 1N} \triangleq RGB_{11}, RGB_{12}, \dots, RGB_{1N}$ where N is the number of sensor responses of the surface patch.

To make the idea simpler, we shall assume that the patch has two sensor responses for now, which are 1 and 2 only. Therefore, the total sensor responses can be denoted by $\{RGB\} \triangleq RGB_{11}, RGB_{12}$. The posterior for a scene that has a patch with such specifications in this case can be formulated as:

$$\begin{aligned}
 & p(a_1, b | \{RGB\}) \\
 & \approx p(\{RGB\} | a_1, b) p(a_1, b) \\
 & \approx p(RGB_{11}, RGB_{12} | a_1, b) p(a_1, b) \\
 & \approx p(RGB_{11} | a_1, b) p(RGB_{12} | a_1, b) p(a_1, b)
 \end{aligned}$$

$$\approx p(a_1, b | RGB_{11}) p(RGB_{12} | a_1, b). \quad (3.7)$$

This means that if a person is moving his/her gaze from a pixel that has a sensor response of type 1 to one that has a sensor response of type 2, then the final posterior would be a product of the posterior given that the person has seen the pixel of sensor response of type 1 and the likelihood of the pixel the person is looking at right now, which is the pixel that has sensor response of type 2. In a similar way, the posterior for a scene containing a surface patch with N sensor responses (as described above) can be derived, as shown in Equation 3.8 and based on the assumptions made in Section 3.1.

$$\begin{aligned} & p(a_1, b | \{RGB\}_{11 \dots 1N}) \\ & \approx p(a_1, b | \{RGB_{11 \dots 1(N-1)}\}) p(RGB_N | a_1, b). \end{aligned} \quad (3.8)$$

This means that the posterior of a surface patch is a product of the posterior of all the surface patch pixels, except the last, that have been seen, and the likelihood of the last pixel that is being looked at right now.

In short, the posterior for each surface is simply a recursive, sequential update of the posterior given each of the sensor responses for that surface. Moreover, the addition of more surface patches in the scene provides more information regarding the illumination, and therefore, about the surface colors themselves.

3.2.2. Multiple Patches. In this section, the case where there are multiple (three) surface patches illuminated with a single light source in the scene is considered. Surfaces 1, 2, and 3 have N , M and K sensor response functions, respectively. Thus, the sensor responses can be denoted as $\{RGB\}_{n1 \dots np} \triangleq RGB_{n1}, RGB_{n2}, \dots, RGB_{np}$ for each surface n , where p is the number of sensor responses for that surface.

3.2 THE NON-UNIFORM FORMULATION

As a start, suppose we move our gaze from the last pixel of surface 1, for example, to the first pixel of surface 2. Therefore, the total RGB set of measurements of the scene, given what we have seen so far, is given by:

$$\{RGB\} \triangleq RGB_{11}, RGB_{12}, \dots, RGB_{1N}, RGB_{21}. \quad (3.9)$$

The posterior for the scene based on what has been seen so far can be formulated as:

$$\begin{aligned} & p(a_1, a_2, b | \{RGB\}) \\ & \approx p(\{RGB\} | a_1, a_2, b) p(a_1, a_2, b) \\ & \approx p(RGB_{11}, RGB_{12}, \dots, RGB_{1N}, RGB_{21} | a_1, a_2, b) p(a_1, a_2, b) \\ & \approx p(RGB_{11}, RGB_{12}, \dots, RGB_{1N}, | a_1, b) p(RGB_{21} | a_2, b) p(a_1, b) p(a_2) \\ & \approx p(a_1, b | RGB_{11}, RGB_{12}, \dots, RGB_{1N}) p(RGB_{21} | a_2, b) p(a_2). \end{aligned} \quad (3.10)$$

Equation 3.10 shows that as we move our gaze from one surface to another, the posterior for the scene becomes a product of the posterior of the scene as seen previously, the likelihood of the first pixel of the new surface that is being looked at, and the prior for that surface. We can observe that the subjective prior term (e.g. $p(a_2)$) only comes in at the start of the recursion for each surface. In other words, the posterior of the previous scene, given that we have seen surface 1 only, which is formulated in Equation 3.8 acts as a prior for the new posterior, which is formulated in Equation 3.10.

The measurements for the entire scene are given by the following equation:

$$\begin{aligned} \{RGB\} \triangleq & RGB_{11}, RGB_{12}, \dots, RGB_{1N}, RGB_{21}, RGB_{22}, \dots, RGB_{2M}, RGB_{31}, \\ & RGB_{32}, \dots, RGB_{3K}. \end{aligned} \quad (3.11)$$

Based on the assumptions made in the previous section, the posterior for the entire scene is again recursive and can be derived in a similar way to that used to derive Equations 3.8 and 3.10. It is stated in Equation 3.12.

$$\begin{aligned}
& p(a_1, a_2, a_3, b | \{RGB\}) \\
& \approx p(a_1, a_2, a_3, b | \{RGB\}_{11\dots 1N}, \{RGB\}_{21\dots 2M}, \{RGB\}_{31\dots 3K}) \\
& \approx p(a_1, a_2, b | \{RGB\}_{11\dots 1N}, \{RGB\}_{21\dots 2M}) p(\{RGB\}_{31\dots 3K} | a_3, b) p(a_3). \quad (3.12)
\end{aligned}$$

This implies that the posterior is computed by multiplying the prior and likelihood for the third surface by the posterior from the first two surfaces, i.e. these posteriors become priors for the third surface. In short, the posterior for each surface is simply a recursive, sequential update of the posterior given each of the sensor responses for that surface.

For both of the cases described here we obtain a sequential update of the posteriors given each of the sensors in turn. The result is that, for each surface patch, each new sensor response leads to an increase in information regarding the color of the surface patch and the illumination. The addition of more surface patches in the scene provides even more information regarding the illumination and thus, in turn, about the surface colors themselves.

CHAPTER 4

Simulation Results

In this chapter, we demonstrate the approach developed in Chapter 3 of this thesis by presenting a series of simulations that were performed. We constructed “Mondrian” scenes consisting of one patch, on one hand, and scenes consisting of several surface patches, on the other hand. A “Mondrian” is a planar surface composed of several, overlapping, matte (lambertian) patches [22]. These types of images are named after the style of painting produced by the artist Piet Mondrian, who died in 1944. A lambertian surface is one that reflects light uniformly in all directions. The light illuminating a Mondrian scene is assumed locally constant; that is the spectral characteristics of the light vary slowly. Almost all color constancy algorithms are defined for the Mondrian world. All scenes were illuminated by a single light source.

Three cases were simulated. We first simulated the Brainard-Freeman case, where there is no filter on the camera lens. Then the multi-sensor environment was introduced, and it was modeled through the simulated placement of an optical filter onto the (simulated) camera lens, which non-uniformly modifies the spectrum of the light falling on the sensors. We then simulated the binary case, in which half of the filter

is tinted yellow with the other half left transparent, and finally simulated the non-uniform case, in which a filter with a Gaussian absorption profile is used; the center is tinted deep yellow, and the color fades smoothly towards the periphery.

Section 4.1 describes the algorithm used in the simulations and how the components of the Bayesian formulation were obtained. Then Section 4.2 illustrates the behavior of our model by presenting the simulation results, which is mainly our model's ability to give a reasonable estimate of surface patches and light source colors.

4.1. The Algorithm

The *RGB* components of each sensor's response were simulated by multiplying the spectrum of the light passing through the optical filter by the sensor spectral sensitivity curves, $R_k(\lambda)$, where k is the sensor number and λ is the wavelength. In order to simulate the real experiments, which we shall see in Chapter 5, noise was added to the *RGB* measurements. Sources of noise in the real world could be the flickering of the light source while taking the images, dust on the camera lens, and noise in the spectral response of the camera color filter. This noise is random and has a normal distribution, which has a standard deviation of 10% of the entire range of measurements for each of the red, green, and blue. In other words, the standard deviation is $0.1 * (max_{R,G,B} - min_{R,G,B})$. For the purposes of our simulations, we used the Stockman and Sharpe estimates for the sensitivities of the cones in the human retina [39]. Randomly selected *matte* Munsell color patches (see Appendix A for details on Munsell patches) [2] and random combinations of the Parkkinen and Silftsen daylight and skylight spectra [35] to generate the spectrum of the light falling on the sensor array were used.

As for the Bayesian formulations, the likelihoods were computed using the model predictions of the sensor measurements, which are given by Equation 2.3:

$$\rho_k^x = \int E(\lambda) S^x(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, \dots, p.$$

The illuminant spectrum $E(\lambda)$ is modeled with five basis functions, while the surface spectrum $S^x(\lambda)$ is modeled with eight basis functions. They were assumed to have Gaussian distributions. The basis functions for the surface spectrum model were taken to be the principal components (see Section 2.1.1 for details on principal components analysis) of the spectra of the 1250 Munsell color patches as measured by Parkkinen *et al.* [36]. The basis functions for the light source spectrum model were taken to be the principal components (see Section 2.1.1) of the set of daylight and skylight spectra as measured by Parkkinen and Silftsen [35]. The first five basis functions of each of the 1250 Munsell color patches and the daylight and skylight spectra are plotted in Figure 4.1. The number of basis functions plotted was limited to five for reasons of clarity. Note that the basis functions start to become noisy after the first five. Only the first eight are used for modeling the surface spectrum, while only the first five are used for modeling the light source spectrum in the algorithm. The prior distributions for the spectral model parameters are assumed to be independent and Gaussian functions. The means and variances of the priors were computed from the distribution of weights corresponding to the 1250 Munsell spectra and the 37 different skylight and daylight spectra that were used. These weights were obtained by projecting the measured spectra onto the basis function sets.

The location of the maximum value of the posterior distribution was estimated by a standard MATLAB optimization package. Since MATLAB does not have a function that searches for a maximum value, the minimum of the negative of the posterior distribution was searched for instead. The function used was “lsqnonlin”,

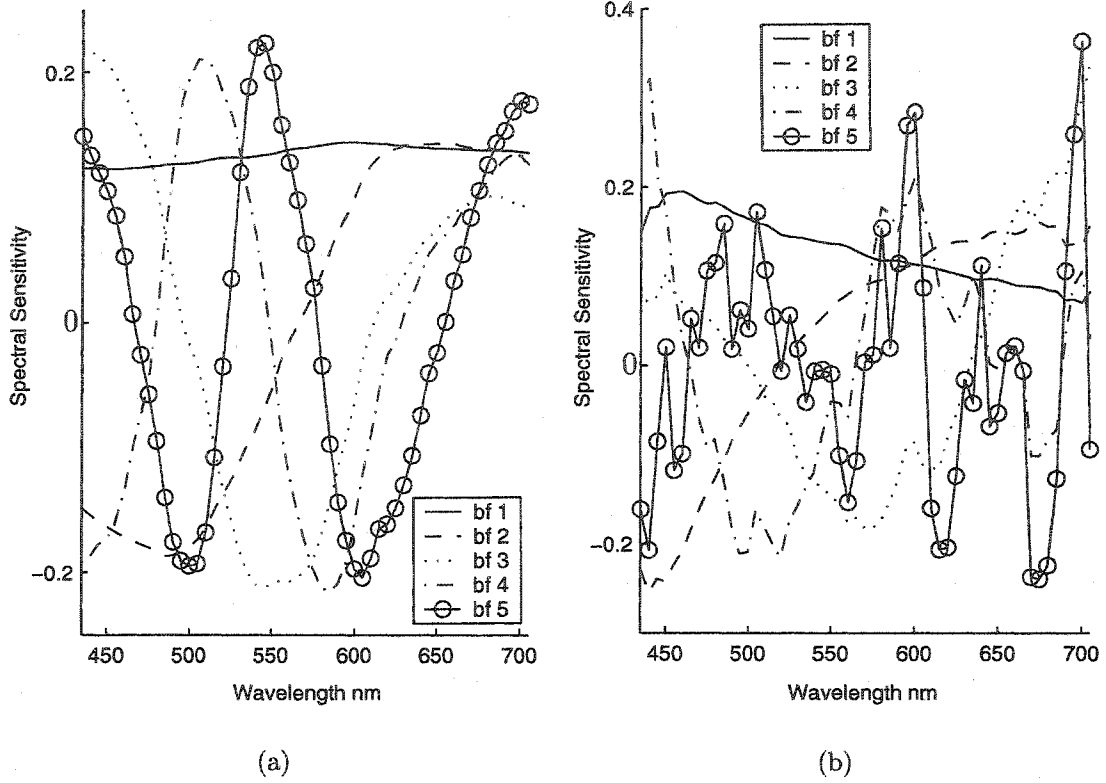


FIGURE 4.1. (a) Surface basis functions and (b) Daylight and Skylight basis functions. “bf” stands for “basis function”.

which solves nonlinear least squares problems of the form:

$$\min_x f(x) = f_1(x)^2 + f_2(x)^2 + f_3(x)^3 + \dots + f_m(x)^2 + L, \quad (4.1)$$

where L is a constant. In Equation 4.1, x is the solution or the global minimum to which the optimization algorithm converges after an initial value x_0 is given to it. In our application, x_0 is a vector of surface patch and light source spectra weight vectors. The limitations of the algorithm is that it may only give local solutions and that the cost function $f(x)$ in Equation 4.1 must be continuous. Therefore, choosing the initial point is a very crucial issue. The searching scheme the “lsqnonlin” uses is the “Levenberg-Marquardt” algorithm [31]. A point worth mentioning here is that

the posterior function is a product of Gaussian functions; therefore, the logarithm of this function is taken in order to convert it into a quadratic function, which would have a form similar to that of Equation 4.1. After convergence, the optimization results in a set of estimated surface patches and light source spectra weight vectors. As mentioned in Chapter 2, the MAP (maximum a posteriori) estimator is used although it is not the best one, but it is not very computationally expensive.

4.2. Simulation Results

Experiments with either one or several surface patches in the scene, illuminated with a single light source (as described in Chapter 3) were performed, and the resulting estimates for the model and the actual spectra for the surface patches and the light sources were plotted and the root mean square (RMS) errors were computed.

4.2.1. Single versus Multiple Patches. In this case, two cases were compared. On one hand, there was only one patch in the scene, illuminated with a single light source. On the other hand, there were multiple patches in the scene, illuminated with the same light source. Two scenarios were considered: the Brainard and Freeman case or equivalently, the ‘no filter’ case, and the binary filter case. Figures 4.2 and 4.3 show the results, and Table 4.1 depicts the RMS errors between the measurement and the model spectra of the illuminant and surface patch spectra for the ‘no filter’ case.

In this case there is a slight improvement for the surface patch estimation but not the illuminant as can be seen upon examining the RMS errors. However, we can notice the improvement in the RMS error upon introducing the filter by comparing the results in Tables 4.1 and 4.2.

In short, introducing more surface patches into the scene means having more scene parameters: 8 for each surface and 5 for the light source, which makes a total

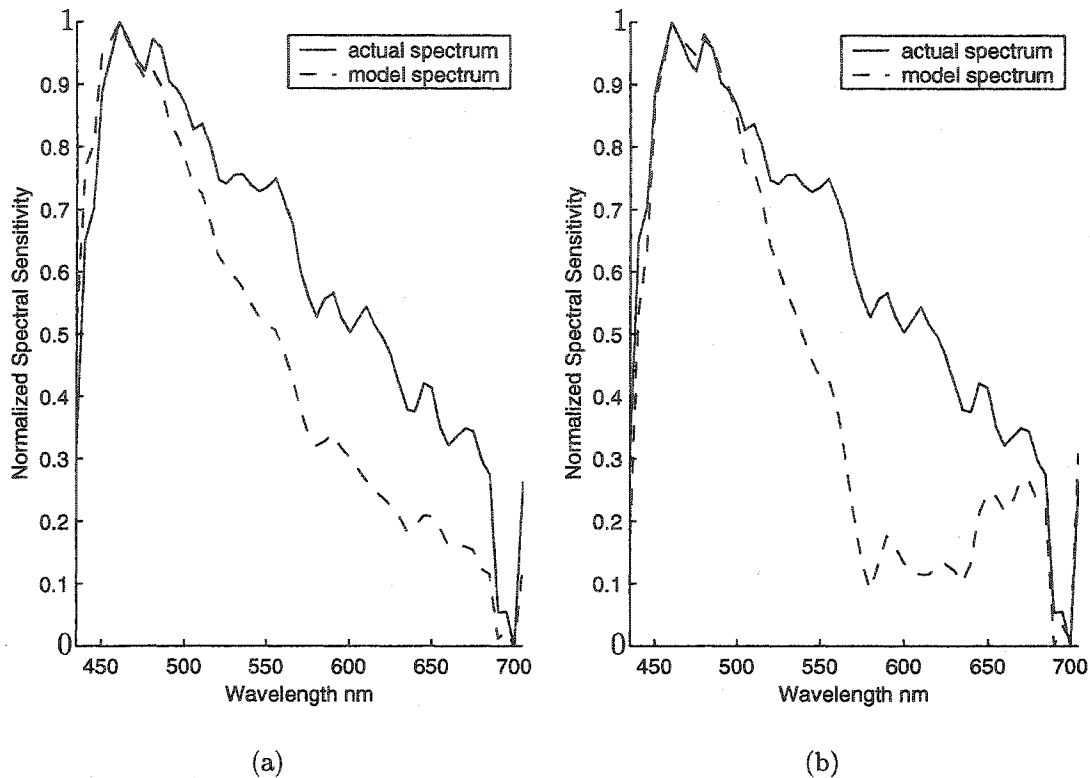


FIGURE 4.2. The estimated and actual spectra of the illuminant (number 9 from Parkkinen's set of spectra) for the 'no filter' case: (a) one patch in the scene, (b) multiple patches in the scene.

Case	Illuminant	Surface Patch
Single Patch	0.1705	0.2313
Multiple Patches	0.2326	0.2713

TABLE 4.1. RMS error for the illuminant (number 9 from Parkkinen's set of spectra) and surface patch (Munsell patch 875) spectra for the 'no filter' case when there is either a single patch or multiple patches in the scene.

of 29 parameters. On the other hand, in the one surface patch scenario, there are only $8 + 5 = 13$ scene parameters. The only information available from the image of the scene is an *RGB* value for each patch. For the three patches case, there are 9 measurements, while for the one patch case, there are 3 measurements. Brainard and Freeman mention in [7] that the inverse problem formulated in Equation 2.3 is difficult

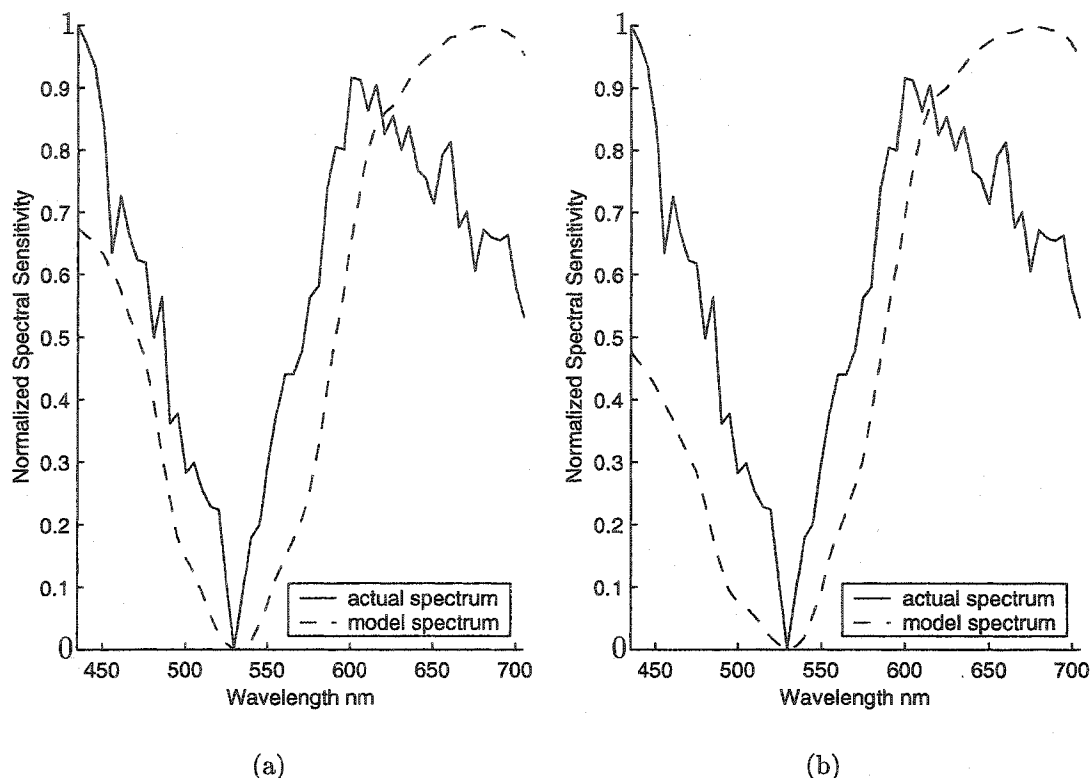


FIGURE 4.3. The estimated and actual spectra of a surface patch (Munsell patch 875) for the 'no filter' case: (a) one patch in the scene, (b) multiple patches in the scene.

because it is underdetermined and nonlinear. An inverse problem is underdetermined if there are more scene parameters than there are degrees of freedom in the data. This can be observed in our case in both the multiple patch case where there are 29 scene parameters versus 9 degrees of freedom in the data, on one hand, and in the single patch case where there are 13 scene parameters versus 3 degrees of freedom in the data, on the other hand. Having more scene parameters poses a harder problem for the optimization to solve and for us to choose an initial value that will not make it get trapped in local minima. Trying multiple initial values helps but does not solve the problem completely because we cannot try the whole space of starting points. Since multiple objects are what we see in everyday life, we focus on this scenario in the

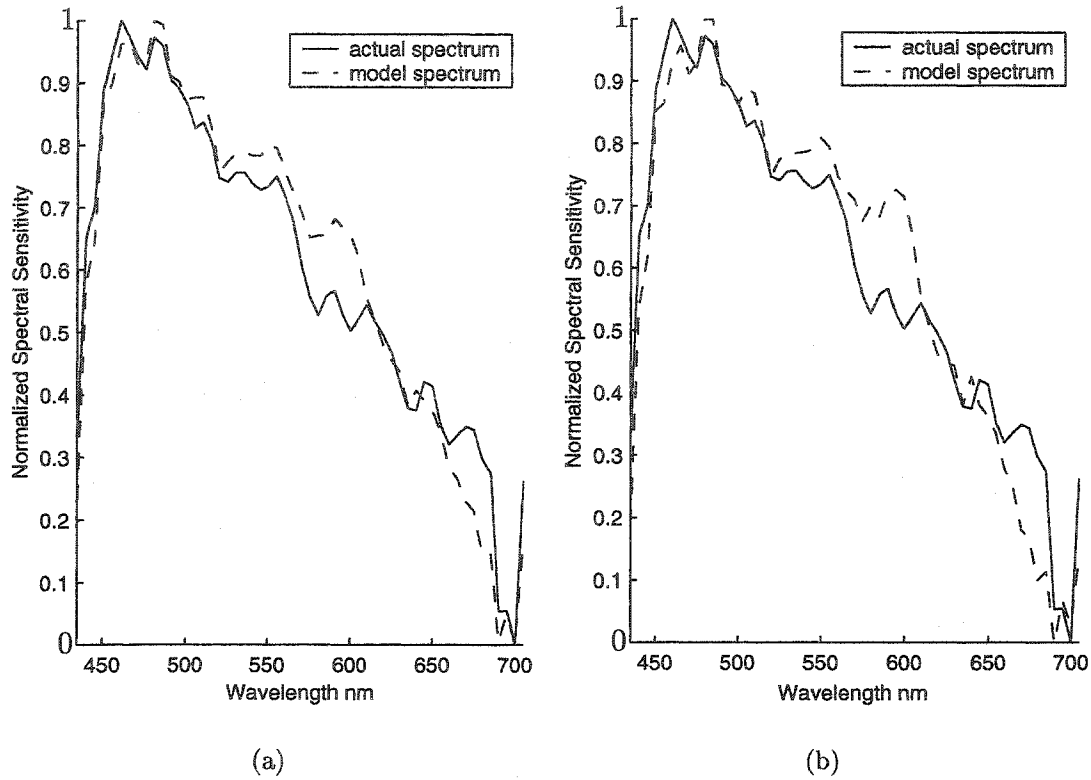


FIGURE 4.4. The estimated and actual spectra of the illuminant (number 9 from Parkkinen's set of spectra) for the binary filter case: (a) one patch in the scene, (b) multiple patches in the scene.

sections yet to follow. We try to improve the performance of Brainard and Freeman's algorithm by examining what happens upon the introduction of the different types of filters and moving the camera, i.e. the notion of active vision.

Case	Illuminant	Surface Patch
Single Patch	0.0666	0.1910
Multiple Patches	0.0895	0.1887

TABLE 4.2. RMS error for the illuminant (number 9 from Parkkinen's set of spectra) and surface patch (Munsell patch 875) spectra for the binary case when there is either a single patch or multiple patches in the scene.

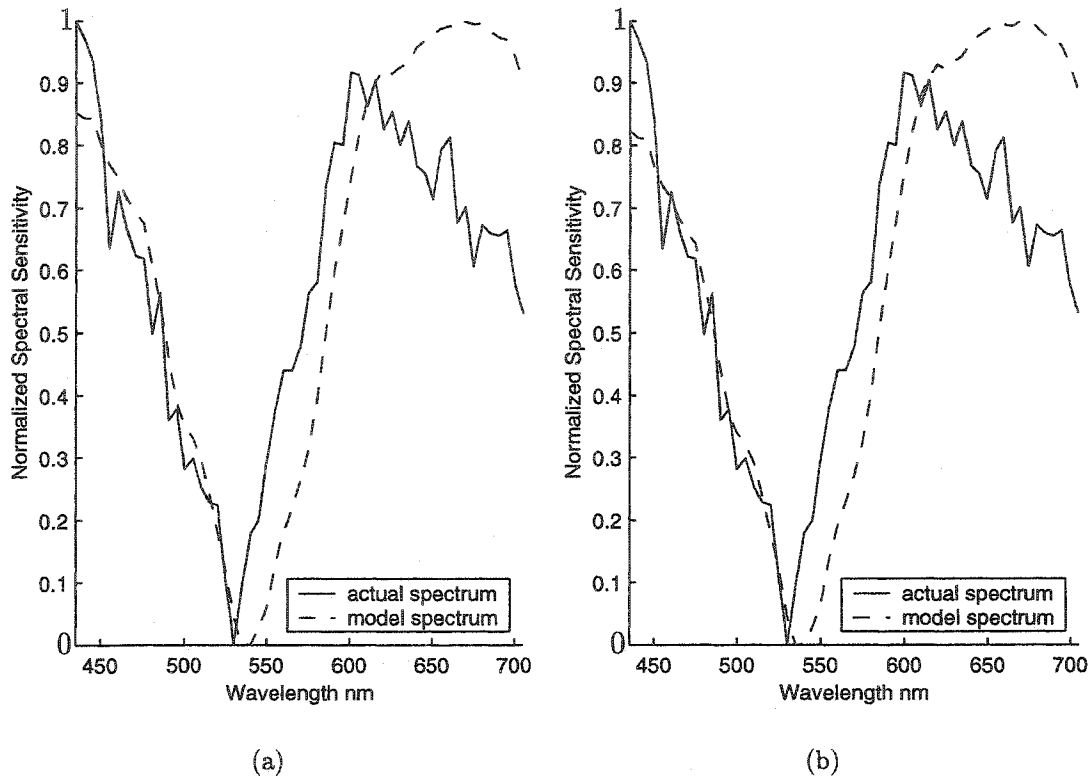


FIGURE 4.5. The estimated and actual spectra of a surface patch (Munsell patch 875) for the binary filter case: (a) one patch in the scene, (b) multiple patches in the scene.

4.2.2. The Different Sensors: A Comparison. Experiments with several surface patches in the scene were performed. The resulting estimates for the model and the actual spectra for one surface patch in the scene as well as the spectra for the light source can be found in Figures 4.6 and 4.7 respectively. In each case, the figures illustrate the spectra for the two different types of optical filters as compared to the case with no filter. The results indicate that there is considerable improvement with the introduction of the multi-sensor method. In fact, the more sensors introduced, the better the estimates. This can also be seen by examining the RMS errors depicted in Table 4.3.

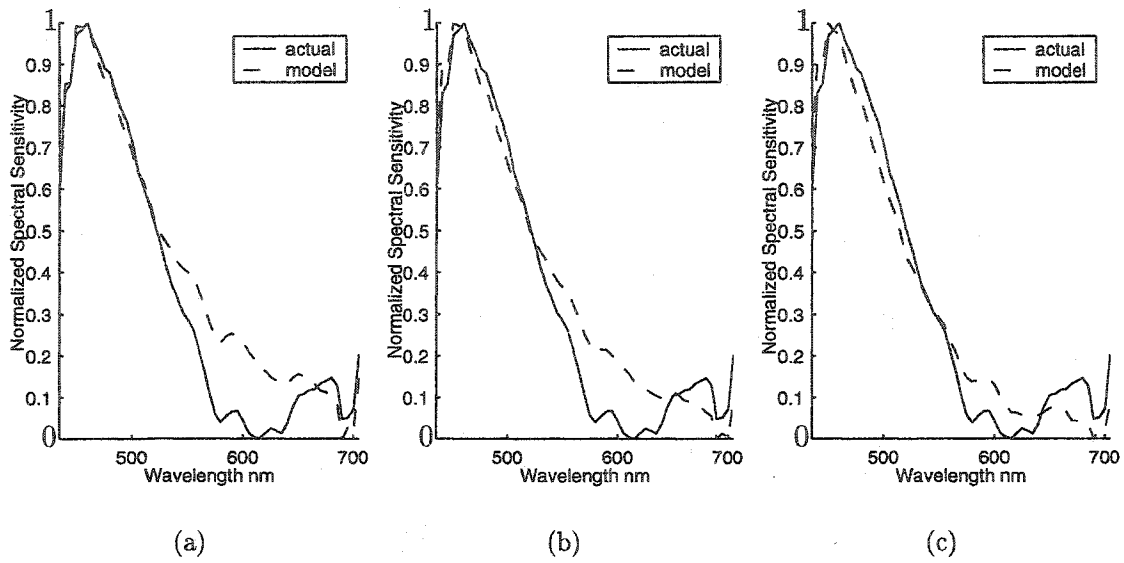


FIGURE 4.6. The estimated and actual spectra of the illuminant (number 2 from Parkkinen's set of spectra) when there are multiple surface patches in the scene in the cases of (a) 'no filter', (b) binary filter, (c) Gaussian filter.

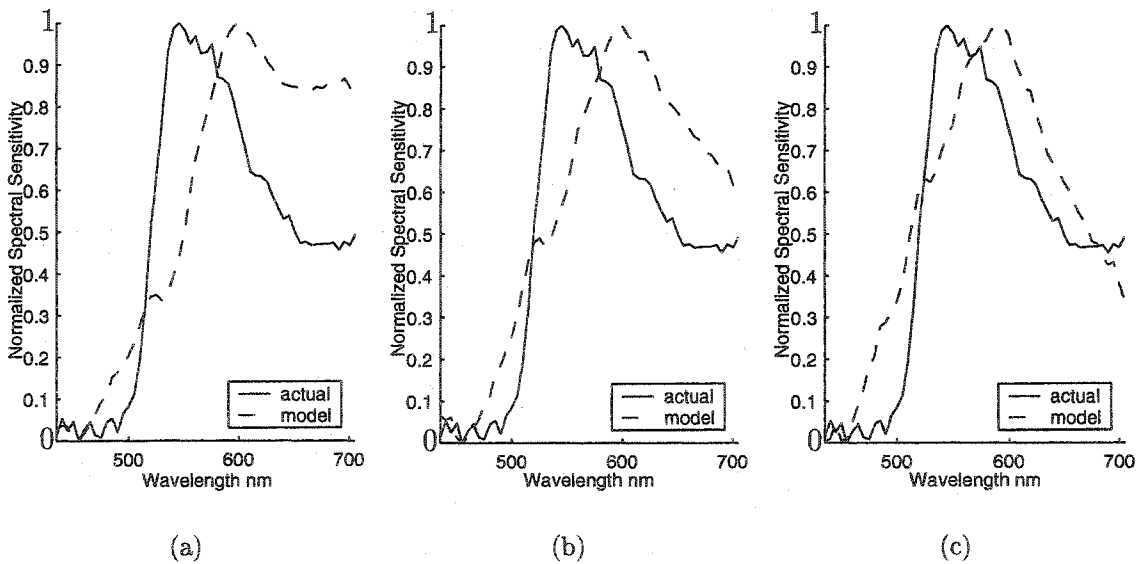


FIGURE 4.7. The estimated and actual spectra of a surface patch (Munsell patch 300) when there are multiple surface patches in the scene in the cases of (a) 'no filter', (b) binary filter, (c) Gaussian filter.

Case	No Filter	Binary Filter	Gaussian Filter
Illuminant	0.1007	0.0867	0.0634
Surface	0.2885	0.2232	0.1584

TABLE 4.3. RMS error for the illuminant and surface patch spectra for the ‘no filter’, binary filter and Gaussian filter cases when there are multiple surface patches in the scene.

4.2.3. Moving Camera. Another type of simulation which is also described in Section 3.1.2.2 is the “moving camera” case. There are three surface patches in the scene with two types of sensors, that is a binary filter on the lens having a clear part and a tinted-yellow part. Then the camera is moved: one move and then two moves. With each movement, new measurements for each surface patch are obtained, and thus more information is gathered about the scene. This gives a tendency to improve the results. Figures 4.8 and 4.9 show the model and the actual spectra for both the light source as well as one of the surface patches of the scene respectively. The RMS errors are depicted in Table 4.4. For comparison purposes the ‘no filter’ case is shown as well.

Case	Illuminant	Surface Patch 400
No Filter	0.1886	0.1310
Binary Filter, No Motion	0.1715	0.1100
Binary Filter, 1 Move	0.0641	0.0742
Binary Filter, 2 Moves	0.0459	0.0519

TABLE 4.4. RMS error for the illuminant (number 30 from Parkkinen’s set of spectra) and surface patch spectra (Munsell patch 400) when there are multiple surface patches in the scene for the ‘no filter’, and binary filter case when the camera is: not moved, moved once, and then moved twice.

In short, after observing the RMS errors in Table 4.4, we might say that the improvement is not much and that obtained from the Gaussian filter (see Table 4.3) is better. However, here we only considered moving the camera twice for illustration

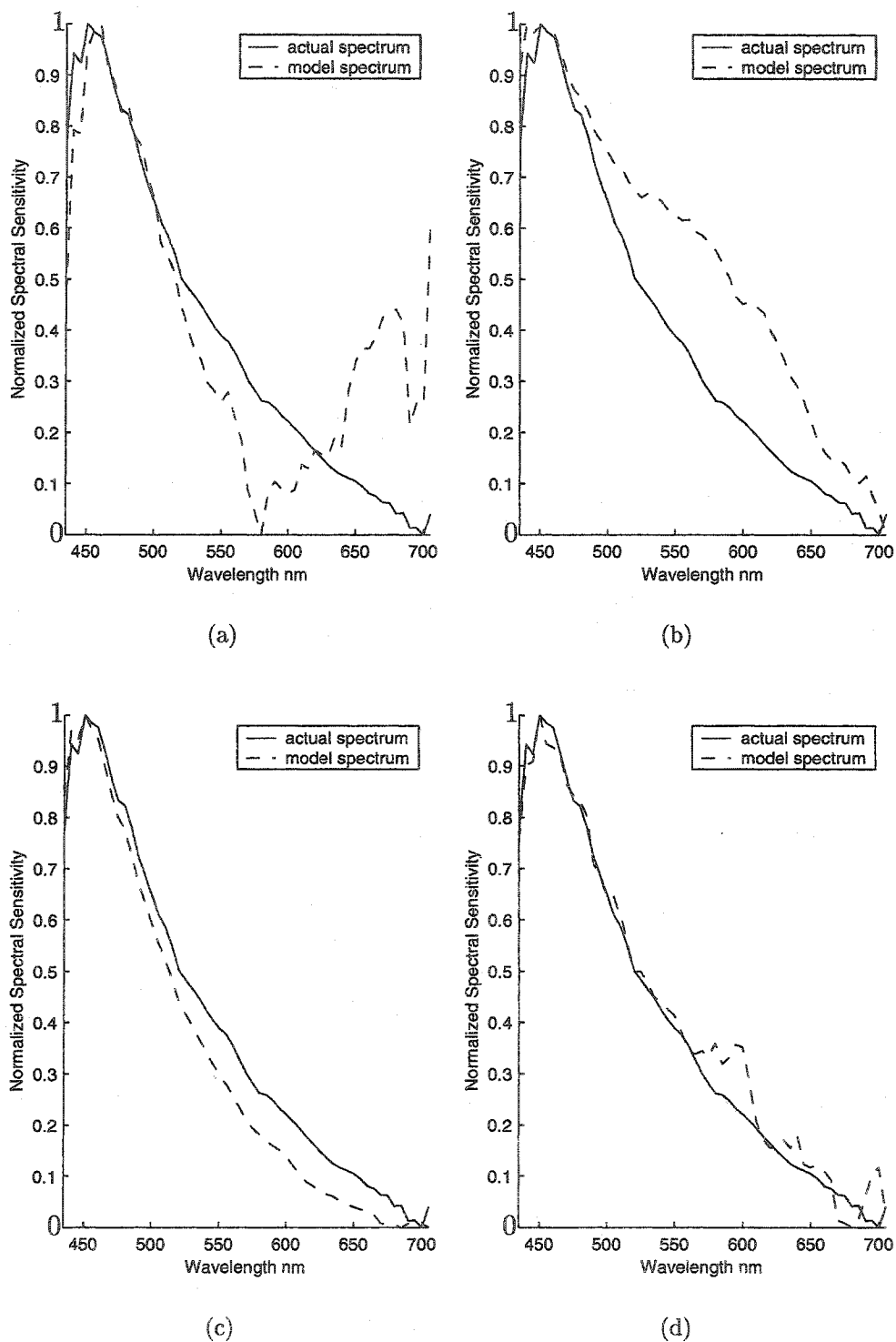


FIGURE 4.8. The estimated and actual spectra of the illuminant (number 30 from Parkkinen's set of spectra) when there are multiple patches in the scene in the (a) 'no filter' case and in the binary filter case when the camera is: (b) not moved, (c) moved once, (d) moved twice.

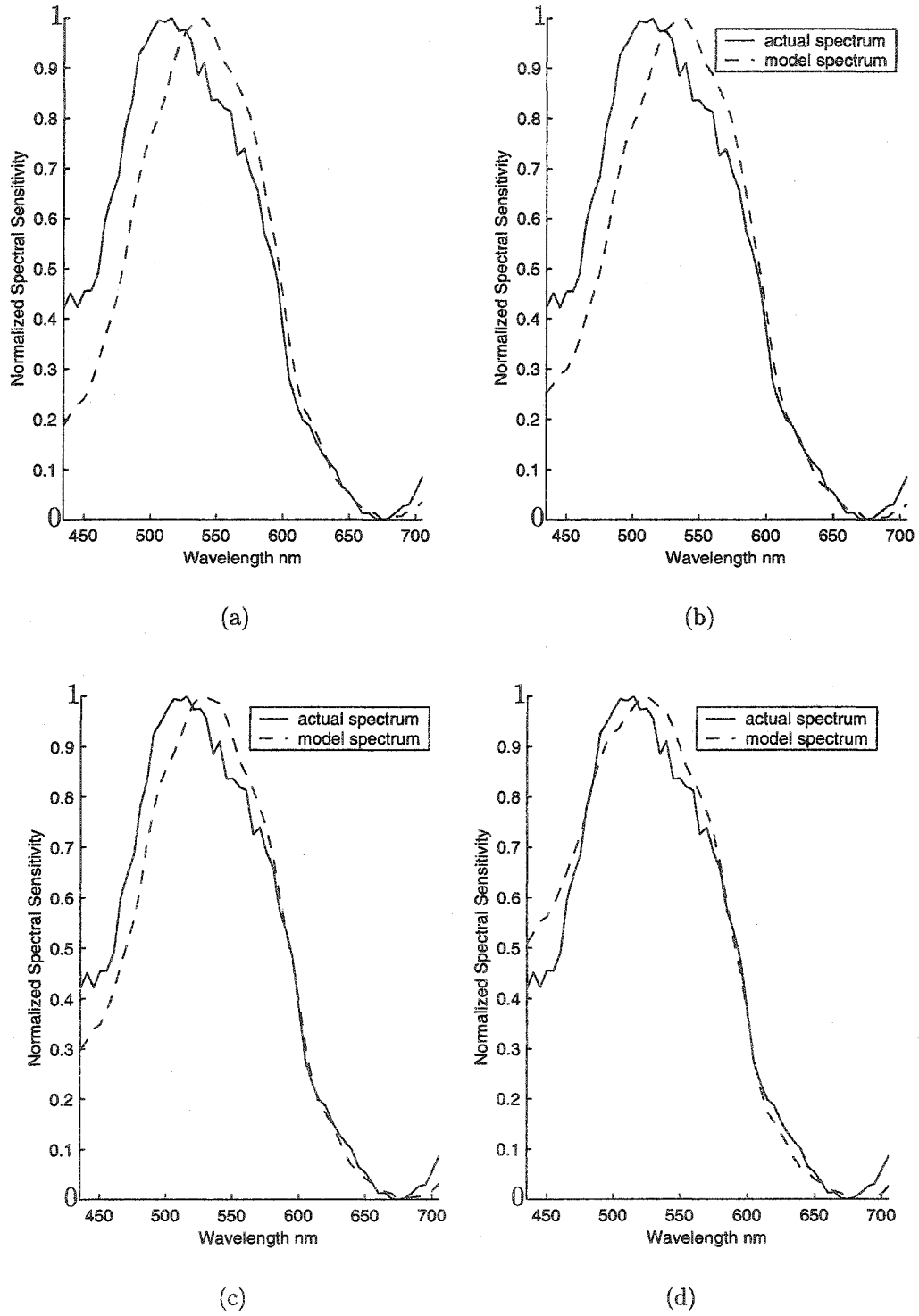


FIGURE 4.9. The estimated and actual spectra of a surface patch (Munsell patch 400) when there are multiple patches in the scene in the (a) 'no filter' case and in the binary filter case when the camera is: (b) not moved, (c) moved once, (d) moved twice.

purposes. Had we moved the camera multiple times, we would have obtained new information with each new movement and thus improved on the results and on the RMS error. There is also the problem that we might sometimes improve the estimate of one surface patch, on one hand, but not get a good estimate of another, on the other hand, in this case and in the case described in Section 4.2.1.

4.3. Summary

In order to get a better idea of the performance of the binary filter case and the “moving camera” case, we compute an average RMS error of the model estimates of the illuminant and surface patch compared to the measured spectra. Ten scenes of three *matte* Munsell color patches [2] each were constructed. Munsell patch 400 was in all scenes while the others were chosen at random. The scenes were illuminated with one light source chosen at random from Parkkinen’s set of 37 daylight and skylight spectra [35]. Then the colors of the Munsell patch (400) and the light source were estimated in the ‘no filter’ case and compared to the binary filter case: static camera, camera moved once, and camera moved twice. Table 4.5 shows the average RMS errors obtained in each case.

Case	Illuminant	Surface Patch 400
No Filter	0.2386	0.1310
Binary Filter, No Motion	0.1392	0.1059
Binary Filter, 1 Move	0.0982	0.0682
Binary Filter, 2 Moves	0.0839	0.0552

TABLE 4.5. Average RMS error for Munsell Patch 400 and the illuminant spectra when there are multiple patches in the scene for the ‘no filter’ case and the binary filter case when the camera is: not moved, moved once, and moved twice.

CHAPTER 5

Experiments and Analysis

The sequential multiple sensor model introduced in this thesis (Chapter 3) is an extension to the Brainard-Freeman model [7] described in Section 2.2. In order to make these two models more plausible, which would in turn help in making our model clearer, some experiments were carried out on real data in order to test these hypotheses. These experiments were similar to the simulations, except for the changes that were made to the components of the models in order to make them compatible with the real world. The experiments are detailed in the following sections. In terms of the simulations described in Chapter 4, we simulate the ‘no filter’ case and the binary filter case, in particular, the *static camera* case.

5.1. The Experimental Setup

The main components of the experiment are the light source, the *glossy* Munsell color patches [1], the yellow filter, and the camera. The experiment was done in a dark room at night to ensure that the only illumination was that coming from our light source. The only exception to that was the computer monitor, which was turned away from the setup as much as possible.

The Munsell color patches were chosen at random from the Munsell Book of Color, *Glossy Finish Collection*, which contains 1600 patches [1]. These are described in detail in Appendix A.

The yellow filter was a Kodak Wratten gelatin filter, which was placed on the surface patches, in such a way as to cover half of the scene. The spectral sensitivities function of the filter in the visible wavelength range (435 nm - 705 nm), which was used in the experiments, was provided with the filter.

The camera with which the images were taken was a Panasonic camera, mounted on a tripod. It was “looking” down at the scene forming approximately a 90° angle with the platform on which the Munsell patches were placed.

The light source was a General Electric (GE) 40-Watts tungsten bulb screwed into a desk lamp. Since the color patches were glossy, they would only exhibit appreciable specular reflectance as the incident light angle with the surface normal to the platform, on which the surface patches were placed, approaches 90° [25]. However, that is unpractical because then the light source would be along the platform. Therefore, the light source was placed such that the incident light was at an angle of approximately 45° from the surface normal of the platform.

5.2. The Algorithm

Going back to Equation 2.3, the response recorded at a location x of an array of p sensors is given by:

$$\rho_k^x = \int E(\lambda) S^x(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, \dots, p, \quad (5.1)$$

where $E(\lambda)$ is the spectral power distribution of the ambient light in the scene, and $S^x(\lambda)$ is the surface spectral reflectance. In the simulations case, noise was added to

the *RGB* measurements in order to have a reasonable simulation of the real experiments. Therefore, in this case no noise was added. In the experiment, the sensor is the camera, as opposed to Chapter 4, where the sensor was the eye. Therefore, the sensor spectral sensitivity curves, $R_k(\lambda)$, are those of the camera. The manufacturer (Panasonic Company) provided us with the spectral response curves of the color filter. These responses were in *CMYG* (Cyan, Magenta, Yellow, Green) scale. Therefore, the coordinates of some points on each curve were obtained by hand and then the rest of the curves were interpolated using the 'spline' function in Matlab [10], to get the functions over the visible wavelength range and at 5nm intervals. Figure 5.1 shows the *CMY* (or *CMYG*) spectral responses of the camera.

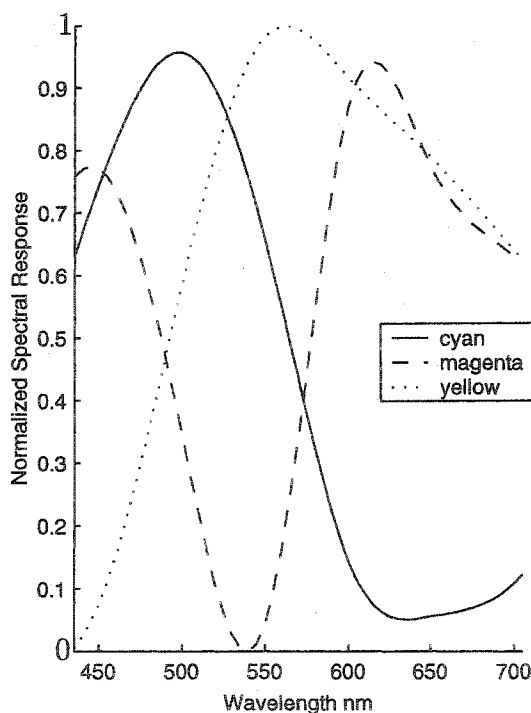


FIGURE 5.1. The camera color filter spectral responses in CMY scale.

Some cameras come with IR (InfraRed) cutoff filters. We did not know whether this was the case with our camera. Therefore, we determined the presence or absence

of the IR filter with the following simple experiment. In a dark room, the button of a television remote control was pressed in front of the camera. Since the IR beam coming out of the remote control did not show up in the image, it was confirmed that the camera has an IR cutoff filter. The IR cutoff filter curve was obtained from the camera manufacturer. In a way similar to the interpolation of the *CMY* curve, the coordinates of some points on the curve were obtained by hand, and then the whole curve was obtained by interpolation using the 'spline' function in Matlab [10]. Figure 5.2 shows the spectral response of the filter that was multiplied by the response obtained from Equation 5.1. It can be noticed from the figure that the curve goes to zero at very high wavelengths, that is towards the infrared color. This makes sensitivities at those wavelengths go to zero.

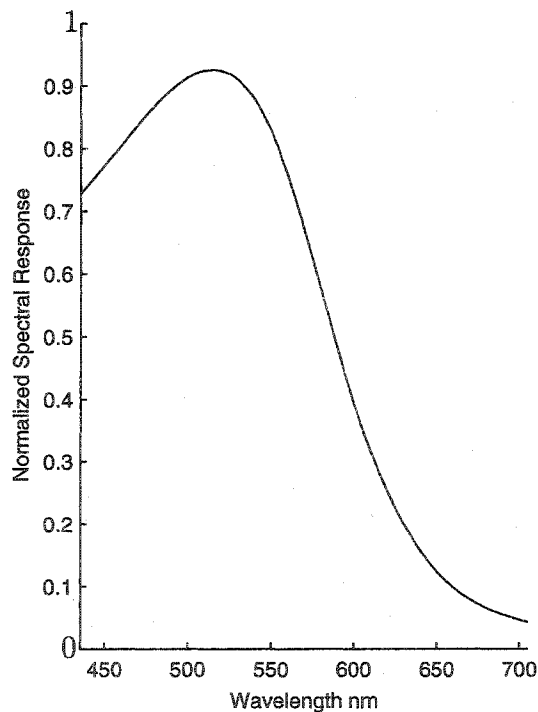


FIGURE 5.2. The IR cutoff filter.

The whole spectral sensitivities curve over the desired range of wavelengths for the yellow filter is obtained in a similar way to that used to obtain the *CMY* one. As for the light source, a regular tungsten halogen bulb was used. The spectrum for the illuminant was obtained from a GE set of Spectral Power Distribution (SPD) curves. These curves provide a visual profile of the characteristics of a light source, in other words, the measurement. The SPD curve used was the one corresponding to incandescent lamps at 2800K (Kelvin). It was obtained in a way similar to that used to get the camera color filter and IR filter responses and it is shown in Figure 5.3.

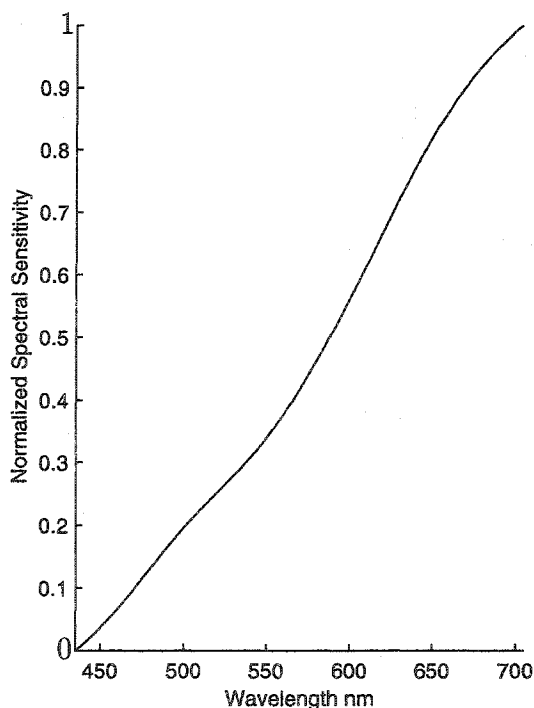


FIGURE 5.3. The tungsten light source measurement.

As for the Bayesian formulation, its components were computed in a similar method to that described in Section 4.1.

A few glossy Munsell color patches [1] were used in groups or individually. The Munsell patches were placed on a black platform, which would absorb all the light

and not reflect any off the surface. This would minimize the tendency of interference of light with the colors of the surface patches. Two scenarios were considered: one in which there was a single patch in the scene and one in which there were multiple patches in the scene. Both types of scenes were illuminated with our light source, which was described above.

5.3. Experimental Results

The correspondence between the labels on the patches and Orava's labels [33] for the measurements of the 1600 *glossy* Munsell color chips [1] is explained in Appendix A. The labels for the Munsell patches used in the experiments are also tabulated in the appendix as well as Orava's corresponding labels. Two scenarios are considered in this section: one in which there is a single patch in the scene and one in which there are three patches in the scene, illuminated with our light source. In each scenario, two cases were tested, which were also simulated in Chapter 4: the 'no filter' case, or equivalently, the Brainard and Freeman case and the binary filter case.

5.3.1. A Single Patch. The measurement and model spectra for surface patches for which images were taken individually are plotted in Figures 5.4, 5.5, 5.6, and 5.7 showing the results for each of the 'no filter' and the binary filter cases, and indicating the colors below each figure. The results show considerable improvement in the estimate of the surface color with the introduction of the filter into the scene. This can be seen by examining the individual as well as the average RMS errors for the surface patch estimates in both the no filter and the binary filter cases, which are given in Table 5.1.

As for the tungsten light source, the average RMS error of its estimate for the 'no filter' case is 0.1104, while that for the binary filter case is 0.2423. In other words, the light source estimate does not improve with the introduction of the filter. For

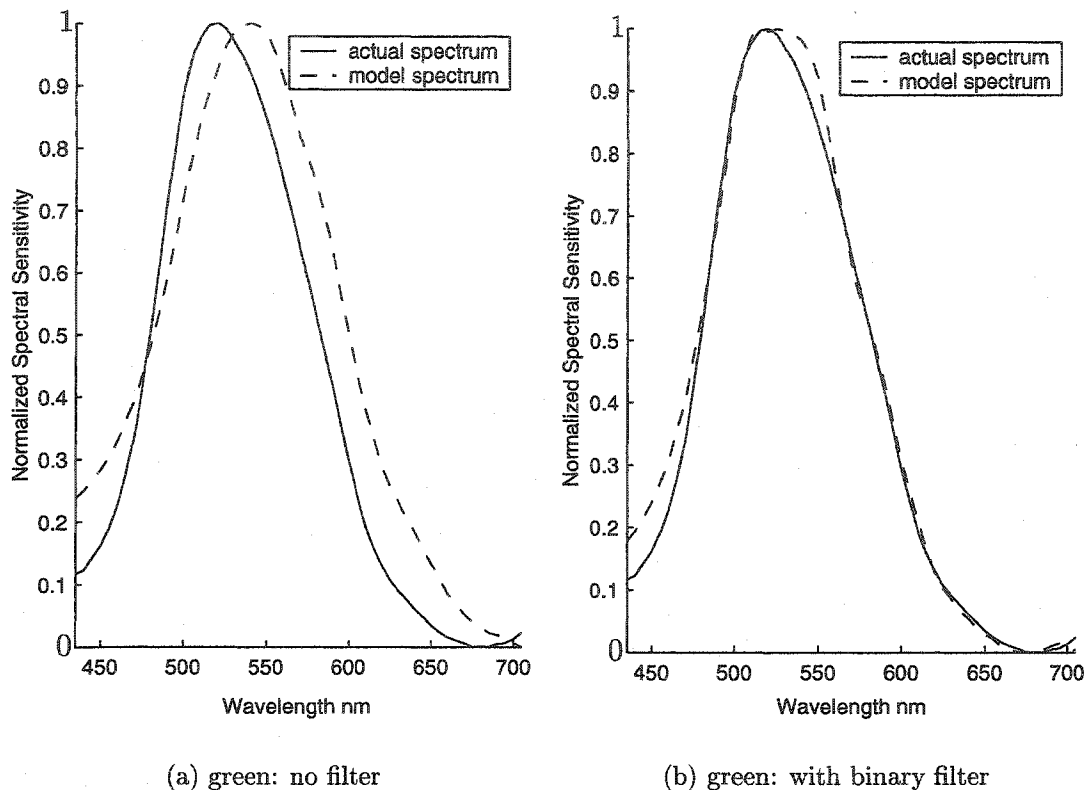


FIGURE 5.4. The estimated and actual spectra of a green Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.

Patch Color	No Filter	Binary Filter
Green	0.1279	0.0366
Red	0.2416	0.0886
Blue	0.1864	0.1024
Purple	0.2442	0.1372
AVERAGE	0.2000	0.0912

TABLE 5.1. RMS error for each of the Munsell surface patch spectra when there is a single patch in the scene for the 'no filter' and the binary filter cases.

each case, a model having an RMS error that is close to the average one is plotted in Figure 5.8.

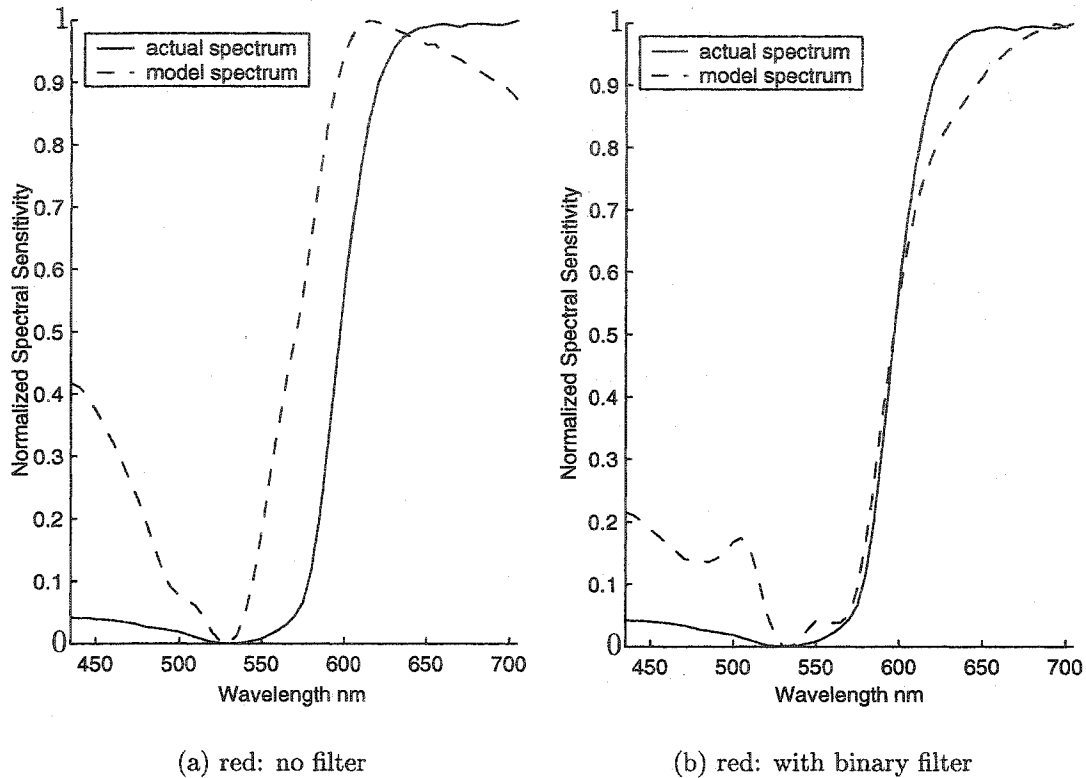


FIGURE 5.5. The estimated and actual spectra of a red Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.

5.3.2. Multiple Patches. In this section, the scenario for which there are three Munsell surface patches in the scene, illuminated with our tungsten light source, is examined. The measurement and model spectra for the surface patches are plotted and the RMS errors are depicted in tables as well.

We shall start with a scene which had a green, blue, and red patch. The green patch was partially covered with the filter, the blue patch was totally covered, and the red patch was not covered at all; that is, it was under the clear part. The results for this case are shown in Figure 5.9 with the color of the surface patch indicated below each figure. The corresponding RMS errors between the measurement and model spectra are given in Table 5.2.

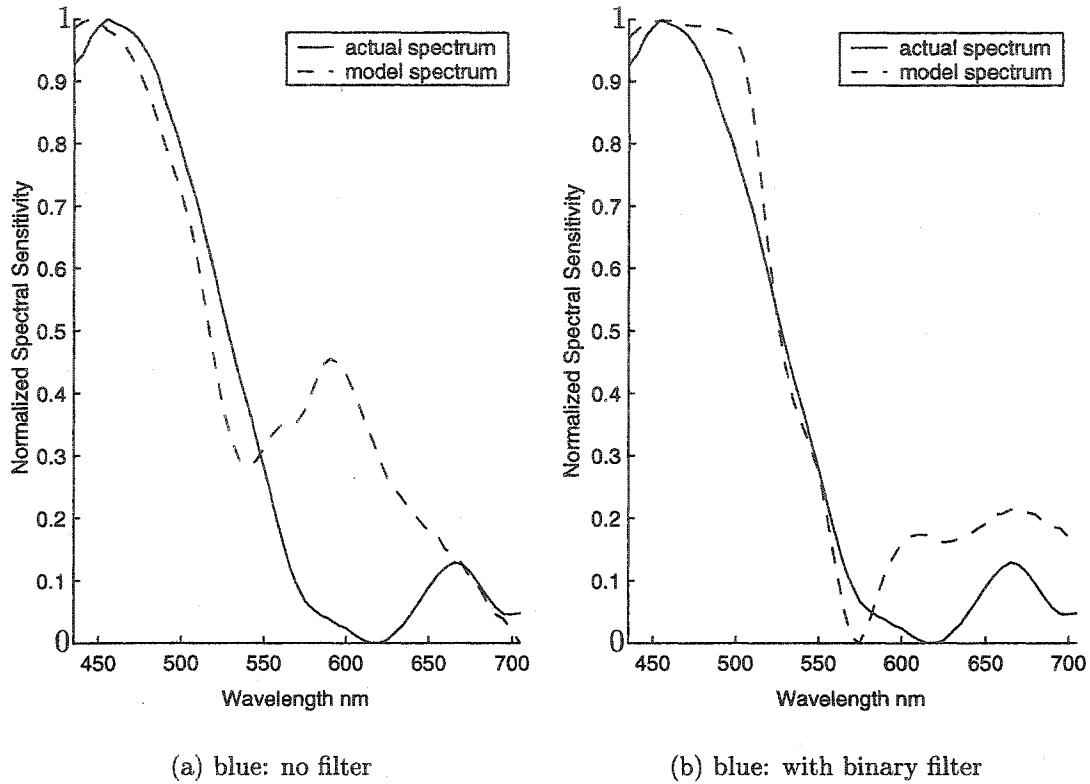
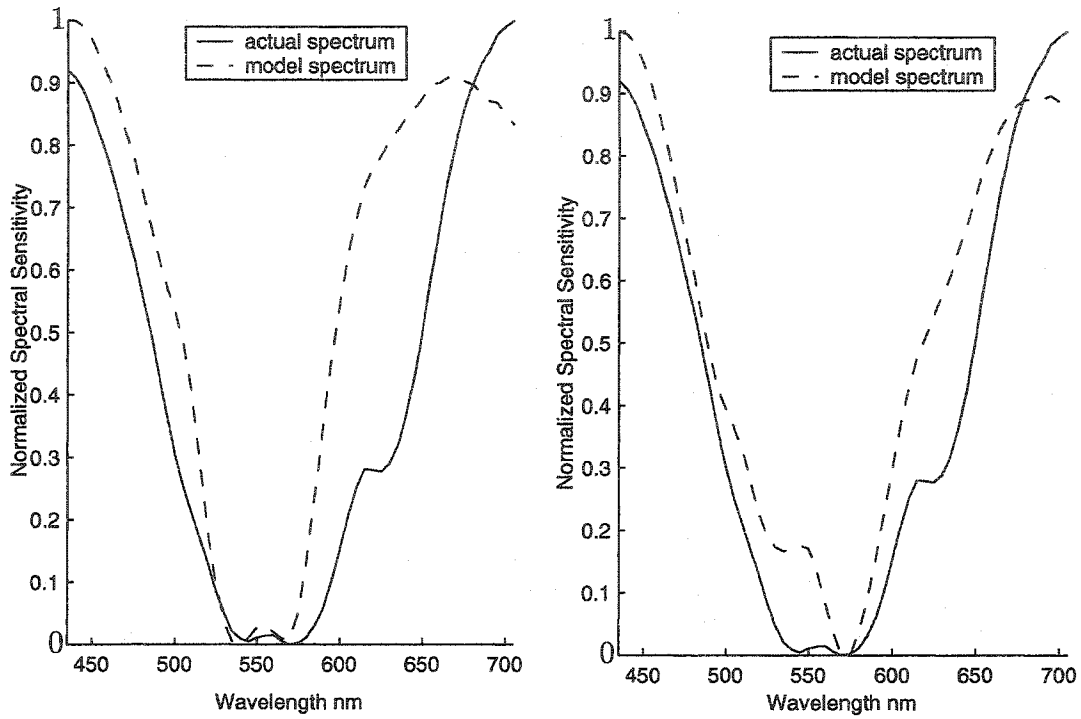


FIGURE 5.6. The estimated and actual spectra of a blue Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.

Patch Color	Green	Blue	Red
No Filter	0.1271	0.2242	0.1715
Binary Filter	0.0936	0.1735	0.1100

TABLE 5.2. RMS error for each of the Munsell surface patch spectra (green, blue, red) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.

Next, images of the purple, green, and yellow patches are taken together. The purple patch was partially covered with a filter, the green patch was totally covered, while the yellow patch was not covered at all; that is, it was under the clear part. The results for this case are shown in Figure 5.10 with the color of the surface patch indicated below each figure. The corresponding RMS errors between the measurement



(a) purple: no filter

(b) purple: with binary filter

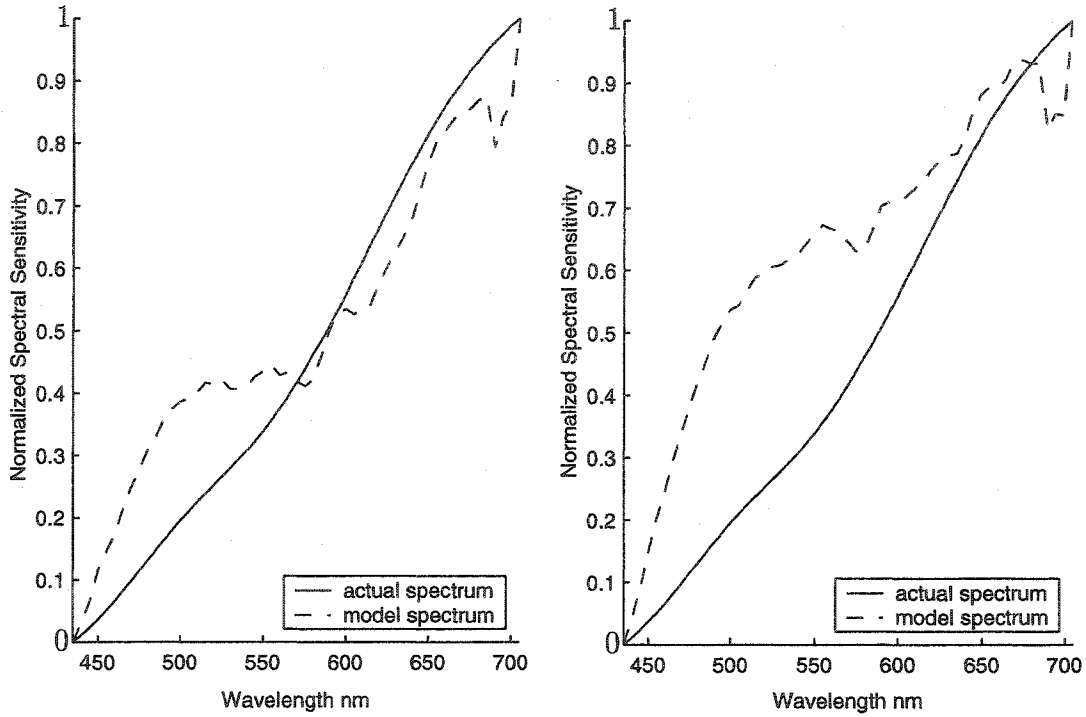
FIGURE 5.7. The estimated and actual spectra of a purple Munsell patch when there is a single patch in the scene in the cases of (a) 'no filter', (b) binary filter.

and model spectra are given in Table 5.3.

Patch Color	Purple	Green	Yellow
No Filter	0.2055	0.1263	0.1566
Binary Filter	0.3214	0.0916	0.1480

TABLE 5.3. RMS error for each of the Munsell surface patch spectra (purple, green, yellow) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.

Finally, images of the purple, blue, and red patches are taken together. The purple patch was partially covered with a filter, the blue patch was totally covered, while the red patch was not covered at all; that is, it was under the clear part. The results for



(a) illuminant: no filter

(b) illuminant: with binary filter

FIGURE 5.8. The estimated and actual spectra of the illuminant when there is a single surface patch in the scene in the cases of (a) 'no filter', (b) binary filter.

this case are shown in Figure 5.11 with the color of the surface patch indicated below each figure. The corresponding RMS errors between the measurement and model spectra are given in Table 5.4.

Patch Color	Purple	Blue	Red
No Filter	0.2127	0.2281	0.1609
Binary Filter	0.2917	0.1456	0.0863

TABLE 5.4. RMS error for each of the Munsell surface patch spectra (purple, blue, red) when there are multiple patches in the scene for the 'no filter' and the binary filter cases.

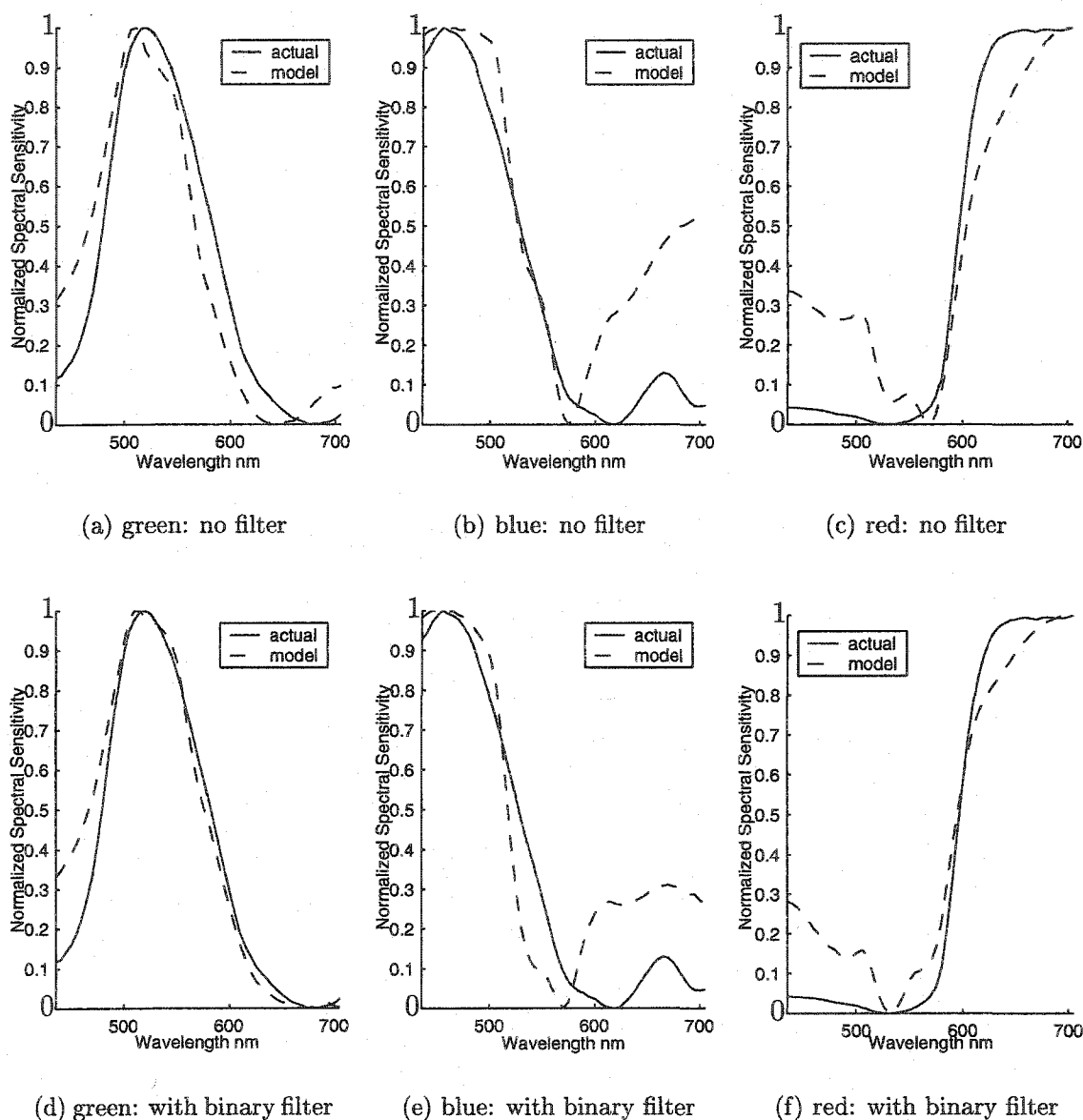


FIGURE 5.9. The estimated and actual spectra of the Munsell patches (green, blue, red) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.

Upon examining the RMS errors which are depicted in Tables 5.2, 5.3, and 5.4, we can notice considerable improvement in the surface patch spectra estimates with the introduction of the binary filter into the scene, with a few exceptions. As for the tungsten light source, the average RMS error between the model and the measurement

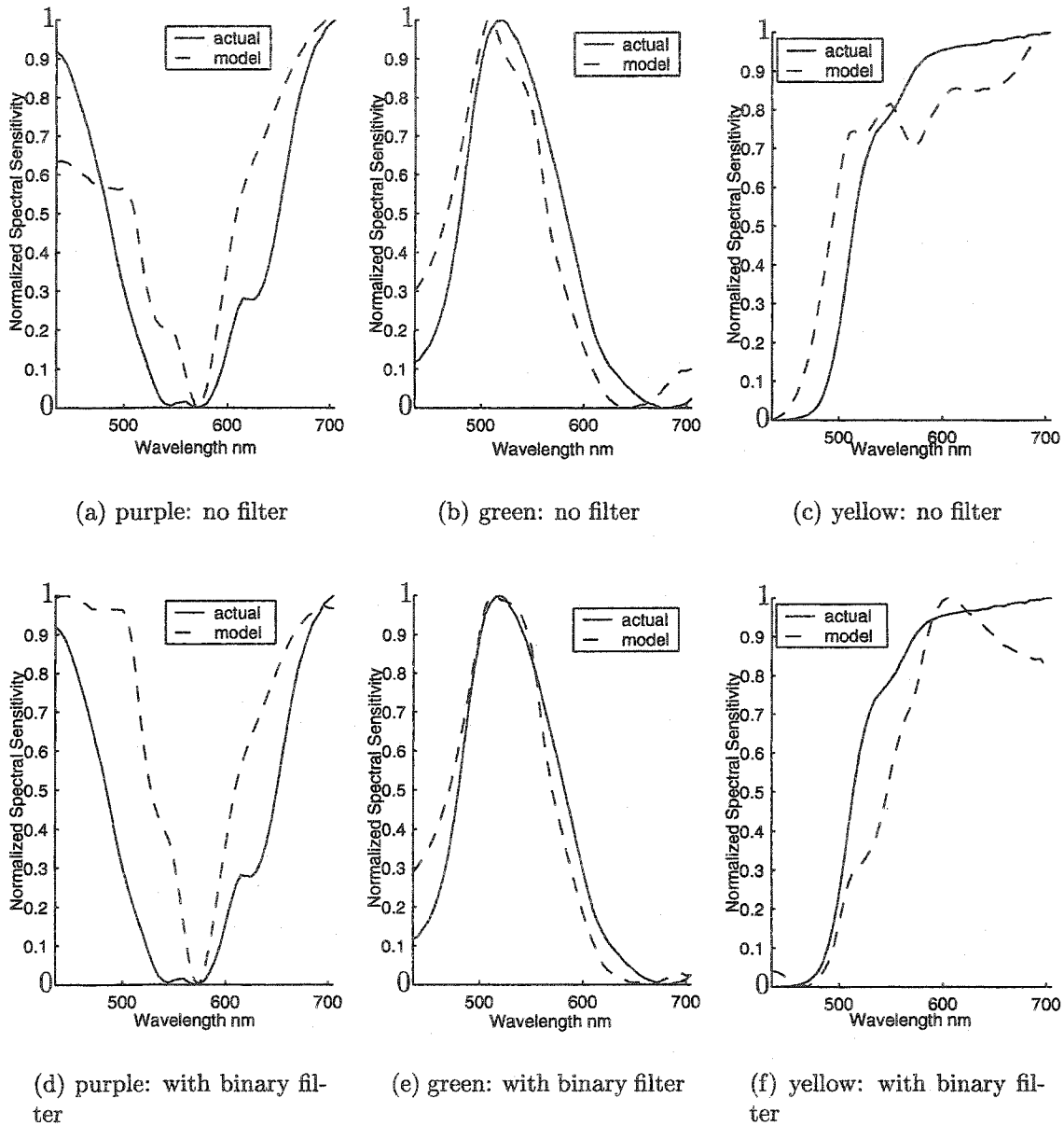


FIGURE 5.10. The estimated and actual spectra of the Munsell patches (purple, green, yellow) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.

for the 'no filter' case of all three combinations considered above is 0.1477. The estimated and actual spectra for the tungsten light source in this case are plotted in Figure 5.12 when there are three patches in the scene: purple, blue, and red. However,

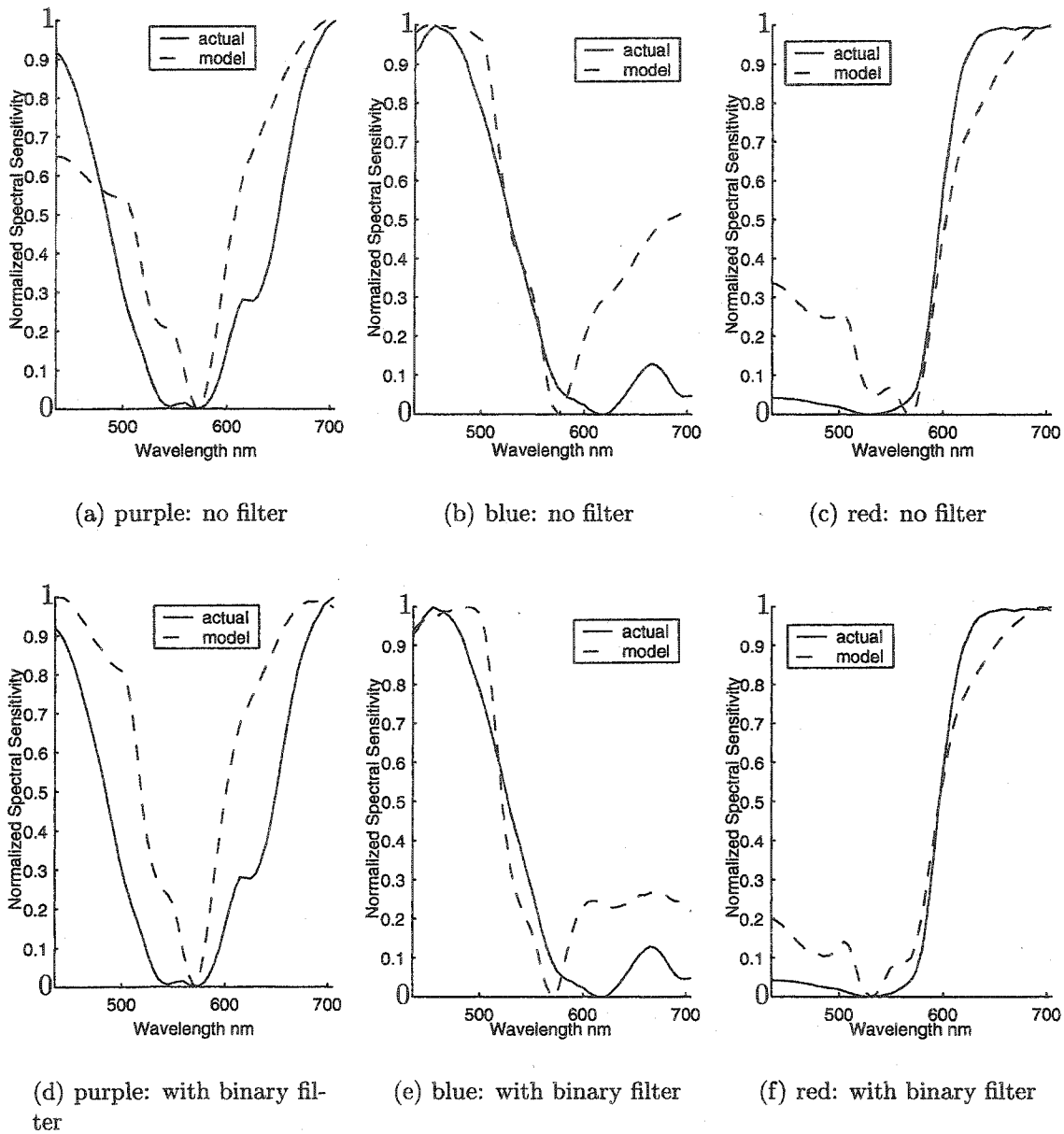


FIGURE 5.11. The estimated and actual spectra of the Munsell patches (purple, blue, red) when there are multiple patches in the scene in the cases of (a)-(c) 'no filter', (d)-(f) binary filter.

the introduction of the filter did not improve upon the estimate of the light source spectrum. A possible explanation for this is that the optimization routine might be getting trapped in a local minimum. Furthermore, the initial value might be one that

leads it to such a minimum. However, as mentioned previously, it is unpractical to try infinitely many starting points. Another source of error might be the fact that the set of basis functions used to approximate the tungsten light source spectrum were those of Parkkinen's set of daylight and skylight spectra [35] and not an indoors set of light spectra.

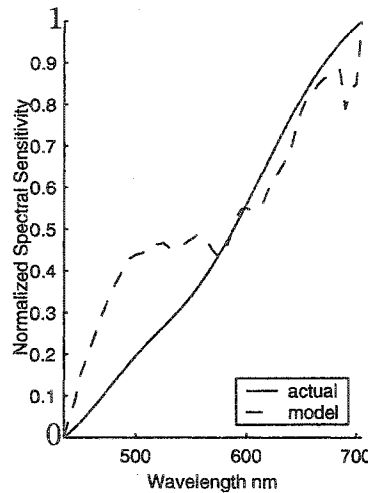


FIGURE 5.12. The estimated and actual spectra of the illuminant when there are multiple patches in the scene in the 'no filter' case.

In closing, we can say that the real experiments produce similar results to the simulations in the sense that the RMS errors are in the same range for the 'no filter' case and the binary case, in particular, the binary filter case. The introduction of the filter into the scene has improved the results considerably in both the single patch and multiple patch cases, especially for the surface patch spectra estimates. However, no improvement was observed for the light source estimates. Possible reasons were mentioned above. A few points are worth mentioning. First, trying random multiple initial values for the optimization did not help because the optimization algorithm converged at nearby points every time. There are many factors that influence the results in the real world such as noise (see Section 4.1), which is very difficult to control

or eliminate. Another experimental error which arises here, which does not occur in the simulations (Chapter 4), is obtaining the coordinates of the points of the camera response (the color filter, for example), which is done by hand, and then interpolating the whole function. This would give a function which is close to the real function but not the exact one itself. Moreover, there are interreflections between surfaces in the real world, and we have to expect surface patches to affect each others' colors. On other hand, the spectral sensitivities of the glossy Munsell patches were measured each independently, that is not taking into account interference of colors from other patches. This could be a source of difference between the measurement and the model spectra and would thus produce a larger RMS error. Finally, our model assumes that there is no interreflection between surfaces, and that the spectral reflectances of the surface and lighting vectors are sufficient statistics for the measurement of that surface, which is not the case in the real world.

CHAPTER 6

Conclusions and Future Work

Color constancy, which is the perception of a constant color across variations in illumination, has been the topic of research in this thesis. We have proposed an *active Bayesian* strategy that is an extension to the Brainard-Freeman Bayesian approach to color constancy [7]. Brainard and Freeman used a Bayesian technique to regularize [40] the problem of computing the values of the illuminant and surface reflectance parameters. This technique uses the bilinear model of Maloney and Wandell [30] to provide a parametrization of the illuminant and surface spectra.

Our model enhances Brainard and Freeman's model obtaining insight from the characteristics of human vision, and the human eye, in particular. The human retina is non-homogeneous, and the spatial distribution of the photoreceptors on the retina is non-uniform. Also the spectral absorption of the lens and macula varies across the retina. In our approach, measurements are acquired sequentially from each sensor of a set of independent sensors, similar to when a person moves his/her gaze across a surface in the scene. Two cases were modeled in this thesis: one in which there are only two different types of sensors (binary filter), and one in which there is a smooth variation in spectral sensitivity curves across the sensor array (Gaussian filter). The Gaussian filter case is an approximate model of the eccentric drop-off in the thickness

of the macular covering, as well as the eccentric drop-off in the path length of the light that passes through the lens. The binary case was extended to moving the camera across the scene. This is acquired from the notion of active vision that is a very important characteristic of human vision.

The simulations in Chapter 4 indicate that there is an improvement in the light source estimate with the introduction of more surface patches in the scene. Furthermore, there is considerable improvement with the introduction of the multi-sensor method compared to the ‘no filter’ method, which is that of Brainard and Freeman. In fact, the more sensors introduced, the better the estimates of the light source and surface patches, but the more the time it would take to get to the solution or for the optimization to converge. Another result shown in Chapter 4 is that of the “moving camera”: the more the number of moves of the camera, the better the estimates. In fact, if the camera were moved more than twice, which was one of the cases simulated in this thesis, we could have obtained better estimates of the light source and surface patches.

In Chapter 5, the experiment was done in our laboratory on *glossy* Munsell patches [1]. The results showed some improvement in the surface patch spectra estimates upon the introduction of the binary filter. Future work would involve moving the camera and introducing the Gaussian filter case. The binary filter case, particularly the *moving camera* case, can be modeled by placing the yellow filter on half of the scene and then moving it around such that different patches are under different parts of the filter each time and such that there is a new set of data gathered with every new movement. The Gaussian filter case can be modeled by placing a drop of yellow food coloring on the middle of the lens, which would be concentrated at the center and would fade away towards the peripheries.

In conclusion, Finlayson in [18] states that for scenes containing a small number of surfaces (1, 2, 3, or 4) many illuminants might be equally likely according to the Bayesian probability formulation [17]. This is what we observe in the light source estimate; that is as we introduce more surface patches into the scene, no improvement is observed. Furthermore, the results can be explained by the fact that some assumptions have been violated. The most important assumption is that we assumed that there is no interreflection between surfaces; and therefore, the measurements are assumed to be conditionally statistically independent (Chapter 3). However, this is not always true because we cannot say that surfaces do not reflect light onto other surfaces and affect each others' colors. Tominaga and Wandell state in [41] that the visual information we use to perceive the color of an object is derived from the light reflected from the object itself and the light reflected from nearby objects. This may be the reason behind getting unsatisfactory estimates for a surface patch in the scene when there are many surface patches. However, if we were to filter out the effect of the reflectances of the neighboring surfaces, we will not be taking into account a multiple surface patch scene anymore. Another challenging problem is finding a reasonable starting point for the optimization problem, which will not make it get trapped in local minima. These remain interesting questions for the future.

REFERENCES

- [1] *Munsell book of color-glossy finish collection*, Baltimore, MD: Munsell Color, 1976.
- [2] *Munsell book of color-matte finish collection*, Baltimore, MD: Munsell Color, 1976.
- [3] J. Aloimonos and A. Bandyopadhyay, *Active vision*, Proceedings of the IEEE International Conference on Computer Vision, 1987.
- [4] T. O. Berger, *Statistical decision theory and bayesian analysis*, New York, NY: Springer-Verlag, 1985.
- [5] M. Bertero, T. A. Poggio, and V. Torre, *Ill-posed problems in early vision*, Proceedings of the IEEE **76** (1998), no. 8, 869–889.
- [6] R. A. Bone, J. T. Landrum, and A. Cains, *Optical density spectra of the macular pigment in vivo and in vitro*, Vision Research **32** (1992), 105–110.
- [7] D. H. Brainard and W. T. Freeman, *Bayesian color constancy*, Journal of the Optical Society of America A **32** (1997), no. 7, 1393–1411.
- [8] J. J. Clark and J. K. O'Regan, *A temporal difference learning model for perceptual stability in color vision*, IEEE International Conference on Pattern Recognition (Barcelona, Spain), September 2000, pp. 503–506.

- [9] F. G. Cooper, *Munsell manual of color*, Baltimore, MD: Munsell Color, 1929.
- [10] C. de Boor, *A practical guide to splines*, Springer-Verlag, 1978.
- [11] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*, Stanford, CA: John Wiley & Sons, 1973.
- [12] M. D. D’Zmura, *Color constancy: Surface color from changing illumination*, Journal of the Optical Society of America A **9** (1992), no. 3, 490–493.
- [13] M. D. D’Zmura and G. Iverson, *Color constancy I. basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces*, Journal of the Optical Society of America A **10** (1993), no. 10, 2148–2165.
- [14] ———, *Color constancy II. results for two-stage linear recovery of spectral descriptions for lights and surfaces*, Journal of the Optical Society of America A **10** (1993), no. 10, 2166–2180.
- [15] ———, *Color constancy III. general linear recovery of spectral descriptions for lights and surfaces*, Journal of the Optical Society of America A **11** (1994), no. 9, 2389–2400.
- [16] M. D. D’Zmura and P. Lennie, *Mechanisms of color constancy*, Optical Society of America A **10** (1986), no. 3, 1662–1672.
- [17] G. Finlayson and S. Hordley, *Selection for gamut mapping colour constancy*, Image and Vision Computing **17** (1999), no. 8, 597–604.
- [18] G. D. Finlayson, *Color in perspective*, IEEE transactions on Pattern Analysis and Machine Intelligence (1996), 1034–1038.
- [19] G. D. Finlayson, M. S. Drew, and B. V. Funt, *Diagonal transforms suffice for color constancy*, IEEE International Conference on Computer Vision (Berlin, Germany), IEEE Computer Society Press, May 1993, pp. 164–171.

REFERENCES

- [20] G. D. Finlayson, B. V. Funt, and K. Barnard, *Color constancy under varying illumination*, Proceedings of the International Conference on Computer Vision (Cambridge, Massachusetts), June 1995, pp. 720–725.
- [21] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, *Color by correlation: a simple unifying framework for color constancy*, IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001), no. 11, 1209–1221.
- [22] G. H. Finlayson, *Colour object recognition*, Master's thesis, Simon Fraser University, Burnaby, British Columbia, April 1992.
- [23] B. Funt, V. Cardei, and K. Barnard, *Learning color constancy*, Proceedings of the IS&T/SID Fourth Color Imaging Conference (Scottsdale, Arizona), November 1996, pp. 58–60.
- [24] J. Hadamard, *Lectures on the cauchy problem in linear partial differential equations*, 1923.
- [25] D. Hearn and M. P. Baker, *Computer graphics*, ch. 14, pp. 494–563, Englewood Cliffs, NJ: Prentice Hall, Inc., 1994.
- [26] H. Helson, *Fundamental problems in color vision. I. the principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination*, Journal of Experimental Psychology **23** (1938), no. 5, 439–476.
- [27] H. E. Ives, *The relation between the color of the illuminant and the color of the illuminated object*, Transactions of the Illuminating Engineering Society (1912b), 62–72.
- [28] D. B. Judd, *Hue saturation and lightness of surface colors with chromatic illumination*, Journal of the Optical Society of America **3** (1940), 2–32.
- [29] E. H. Land, *The retinex theory of color constancy*, Scientific American (1977), 108–129.

REFERENCES

- [30] L. T. Maloney and B. A. Wandell, *Color constancy: A method for recovering surface spectral reflectance*, Journal of the Optical Society of America A **3** (1986), no. 1, 29–33.
- [31] J. J. Moré, *Numerical analysis*, 630 ed., Lecture Notes in Mathematics, ch. The Levenberg-Marquardt Algorithm: Implementation and Theory, pp. 105–116, New York, NY: Springer-Verlag, 1977.
- [32] E. Oja, *Subspace methods of pattern recognition*, Letchworth, England: Research Studies Press, 1983.
- [33] J. Orava, *The reflectance spectra of 1600 glossy munsell color chips*, http://www.cs.joensuu.fi/~spectral/databases/download/munsell_spec_glossy_all.htm.
- [34] A. Papoulis, *Probability, random variables, and stochastic processes*, New York, NY: McGraw-Hill, 1984.
- [35] J. P. Parkkinen and P. Silfsten, *Database of spectra*, University of Joensuu, Finland, <http://cs.joensuu.fi/~spectral>.
- [36] J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen, *Characteristic spectra of munsell colors*, Journal of the Optical Society of America A **6** (1989), no. 2, 318–322.
- [37] G. Sapiro, *Bilinear voting*, International Conference on Computer Vision, 1998, pp. 178–183.
- [38] S. Skaff, T. Arbel, and J. J. Clark, *Active bayesian color constancy with non-uniform sensors*, IEEE International Conference on Pattern Recognition (Quebec City, QC), vol. 2, August 2002, pp. 681–684.
- [39] A. Stockman, L. T. Sharpe, and C. C. Fach, *The spectral sensitivity of the human short-wavelength cones*, Vision Research **39** (1999), 2901–2927.

REFERENCES

- [40] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, Washington, D.C.: V. H. Winston, 1977.
- [41] S. Tominaga and B. A. Wandell, *Standard surface-reflectance model and illuminant estimation*, Optical Society of America A **6** (1989), no. 4, 576–584.

APPENDIX A

Munsell Patches

The purpose of this appendix is to introduce the Munsell patches and their specifications. The first section of this appendix describes the notation system of the Munsell color patches and the way they are grouped in the Munsell Book of Color. There are two kinds of color patches: *matte* and *glossy*. The matte ones were used in the simulations (Chapter 4) and the glossy ones were used in the experiments (Chapter 5). The second section of this appendix describes the correspondence between the labels of the 1600 glossy Munsell patches themselves and the labels given by Orava who measured their spectra.

A.1. The Notation System

The system of colors, developed by Munsell and used in the Munsell Book of Color [1], identifies color in terms of three attributes: hue, value, and chroma [9].

The hue (H) is the first characteristic of a color that the eye detects and it is scaled from one to ten. There are ten major hues, of which five are principal and five are intermediate. These hues are positioned ten steps apart within this scale. The names of the major ones are: Red (R), Yellow-Red (YR), Yellow (Y), Green-Yellow

(GY), Green (G), Blue-Green (BG), Blue (B), Purple-Blue (PB), Purple (P), and Red-Purple (RP).

The value (V) notation indicates how *light* or how *dark* a color is in relation to a natural gray scale. The value scale extends from absolute black, which is symbolized by zero, to absolute white, which is symbolized by ten. For example, yellow is usually a light color, so it is closer to white than to black. Figure A.1 is a simple diagram of the value scale.

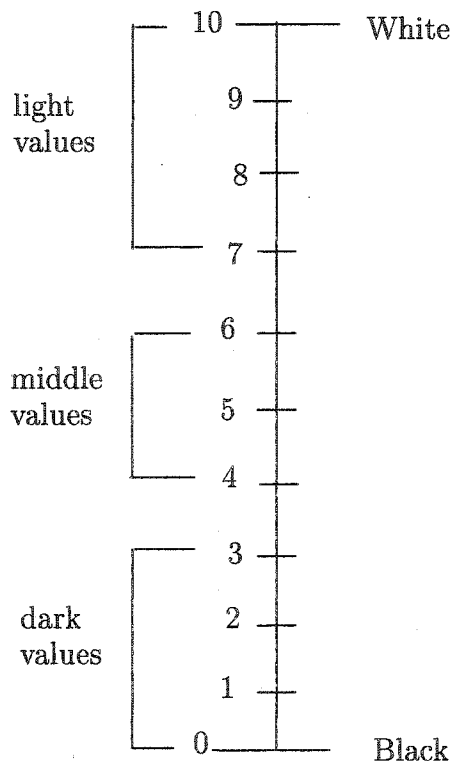


FIGURE A.1. The value scale.

The chroma (C) notation is a measure of the degree of color strength or intensity. It indicates how much of a green hue is different from a neutral gray of the same value. That is, two colors may be the same in hue (for example, both green) and the same in value (that is neither is darker or lighter than the other), and yet be different

A.2 CORRESPONDENCE: MUNSELL LABEL TO ORAVA'S LABEL

in color strength. One may be strong green and the other weak, grayish green. This difference is in the dimension of the chroma. Chromas close to neutral are known as *weak* chromas, those close to maximum strength are known as *strong* chromas, and those between the weak and the strong are known as *moderate* chromas. The scale of chromas extends from 0 to 10, 12, 14 or farther. In the case of the glossy Munsell patches used in the experiments in this thesis (Chapter 5), the scale is 0 to 16.

To sum up, hue (H) is the name of a color, value (V) is the amount of light in a color, and chroma (C) is the degree of strength in a color. Therefore, a Munsell patch or color chip has a label of the following form: HV/C .

A.2. Correspondence: Munsell Label to Orava's Label

The spectral sensitivities of the 1600 glossy Munsell color patches were obtained from Joni Orava's website [33]. He measured the reflectance spectra of the color chips with a Perkin-Elmer Lambda 18 UV/VI spectrophotometer over a wavelength interval of 380 nm to 780 nm and a resolution of 1 nm. The correspondence between the Munsell patch labels (explained in Section A.1) and Orava's labels is explained below.

Since it is easier to understand the correspondence by example, we shall explain it with the following one. The label on a red Munsell patch (used in the experiments in Chapter 5) from the Munsell book of color is $5R\ 5/12$. The first three letters in his labels represent the hue. Since $B = 5$ and $RR = Red$ (it is R in the label), the first three letters in Orava's representation would be BRR . The value is represented with the fourth and fifth letter. Since 5 corresponds to 50, then the label, putting the hue and value together, would become $BRR50$. Finally, the chroma is represented with the sixth and seventh letter. Since 12 corresponds to 12, putting all the components of the label together, the label of the red Munsell patch having a label $5R\ 5/12$

A.2 CORRESPONDENCE: MUNSELL LABEL TO ORAVA'S LABEL

would be represented in Orava's representation by *BRR5012*. Having the label of the patch, the number of the patch could be obtained from Orava's list and hence its measurement from the Matlab file provided on his website [33]. Below are Orava's lists of the corresponding indexing between his labels and the Munsell patches labels. The values to the left of the "=" operator are his given indices of the labels and the values to the right of the "=" operator are the Munsell patches indices of the labels.

MUNSELL HUE (First letter)

$$A = 2.5$$

$$B = 5$$

$$C = 7.5$$

$$D = 10$$

for supplementary colors:

$$E = 1.25$$

$$F = 3.75$$

$$G = 6.25$$

$$H = 8.75$$

Second and third letter:

$$RR = Red$$

$$YR = Yellow - Red$$

$$YY = Yellow$$

A.2 CORRESPONDENCE: MUNSELL LABEL TO ORAVA'S LABEL

GY = Green–Yellow

GG = Green

BG = Blue–Green

BB = Blue

PB = Purple–Blue

PP = Purple

RP = Red–Purple

MUNSELL VALUE (Fourth and Fifth letter)

10 = 1

20 = 2

30 = 3

40 = 4

50 = 5

60 = 6

70 = 7

80 = 8

85 = 8.5

90 = 9

A.2 CORRESPONDENCE: MUNSELL LABEL TO ORAVA'S LABEL

MUNSELL CHROMA (Sixth and Seventh letter)

$$01 = 1$$

$$02 = 2$$

$$04 = 4$$

$$06 = 6$$

$$08 = 8$$

$$10 = 10$$

$$12 = 12$$

$$14 = 14$$

$$16 = 16$$

After explaining Orava's labelling system and giving his lists of corresponding indexing between his labels and the Munsell patch labels, we list the colors of the glossy Munsell patches used in our experiments, their labels, and Orava's corresponding labels in Table A.1.

A.2 CORRESPONDENCE: MUNSELL LABEL TO ORAVA'S LABEL

Patch Color	Munsell Label	Orava's Label
Green	2.5G 6/8	AGG6008
Red	5R 5/12	BRR5012
Purple	2.5P 5/8	APB6008
Purple-Blue	2.5PB 6/8	APP5008
Yellow	5Y 8/10	BYY8010

TABLE A.1. Table of correspondence between the Munsell patch labels and Orava's labels for the glossy Munsell patches used in our experiments.