COEFFICIENT COLOR CONSTANCY

by

Graham David Finlayson

B.Sc University of Strathclyde 1989

M.Sc Simon Fraser University 1992

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
April 1995

APPROVAL

Name: Graham David Finlayson

Degree: Doctor of Philosophy

Title of thesis: Coefficient Color Constancy

Examining Committee:

Dr. Rick Hobson
Chair

Dr Brian V. Funt, Senior Supervisor

Dr Ze-Nian Li, Supervisor

Dr Jacques Vaisey, Supervisor

Dr David Fracchia, SFU examiner

Dr David A. Forsyth, External Examiner

Date Approved: _____

Abstract

The goal of color constancy is to take the color responses (for example camera *rgb* triplets) of surfaces viewed under an unknown illuminant and map them to illuminant independent descriptors. In existing theories this mapping is either a general linear 3×3 matrix or a simple diagonal matrix of scaling coefficients. The *general* theories have the advantage that the illuminant can be accurately discounted but have the disadvantage that nine parameters must be recovered. Conversely while the *coefficient* theories have only three unknowns, a diagonal matrix may only partially discount the illuminant.

My staring point in this thesis is to generalize the coefficient approach; the goal is to retain its inherent simplicity while at the same time increasing its expressive power. Under the generalized coefficient scheme, I propose that a visual system transforms responses to a new sensor basis before applying the scaling coefficients. I present methods for choosing the best coefficient basis for a variety of statistical models of color responses. These models are rich enough that the generalized coefficient approach suffices for almost all possible sensor sets.

To achieve color constancy the correct coefficients must be recovered. Existing algorithms can do so only when strong constraints are satisfied. For example it is often assumed that there is a white reflectance in every scene. In the second part of my thesis, I develop a new coefficient algorithm, which I call *color in perspective*, based on very weak (and very reasonable) assumptions about the world. I assume only that the range of color responses induced by different reflectances varies with a change in illumination and that illumination itself can vary only within certain bounds. I tested the algorithm on real images taken with a color video camera—extremely good

constancy is delivered. Indeed the degree of constancy compares favorably with the best which is theoretically possible.

The methods developed in this thesis can be applied to a variety of other areas including color graphics, color reproduction and color appearance models.

Acknowledgements

My first and foremost thanks go to my thesis supervisor Dr. Brian V. Funt. He has been exceptionally supportive, directing me to relevant literature in the beginning, providing constructive criticism at every stage and tutoring me in the art of research throughout.

On an academic front I have many other people to thank. I am particularly grateful to Dr. Mark Drew with whom I have collaborated on many projects. Dr. Drew has enviable mathematical acumen and he patiently guided me through many mathematical results. Thanks also go to Dr. Binay Bhattacharya for introducing me to computational geometry (a key component of my thesis), Dr. Michael H. Brill for sharing his encyclopedic knowledge of color science with me and to Drs. Ze-Nian Li and Jacques Vaisey for their thesis supervision.

For continued trips for tea and coffee: Sanjeev Mahajan (he started it all off), Sumeet Bawa, Pinaki Mitra, Sheelagh Carpendale and Subho Chatterjee have provided excellent company. On the football (or soccer) pitch thanks go to my fellow *Pandemonium* team-mates; these include Jacques Vaisey, Petr Kubon, Jean Varaldi, Warren Moors and Tamas "the chief" Erdelyi. For help and support in the lab: Brigitte Dorner, Kobus Barnard and Janet Dueck have been first class. Partners on numerous and various extra curricular activities include T. Pattabhiraman, Glenn Macdonald and Stephan Wehner. For asking to be acknowledged thanks go to Tammy Laberge. I am especially grateful (again) to three good friends who I have known since arriving in Canada. These comprise Mark Mezofenyi, Allan Bennet-Brown and Gilles Dionne.

A final and heartfelt thanks go to my family all those miles away in Scotland.

Foreword: Thesis Organization

Common practice dictates that a thesis should be organized linearly: a smooth progression from introduction, to background, theoretical results, experiments and conclusion. This thesis is **not** organized in this manner. As such, I feel compelled to comment on the layout of my thesis.

Like many bodies of work my thesis encompasses not a single idea but many. While these are all related they are each individually interesting and important. However if I followed common practice I would strip away individual identity in creating the linear whole. Specifically each section would cover all ideas at the same time. The background section would be the union of background material for the many ideas, the theoretical results the union of derivations, and the experimental results the union of experiments.

In this thesis I take a more *lateral* approach and present each idea by itself. Specifically (excepting the introduction and conclusion) each chapter tackles a single major idea. Only the background, theoretical derivations and experiments relevant to the idea at hand are presented. As such each idea is placed in the context in which it was investigated and each chapter is completely self contained. Moreover, because later chapters build on earlier ones the thesis is still linear though in a different (and I would suggest more appropriate) sense.

Of course one can argue the merits for either organizational strategy. For example in the linear model definitions are presented singularly; here they are necessarily duplicated if they are germane to more than one idea. Thus the linear thesis is more compact. However I would contend that the small degree of repetition in lateral theses is worthwhile and constitutes reinforcement of ideas; in essence making things clearer.

Moreover background material for the subject at hand can always be found locally; there is no need for a tiresome search through earlier chapters.

I have to confess that the lateral oragnizational strategy was not my own idea. Rather I was fully prepared to adhere to the strictures of the linear approach until I read the thesis of Lucassen[Luc93]. In reading that thesis I was impressed at the clarity of exposition and this, I feel, was in part due to the separate presentation of ideas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Color constancy is an important problem for any vision system requiring stable perception of the world. Common illuminants such as daylight and tungsten light, or even just clear sky versus cloudy sky, differ markedly in their spectral properties; and since the spectrum of the light impinging on an eye or camera is the product of the incident illumination and the percent spectral reflectance of the surface, the illumination must be accounted for and discounted if there is to be stable perception of the surface color. Despite extensive research into color constancy in both the machine vision community and elsewhere [For90, FDH91, FD88, D'Z92, DL86, MW86, Lan77], there still does not exist a color constancy algorithm that performs sufficiently well that it either matches human color perception or provides a robot with adequate color recognition.

A pictorial representation of the color constancy problem is shown in Figure 1.1; the input consists of an image of a scene viewed under unknown lighting conditions and the output the image of the same scene viewed relative to a known canonical illuminant. I call the input illumination-dependent color responses *color observations* and the illumination-independent output *color descriptors*. The goal of color constancy is to map observations to descriptors.

The color constancy problem can be split into two related parts: the image model and the recovery algorithm. The image model describes the interdependence between illumination, reflectance and the visual system's (camera or human observer) response.

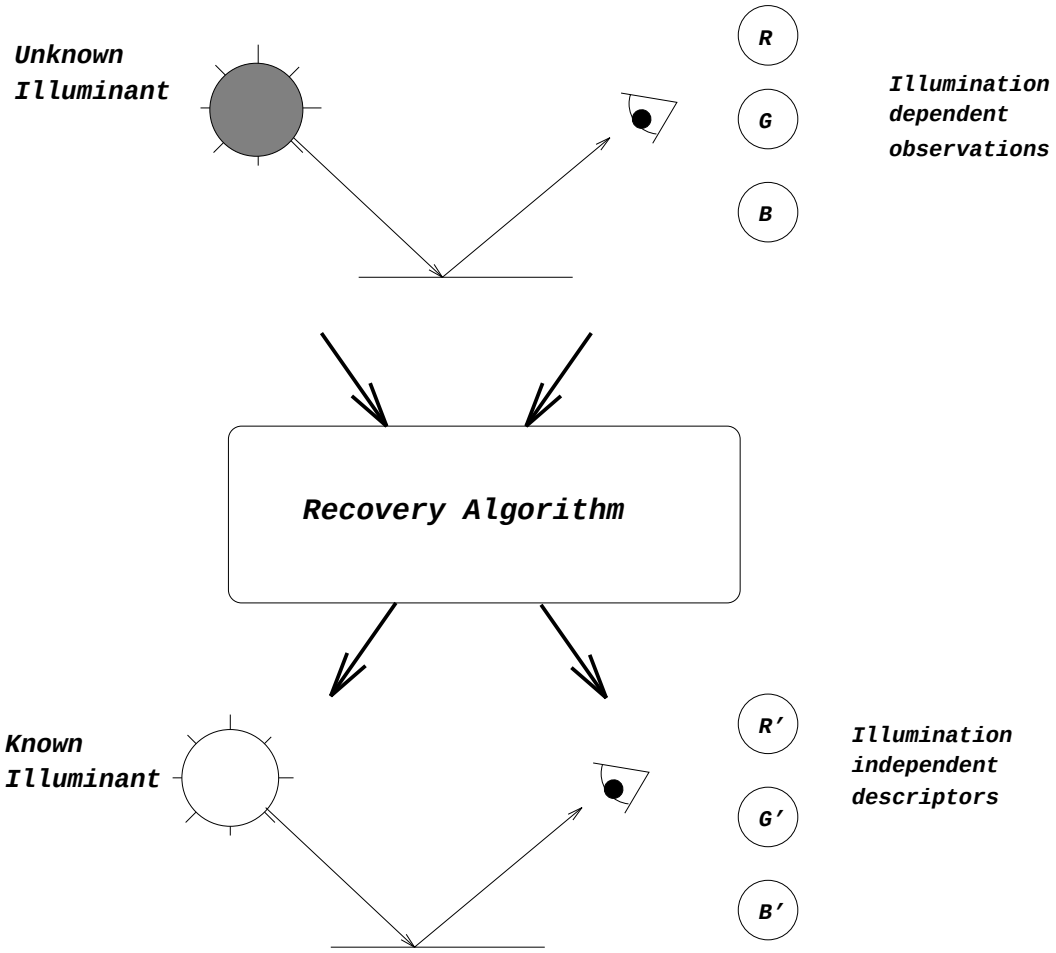

Figure 1.1: The Color Constancy Problem: the human eye (or color camera) views a scene under an unknown illuminant. A recovery algorithm transforms the scene to a known illuminant, solving the color constancy problem.

The recovery algorithm attempts to extract reflectance information from an image. It is only with a good image model that it is possible to recover the mapping taking observations to descriptors.

Surprisingly, most existing theories of color constancy have been developed with scant regard to the question of image model. The conventional approach has been to adopt some model (sometimes without comment) and focus only on the recovery problem. If an unrealistically simple model is employed then only an approximate solution, if any, to the color constancy problem will be possible. Conversely, an over-expressive image model necessitates extremely complex recovery. Moreover there may no longer be a unique solution to the color constancy problem.

In this thesis the question of *which image model?* is placed at centre stage. Sketches of the three basic linear candidate image models—*trivial*, *coefficient* and *general* are shown in Figure 1.2. In the trivial model observations are mapped to descriptors by a simple global intensity scaling. A separate scaling factor is applied to each sensor channel in the coefficient model. The general model is more complex still, each descriptor value is a weighted sum of the three observation responses (in Figure 1.2 the three inputs to the descriptor nodes, R', G' and B' are added together). Given the three cone response functions of the human eye, only the general model is sufficient to map accurately color observations to descriptors. In contrast if a visual system's sensors are narrow-band (they are sensitive to a single wavelength) then the coefficient model is all that is required. Only under extremely specialized (and quite unrealistic) circumstances is the trivial model appropriate [BWC89].

It is well known [WB82, Lan83, Hur89, Hur86] that the suitability of the coefficient model is dependent on the sensor basis employed; i.e. some linear combinations of the R, G, and B color channels are more amenable to a coefficient model than others. However, to date the question of finding the best sensor basis has received little attention save the purely speculative approach of Hurlbert [Hur89] and the heuristic approach of West and Brill [BW82]. In this thesis I develop several methods for finding an appropriate basis for coefficient color constancy. I will show that the coefficient model of image formation is almost always sufficient so long as a suitable sensor basis is employed. This *generalized coefficient* model is sketched in Figure 1.3.

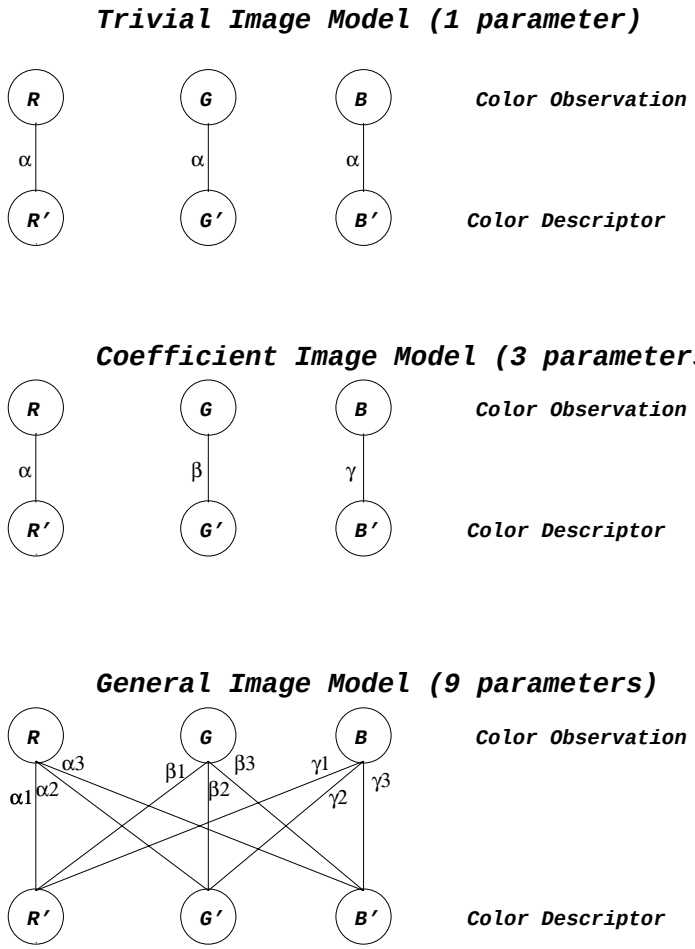

Figure 1.2: Comparison of Image Models. A single scaling coefficient describes illumination change under the trivial model. Three scaling coefficients are required, one per color channel, for the coefficient model. Nine parameters needed for the general model; each color channel output is a weighted sum of the three input color channels.

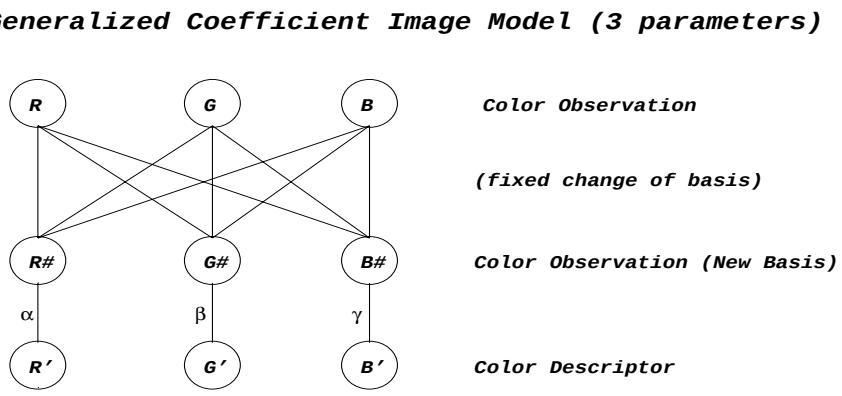

Figure 1.3: There are two steps in the generalized coefficient model: first there is an illumination independent change of sensor basis then illumination dependent coefficient scaling.

In recovering reflectance information from an image, I distinguish between two classes of algorithm: constrained and unconstrained. An unconstrained algorithm sets out to recover reflectance information with no a priori knowledge about the world save the model of image formation. However, because the image model provides only a weak constraint, these algorithms may not be able to provide a unique answer to the color constancy problem. In contrast, constrained algorithms assume that some property holds in the world (e.g. that there is a white reflectance in every scene) and this generally leads to a single solution. The relationship between image model and recovery algorithm is sketched in Figure 1.4.

Arguably unconstrained algorithms are superior in that they are applicable to any image, not only those where some world assumption is satisfied. Moreover, should additional world information be available then this should be used in conjunction with the output from an unconstrained algorithm. The most plausible, and powerful, unconstrained algorithm to date has been developed by Forsyth [For90]. By incorporating the physical realizeability of surface reflectance into a coefficient image model, Forsyth demonstrated that the color constancy problem was highly constrained; especially if there are many reflectances in a given scene. Unfortunately, his algorithm has two major failings: first, the coefficient image model does not in general apply

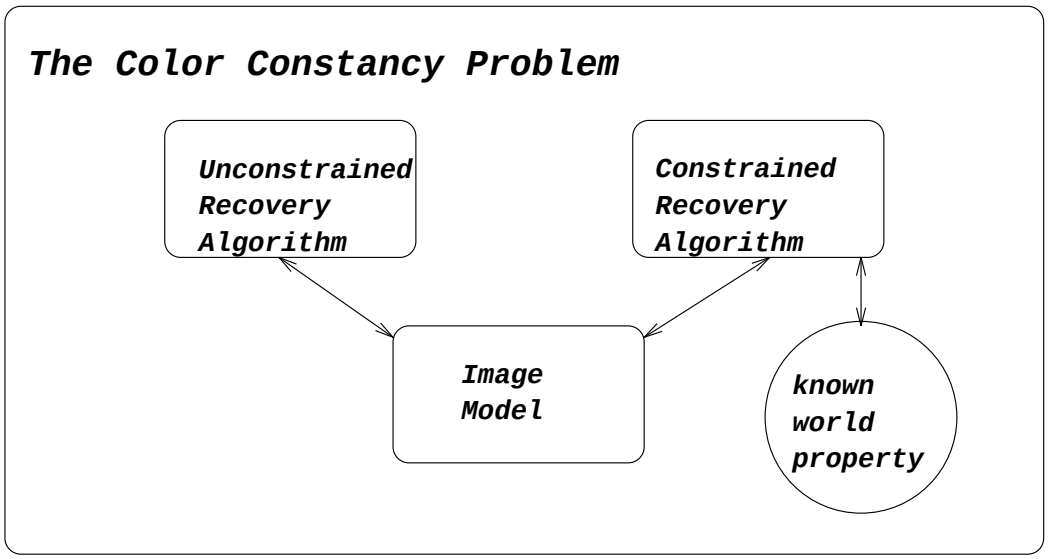

Figure 1.4: There are two classes of recovery algorithm for Color Constancy. The Unconstrained Algorithms proceed only with knowledge of the image model. Constrained algorithms use both the image model and some other known property of the world.

and second the algorithm is quite complex (it is tricky to implement and has a fairly high computational complexity).

In this thesis I extend Forsyth's approach in three important ways: first I allow a change in sensor basis to make the coefficient model appropriate; second I place a physical realizeability constraint on the illuminant further restricting the solution set; and lastly I demonstrate that this new algorithm can be implemented efficiently. Tests on real images show good color constancy is possible even when the number of distinct colored surfaces is small. Indeed such good constancy is possible that additional world assumptions may not be required.

Recent research into color constancy [D'Z92, DI93] has modified the general problem (Figure 1.1). Instead of a single view of the scene under one illuminant the input to the color constancy problem consists of a scene viewed under many illuminants. Ignoring the question of the plausibility of this circumstance (in general it seems quite implausible) I show that if there exists a sensor basis where a coefficient model suffices there can be no benefit in observing the world under many illuminants. The only

exception to this is if the illuminants which occur are quite restricted. In this case however, I show that the generalized coefficient model is always applicable.

The rest of this chapter is devoted to building the foundation of the thesis. I begin by formalizing the notion of an image model and this leads to a formal statement of the color constancy problem. How well a particular recovery algorithm can solve the color constancy problem is intimately related to how well the image model accounts for color observations. This notion is formalized in a discussion of image performance: a recovery algorithm's performance is bounded by that of its underlying image model. I go on to give a brief review of existing solution methods indicating their strengths and weaknesses. The chapter concludes with the outline for the rest of this dissertation.

1.1 Color Image Formation

The interaction of light and reflectance can be extremely complex. Factors such as interreflection and specularities are often difficult to model and deal with. To simplify matters I will initially consider color image formation for the simplified Mondriaan world. This world consists of a planar surface composed of several, overlapping, matte (Lambertian) patches and a single everywhere uniform illumination. In the Mondriaan world the only factor confounding the retrieval of descriptors is the illumination. Later I relax the model and will allow shape, specularities and varying illuminant power.

A Lambertian reflectance $S(\lambda)$ (λ denotes wavelength) illuminated by a spectral power distribution $E(\lambda)$ reflects the color signal $C(\lambda)$:

$$C(\lambda) = E(\lambda)S(\lambda) \tag{1.1}$$

I assume that light reflected from a Mondriaan falls onto a planar array of sensors. At each location X in the sensor array there are 3 different classes of sensors[1]. The value registered by the kth sensor, p_k^X (a scalar), is equal to the integral of its response function multiplied by the incoming color signal. This model accurately describes color

[1] In principle there could be more than 3 sensors. However because human vision and the majority of color cameras are trichromatic I will, unless stated otherwise, assume trichromacy.

image formation for the human eye and color cameras [WS82]. For convenience, we arrange the index X such that each p_k^X corresponds to a unique surface reflectance:

$$p_k^X = \int_\omega C^X(\lambda) R_k(\lambda) d\lambda \tag{1.2}$$

where $R_k(\lambda)$ is the response function of the kth sensor, $C^X(\lambda)$ is the color signal at X and the integral is taken over the visible spectrum ω. Using an underscore to denote vector quantities, and understanding the one-to-one correspondence between scene point and color response (henceforth I drop the X), equation (1.2) can be rewritten as:

$$\underline{p} = \int_\omega C(\lambda) \underline{R}(\lambda) d\lambda \tag{1.3}$$

While the Mondriaan world is certainly restrictive it is extremely useful in developing theories of color constancy. First, comparison can be made with existing Mondriaan based algorithms (the majority are based on the Mondriaan world). Second, because the Mondriaan world is free of factors such as specularities and mutual illumination, their presence cannot be assumed in solving for color constancy and there is a greater need for a competent unconstrained recovery algorithm. Lastly human observers exhibit reasonable color constancy for the Mondriaan world, demonstrating that the color constancy problem is soluble in the absence of other constraints.

As a basis for designing a recovery algorithm, equation (1.3) is not very encouraging. Each observation is a vector of 3 numbers whereas each reflectance spectrum can, in principle, have infinite detail. An observation can only be accurately mapped to its descriptor if this infinite detail can be extracted from an image. Of course this cannot be done. Fortunately the model in (1.3) is overly general and can be replaced by much simpler discrete approximations.

1.1.1 Discrete Approximations

Color signals are not arbitrary functions of wavelength but are band-limited; by Nyquist's [KKOP66] famous sampling theorem, a continuous function of wavelength can be precisely characterized by its values at a discrete number of sample points. In

color science, it is common practice to sample spectral data every 10nm across the visible spectrum. For example the data sets of natural reflectances and Munsell chips measured by Krinov [Kri47] and Nickerson [Nic57] sample spectra every 10nm from 400 to 650nm and from 380 to 770nm respectively.

The precise range over which spectra are sampled varies from study to study and should depend on the sampling properties of the particular visual system under investigation. When considering human vision it is crucial to include data points in the range 400 to 650nm since most of the eye's sensitivity is concentrated there. This range is generally the most important for color cameras; cameras are generally designed to capture color information useful to human observers.

In this sampling framework illuminants, reflectances and sensor response functions are effectively 26- or 31-dimensional vectors and the integral equation (1.3) is computed as a summation:

$$p_k = \sum_{i=1}^{n} E(\lambda_i)S(\lambda_i)R_k(\lambda_i)\Delta\lambda \tag{1.4}$$

where the factor $\Delta\lambda$ accounts for the sample spacing. For the most part sample points from 400 to 650nm at 10nm intervals will be used in this thesis. Henceforth I will assume the term $\Delta\lambda$ is incorporated into the sensor response vectors. For almost all real illuminants and reflectances equation (1.4) approximates the definite integral in equation (1.3) with a vanishingly small error [SSS92].

Because color measuring devices only provide estimates of spectral functions at discrete sample points it is important from an analysis point of view that the color image integral can be accurately described by a vector approximation. Under the discrete formulation the color constancy problem can be formulated using linear algebra. The tools of linear algebra, used extensively in this thesis, have proved themselves of enormous value in studying the color constancy problem.

1.1.2 Finite basis models of illumination and reflectance

The description of illuminant spectral power distribution functions and surface spectral reflectance functions can be further simplified using finite-dimensional models. A

surface reflectance $S(\lambda)$ can be approximated as:

$$S(\lambda) \approx \sum_{i=1}^{d_S} S_i(\lambda)\sigma_i \tag{1.5}$$

where $S_i(\lambda)$ is a basis vector and $\underline{\sigma}$ is a d_S-component column vector of weights. Anything from 3 to 8 basis vectors are required to describe reflectance [Mal86, PHJ89a, VGI94]. Similarly illuminants are well described by a low–dimension basis set:

$$E(\lambda) \approx \sum_{j=1}^{d_E} E_j(\lambda)\epsilon_j \tag{1.6}$$

$E_j(\lambda)$ is a basis vector and $\underline{\epsilon}$ is a d_E dimensional vector of weights. Judd [JMW64] measured 605 daylight illuminants and showed they are well modelled by a set of 3 basis functions.

Basis functions are generally chosen by performing a principal component analysis of each data set (reflectances and illuminants) in isolation [Coh64, Nic57, Mal86]. This type of analysis is weak in the sense that it does not take into account how illuminant, reflectance and sensor interact in forming a color vector (eqn. (1.3)). Recently Marimont and Wandell [MW92] developed a method for deriving reflectance and illuminant basis functions which best model color observations—eqn. (1.3) is the foundation for their method. They conclude that, given the human cones, a 3-dimensional basis set for surface reflectance and a 3-dimensional basis set for illumination is sufficient to model the color observations of the 462 Munsell chips [Nic57] under a wide range of black-body radiator illuminants.

A *lighting matrix* $\Lambda(E(\lambda))$ maps reflectances, defined by the $\underline{\sigma}$ vector, onto a corresponding observation vector:

$$\underline{p} = \Lambda(E(\lambda))\underline{\sigma} \tag{1.7}$$

where $\Lambda(E(\lambda))$ is a $3 \times d_S$ matrix; $\Lambda(E(\lambda))_{ij} = \int R_i(\lambda)E(\lambda)S_j(\lambda)d\lambda$. If $E(\lambda)$ is defined in equation (1.6), then the lighting matrix depends only on the $\underline{\epsilon}$ weight vector and is denoted $\Lambda(\underline{\epsilon})$. The roles of illumination and reflectance are symmetric; we can write a color observation as a *surface matrix* transforming an epsilon vector:

$$\underline{p} = \Omega(S(\lambda))\underline{\epsilon} \tag{1.8}$$

where $\Omega(S(\lambda))$ is a $3 \times d_E$ matrix; $\Lambda(S(\lambda))_{ij} = \int R_i(\lambda)E_j(\lambda)S(\lambda)d\lambda$. If $S(\lambda)$ is defined in equation (1.5) then the surface matrix depends only on the $\underline{\sigma}$ weight vector and is denoted $\Lambda(\underline{\sigma})$.

1.2 Solving for Color Constancy

The aim of any theory of color constancy is to transform the color observation vector $\underline{p}^e$ to its corresponding illuminant-independent descriptor $\underline{p}^c$ (throughout this thesis the superscripts c and e denote dependence on a known fixed canonical light and unknown illuminant respectively).

$$\underline{p}^c = \Psi(\underline{p}^e) \tag{1.9}$$

The symbol Ψ in (1.9) represents a function or computational procedure. Of course the problem is quite unconstrained given a single observation vector. Let us instead suppose there are n matte surfaces in a given Mondriaan. Let us place the n observations in a $3 \times n$ matrix $\mathcal{P}^e$ and the corresponding n descriptors in the matrix $\mathcal{P}^c$. Rewriting (1.9):

$$\mathcal{P}^c = \Psi(\mathcal{P}^e) \tag{1.10}$$

Assuming linear models of reflectance and illumination (equations (1.5) and (1.6)), Forsyth has shown that the color constancy problem is soluble if and only if descriptors and observations are related by a 3×3 linear transform. Color constancy becomes the problem of determining the 9 parameters of the matrix $\mathcal{M}^{e,c}$:

$$\mathcal{P}^c \approx \mathcal{M}^{e,c}\mathcal{P}^e \tag{1.11}$$

Alternately assuming the trivial image model the color constancy involves finding the best scalar $\alpha^{e,c}$:

$$\mathcal{P}^c \approx \alpha^{e,c}\mathcal{P}^e \tag{1.12}$$

Or finding the diagonal matrix $\mathcal{D}^{e,c}$ in the coefficient model which satisfies:

$$\mathcal{P}^c \approx \mathcal{D}^{e,c}\mathcal{P}^e \tag{1.13}$$

Of course, in practice there will always be error in the mappings (1.11), (1.12) and (1.13). Color constancy is the problem of finding the mapping which minimizes the error.

1.2.1 Color Constancy and a Changing Illumination

To formalize the definition of the color constancy problem under varying illumination, consider that there are n matte surfaces viewed under m different unknown illuminants. Place the observations for each illuminant in a $3\times n$ matrix P^i $i = (1, 2, \cdots, m)$. The function Ψ mapping the set of observation matrices to the descriptor matrix $\mathcal{P}^c$ solves the color constancy problem under varying illumination.

$$\mathcal{P}^c = \Psi(\mathcal{P}^1, \mathcal{P}^2, \cdots, \mathcal{P}^m) \tag{1.14}$$

At first glance, it would appear that the multi-illuminant color constancy problem adds more information since the input set is larger. Indeed, the color constancy problem becomes soluble even where there is no linear transform which exactly maps between illuminants [D'Z92], and as such constancy is still possible when there is metamerism (i.e. when two surfaces viewed under one illuminant look the same but appear differently under some other illuminant). The human visual system does not solve for color constancy where there is metamerism (if it did metamerism would not occur).

1.3 Image Model Performance

What does it mean to say that one image model is better than another? To address this question, a measure of image performance is required. To quantify the efficacy of a particular image model I find the map which best takes an exemplar set of observations onto their corresponding descriptors. The Euclidean distance between

each mapped observation and descriptor, normalized to the descriptor's length, will then be used to measure performance.

1.3.1 The Data Set

To calculate image performance a data set of illuminant and reflectance spectra is required. Unless otherwise stated I will use the 426 Munsell chip reflectances, measured by Nickerson [Nic57], D55 [JMW64] (cloudy skylight) for the canonical light, and 4 other daylights [JMW64] (D48, D65, D75 and D100) and the CIE standard A source [WS82] as test illuminants. These illuminants span the range of red, white and blue illuminants; the reddest is CIE A followed by the progressively bluer daylights D48, D65, D75 and D100.

The Munsell set was chosen because it contains a large variety of reflectances: saturated blues, reds and greens and everything in between. Moreover the set has become somewhat of a standard in the color constancy literature and is often used as a benchmark set. My choice of illuminants is more arbitrary. Once again a motivation for using them comes from their wide application in the literature. More crucially, there is a paucity of published illuminant spectra.

Finally I will use the Vos Walraven cone fundamentals [WS82] for response functions. The Vos Walraven curves are sketched in Figure 1.5.

1.3.2 A Least-squares measure of image performance

Given the spectral sensitivities of a visual system (color camera or human eye) and measured reflectances and illuminants, color observations can be simulated using equation (1.4). If there are n reflectances and m illuminants then m, $3 \times n$ observation matrices $\mathcal{P}^e$ are readily constructed. One of these, $\mathcal{P}^c$, is chosen to be the canonical illuminant (Forsyth [For90] chooses white light). For each of the remaining $m-1$ observation matrices the mapping which best takes observations to descriptors is calculated:

$$\mathcal{P}^c \approx \Psi^{\text{model}}(\mathcal{P}^e) \tag{1.15}$$

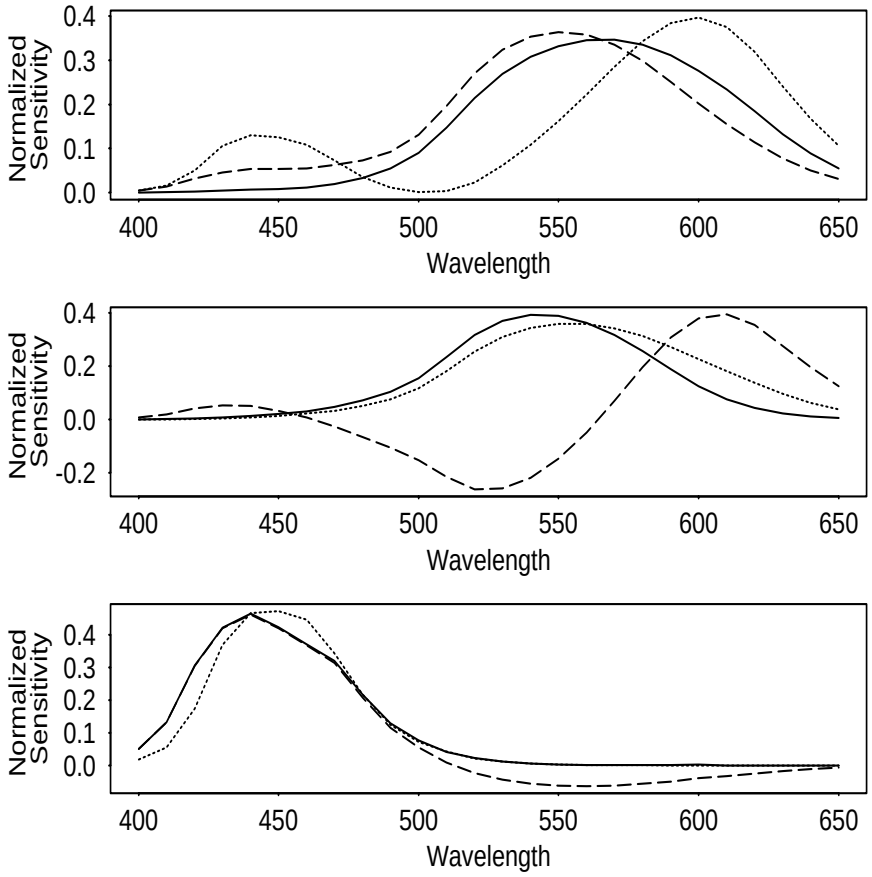

Figure 1.5: Vos Walraven fundamentals (solid line) are contrasted with XYZ color matching curves (dotted lines) and opponent type channels (dashed lines).

Ψ^{model} is an image model dependent *oracle* color constancy procedure. For a particular image model Ψ^{model} transforms $\mathcal{P}^e$ such that the error in (1.15) is minimized with respect to some criterion. The term oracle draws attention to the fact that this transformation is calculated given the correspondence between observations and descriptors (the columns of $\mathcal{P}^c$ and $\mathcal{P}^e$). Of course this correspondence will not in general be known and as such (1.15) bounds the performance of all recovery algorithms for a given image model.

To quantify the performance of an image model I compare fitted observations $\Psi^{\text{model}}(\mathcal{P}^e)$ with their corresponding descriptors. Specifically for each fitted observation, I calculate the Euclidean distance from its descriptor normalized with respect to the descriptors length:

$$\frac{||\mathcal{P}_i^c - \Psi^{\text{model}}(\mathcal{P}^e)_i||}{||\mathcal{P}_i^c||} \tag{1.16}$$

where the subscript i indexes the columns of the descriptor and mapped observation matrices. Equation (1.16) measures the percent normalized fitted distance or NFD for short.

The NFD measured for one mapped observation and one descriptor does not shed much light on the power of a particular image model. I will use cumulative histograms of NFD measurements as a means to compare many such pairs.

1.3.3 An Example of Image Performance

In Figure 1.6, I contrast the NFD cumulative histograms for the mapping performance of the trivial, coefficient and general linear image models. Notice that the general image model vastly outperforms the trivial model for all test illuminants. The coefficient model performs better than the trivial model but markedly poorer compared with the general model.

The coefficient model which has 3 free parameters represents a compromise between the trivial (doing nothing) and general (doing as much as possible) models. While performance is still far behind the general case, it is much improved over doing nothing.

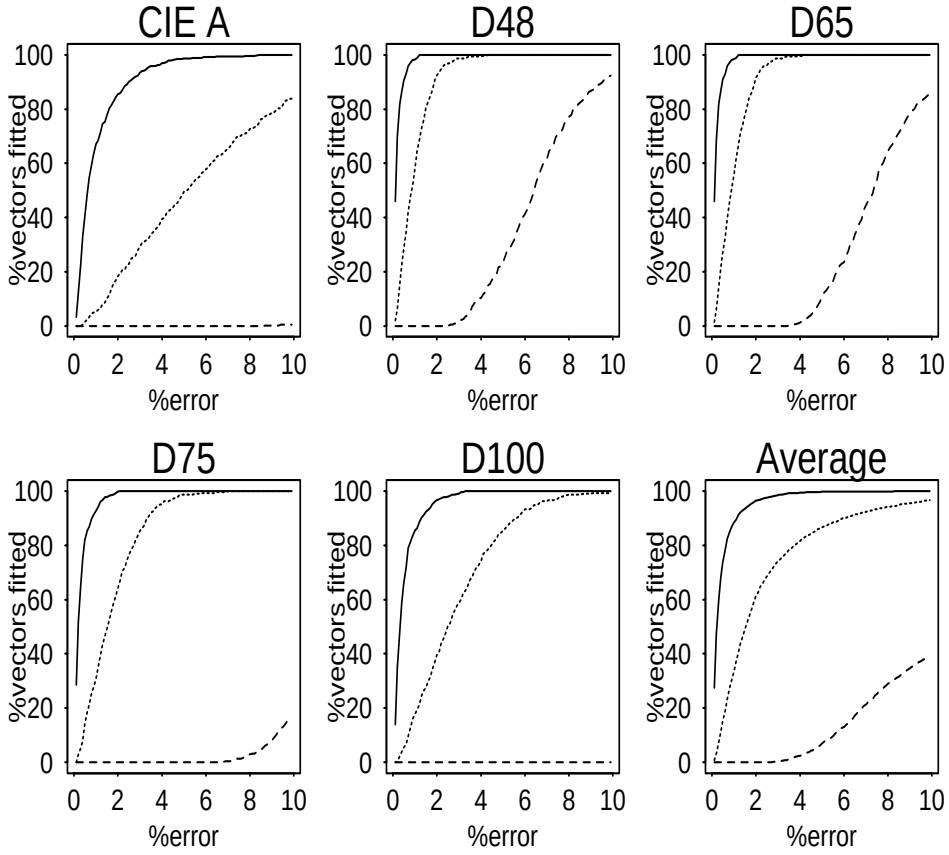

Figure 1.6: Cumulative NFD histogram obtained with each test illuminant (CIE A, D48, D65, D75, and D100) for general (solid lines), coefficient (dotted lines) and trivial image models(dashed lines).

Under the generalized coefficient scheme scaling coefficients are applied to a basis other than the cones. For a different basis the cumulative histogram curve will shift; hopefully toward the general linear performance. In Figure 1.5 I compare the cones with two other bases: the XYZ color matching functions [WS82] and Buchsbaum's [Buc87] color opponent channels. XYZ color matching curves are the standard space for specifying color. Opponent channels efficiently encode color information for transmission. The NFD error for cones, XYZ curves, and opponent channels is shown in Figure 1.7. The NFD curves for XYZ or opponent observations are calculated in two stages. First the best scaling coefficients are applied to XYZ (or opponent) observations mapping them to corresponding XYZ (or opponent descriptors. Both descriptors and mapped observations are now transformed back to the cone basis where the NFD error is calculated. Moving back to the cone space ensures a fair comparison of error.

Both the XYZ and opponent channels perform less well than the cone basis: the respective histogram curves are both below the cone histogram curve. However the performance difference for the XYZ basis is small. This is quite interesting since there is continuing controversy in the color community over whether to apply scaling coefficients to the cone or XYZ bases [MB93]. That the cone basis is currently favoured (e.g. [FB93]) is in part justified by the data in Figure 1.7. Neither the XYZ basis nor the cone basis comes close to the general image model performance.

It is a central aim of this thesis to find bases which shift the coefficient curve toward the general curve; presupposing that such bases exist.

1.4 Constrained Recovery Algorithms

The one-to-one correspondence between observations and descriptors used in quantifying image model performance constitutes the strongest constraint in solving for color constancy. Trivially, when given this correspondence one could simply substitute the correct descriptor for each observation thereby achieving perfect constancy. Of course it is highly unlikely that such a one-to-one correspondence will be known in advance.

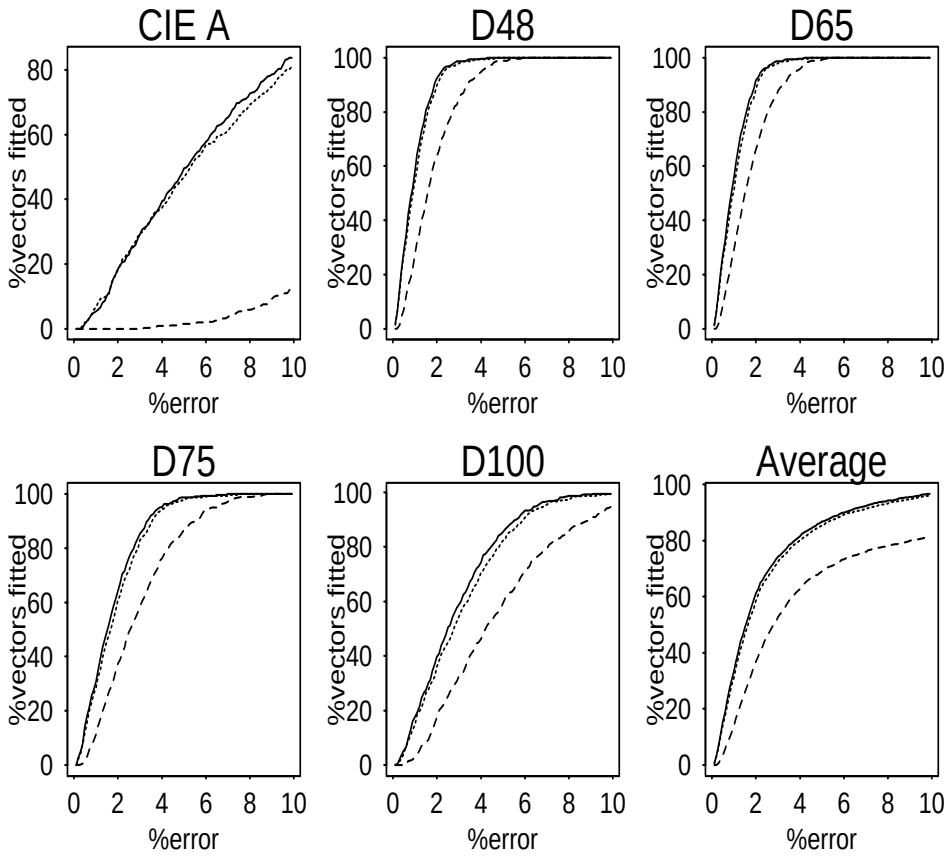

Figure 1.7: Cumulative NFD histogram obtained with each test illuminant (CIE A, D48, D65, D75, and D100) for the coefficient image model operating with respect to three sensor bases: the cones (solid lines), the XYZ matching curves (dotted lines) and opponent channels (dashed lines).

Instead, it is often assumed that the correspondence between observations and descriptors is known for a subset of the surfaces in an image. For example in the supervised color constancy scheme of Novak and Shafer [NS90], a Macbeth [MMD76] color checker card containing 24 known reflectances is placed in each scene. Thus there are 24 correspondence pairs and these are sufficient to estimate accurately the map back to the canonical illuminant. Even those observations not on the calibration card are accurately mapped to their descriptors.

While there are applications where it would be reasonable to place a reference card in a scene, e.g. calibrated imaging or color correction [NSW90, Vrh93], it is a requirement that we ourselves do not require and one that we would not like to impose on a general machine vision system. For this reason many authors have sought to decrease the number of reference patches which must appear in each scene.

Brill [Bri80, Bri78] demonstrates that if surface reflectances are 3-dimensional then only 3 reference patches are necessary to define the general linear mapping back to the canonical illuminant. Even if this constraint were to hold the reference patches must still be located. Of course their location could be fixed—though in this circumstance one might as well use the Macbeth Color checker and recover the illuminant map with higher accuracy.

Many authors have reduced the number of known reference patches to one. Moreover the reference is defined in such a way that its location is easily derived from the image circumventing the location problem of Brill. For example Land [Lan77], in one of the early incarnations of his Retinex theory, of color constancy theory assumes that every scene contains a white patch. Moreover this white is the most reflective over all other whites. The white patch is easily located in an image by searching for the observation vector with the largest *rgb* response[2].

However because a white observation vector is defined by 3 numbers it follows that only 3 parameters of the mapping back to the canonical can only be recovered. Coefficient maps are precisely characterized by three parameters and are used in

[2]Strictly this is true only when illumination is everywhere uniform. Land's Retinex will also work when illumination power varies. In this case the patch with the largest *lightness* is located (this is the patch that would have the largest response under constant illumination).

Retinex theory.

As shown in Figure 1.7 the coefficient model operating on the cones or opponent channels (Land suggests both [Lan83]) results in fairly poor color constancy relative to the general model. This has prompted other authors to search for a more realistic set of mappings which are parameterized by three variables. This search usually begins with finite-dimensional descriptions of illumination and reflectance. Specifically given the observation vector $\underline{p}^{w,e}$ of a white surface w under an unknown illuminant e then, following from the definition of a surface matrix equation (1.8), a 3-parameter description of the illuminant spectrum $E^e(\lambda)$ can be recovered:

$$\underline{\epsilon}^e = [\Omega(S_w(\lambda))]^{-1}\underline{p}^{w,e} \tag{1.17}$$

where $S_w(\lambda)$ is the reflectance spectrum of white and $E^e(\lambda)$ can be reconstructed using $\underline{\epsilon}^e$ and equation (1.6). Assuming a 3-parameter description of reflectance it is straightforward to calculate the lighting matrices for the canonical and unknown illuminants: $\Lambda(E^c(\lambda))$ and $\Lambda(\underline{\epsilon}^e)$ ($\underline{\epsilon}^e$ defined in (1.17)); from which it follows that the mapping from unknown to canonical illuminants is defined as:

$$\mathcal{M}^{e,c} = \Lambda(E^c(\lambda))[\Lambda(\underline{\epsilon}^e)]^{-1} \tag{1.18}$$

Both Buchsbaum [Buc80] and Gershon et al [GJT88] try to solve for color in this manner though they assume a grey reference patch instead of a white. Moreover the grey refers not to a specific surface in the scene but to a property derived from the image as a whole. In Buchsbaum's theory it is assumed that the average color observation in any image will equal the color response for some grey reflectance. Obviously this heuristic is easy to confound; for example if the majority of an image contains only shades of red the mean vector will also correspond to a red reflectance. Versions of Land's Retinex based on a similar grey world assumption [LM71] are also criticized for this flaw [BW86]. Gershon et al [GJT88] proposes that each distinct color observation is counted only once when the average is calculated. Once again this heuristic can quickly run into problems; if the image contains mostly shades of red, the mean observation vector will still be red.

Similarly, Land tried to assuage criticism of his white patch retinex [Lan77] by providing a computational scheme for calculating the white observation even when a white surface does not appear in the scene. He showed that so long as each scene contains reflectances which are maximally reflective (to the same degree as the whitest white) for each individual sensor class (e.g. the red, green and blue cones) then the white observation can be derived. One simply records the maximum response in each sensor channel.

In the context of human vision there is evidence that this maximum normalization is in fact carried out. McCann [McC94] reports on experiments where Mondriaans are displayed to human observers. Changing the composition of the Mondriaan, while keeping a test patch constant, does not alter the perceived color of the test unless the change in composition alters the maximal response in at least one of the cone channels. This apparent sensitivity to the maximum is indicative of imperfect human color constancy.

Other authors have abandoned the reference patch condition and try to construct recovery algorithms which are more robust to scene composition. Non-reference cues are generally sought outside of the Mondriaan world. For example many authors assume that specularities are present in each scene [LBS90, Lee90, Sha85, TW89, Dre93]. Each color observation vector is modelled as the sum of a Lambertian reflection component, defined in equation (1.3), with a specular component—the color response of the illuminant itself. For example if the Lambertian component is $\underline{b}^{i,e} = \int_\omega \underline{R}(\lambda)S_i(\lambda)E^e(\lambda)d\lambda$ and the specular reflection is $\underline{s}^e = \int_\omega \underline{R}(\lambda)E^e(\lambda)d\lambda$ then the color observation is defined as:

$$\underline{p}^{i,e} = \alpha\underline{b}^{i,e} + \beta\underline{s}^e \tag{1.19}$$

where α and β are scalars not less than 0. It follows that the observation vectors corresponding to a single surface lie on a plane ($\underline{b}^{i,e}$ and $\underline{s}^e$ are a basis set) and that the intersection of two surface planes is the specular component $\underline{s}^e$. Given $\underline{s}^e$, strong constraints can be placed on the position of $\underline{b}^{i,c}$ (i.e. the descriptor for $\underline{p}^{i,e}$) though a unique solution is not possible [Tom94]. Psychophysical experiments suggest that human observers do not use specular cues in solving for color constancy [Hur89].

Other non-reference constraints which have been used include mutual illumination [FDH91, FH88]; the idea that in describing the unknown illuminant we should assume it is as white as possible [RS89]; and probabilistic assumptions about the distributions of reflectance and illumination [TV91, DI94, BF94]. There are no reports in the literature of these methods being implemented for color machine vision; nor are there any indication from the psychophysical community to suggest that these cues might be used in human color vision.

1.5 Unconstrained Recovery Algorithms

Unconstrained algorithms aim to solve the color constancy problem with no constraints save those inherent in the image model. Since these are relatively weak, unconstrained recovery algorithms rarely return unique solutions but rather attempt to restrict descriptor assignments.

In the trivial image model a change of illumination results in a simple scaling of observation vectors. Clearly the trivial model does not hold for human cones—see Figure 1.6. For machine vision systems however, it is in principle possible to design sensors for which the trivial image model is valid. Brainard, Wandell and Cowan [BWC89] set forth the necessary *black-light* spectral conditions given low dimensional linear descriptions of reflectance and illumination.

Supposing that the trivial model accurately reflects illumination change then a degree of color constancy can be attained simply by normalizing the length of response vectors. For example the vectors $\underline{p}^c$ and $\alpha\underline{p}^c$ normalized with respect to their length are identical since:

$$\frac{\underline{p}^c}{||\underline{p}^c||} = \frac{\alpha\underline{p}^c}{||\alpha\underline{p}^c||} \tag{1.20}$$

Thus a 2 parameter descriptor is readily recovered (the normalized vector can be described by two angles e.g. azimuth and elevation). Unfortunately Maloney [Mal90] has shown that filters which satisfy the black-light conditions measure almost none of the signal present in real color signal spectra and as such would not provide suitable

input for a vision system. Moreover this deficiency cannot be mitigated by changing the sensor basis [Bri93].

A coefficient model operating on cone observations supports a reasonable degree of color constancy (see Figure 1.6). Moreover it is well known that this performance improves when narrow-band sensors are employed [WB86]; in the limit, sensors which are sensitive to a single wavelength of light are known to support perfect coefficient color constancy [For90].

Under the coefficient model a descriptor under the canonical illuminant can in principle move to any other 3-vector when the illumination changes: there is no a priori constraint on where observations should lie. However under an illumination change all observation vectors shift by the same 3 scaling coefficients; Forsyth [For90] shows that this common bond is a powerful constraint in solving for color constancy.

In particular Forsyth's CRULE (coefficient rule) algorithm operates as follows. As a preprocessing step the gamut of all descriptors is estimated. The gamut is simply the set of responses induced from many different reflectances and is represented as a body in 3-space (3 since cameras and the eye are trichromatic). When a single surface is viewed under an unknown illumination the observation vector it induces must correspond to a descriptor in the canonical gamut. From this, it follows that only those scaling coefficients which map the observation somewhere inside the canonical gamut are possible solutions to the color constancy problem. A second surface observed under the same unknown illuminant leads to a second set of possible solutions. Since the same coefficient scalings must map both observations to the canonical illuminant the color constancy solution must lie in the intersection of the two constraint sets. A third surface further constrains the solution set and so on. In the best case if an image contains all reflectances then CRULE uniquely solves the color constancy problem (all constraint sets have a single point in common).

In practice if an image contains many reflectances the set of possible solutions is small and any of the coefficients mappings will result in good color constancy. Unfortunately many realistic images contain few surfaces and in this case the solution set may be large. Forsyth proposes that the coefficient scalings which maximize the volume of the mapped image colors should be chosen. Of course this is a heuristic

and may or may not lead to good color constancy.

A more serious problem with CRULE is that it can only work for the Mondriaan world. Changes in illumination intensity across a scene, specularities and changing object shape all locally affect color observations and the assumption that the same set of scaling coefficients takes observations to descriptors no longer holds.

The general linear image model has the advantage that it can be directly applied to most sensor sets, including the human cones. The disadvantage is that it has 9 free parameters (3 columns and 3 rows) and these must all be retrieved. Recovery is especially hard for unconstrained algorithms whose only constraint is the image model.

In principal the general linear solution set can be constrained in a similar manner to CRULE. For each observation the set of linear maps taking it into the canonical gamut is readily calculated. Intersecting these constraint sets (one per surface) returns those linear maps which take all observations back to the canonical illuminant. Indeed this procedure is at the heart of Forsyth's MWEXT general linear recovery algorithm.

Unfortunately, because a linear transform has 9 parameters each set of illuminant maps occupies a region in 9-space. To calculate the intersection of many 9-dimensional bodies with each other is an expensive computational task. Indeed there is no evidence in the literature that MWEXT has in fact been implemented.

To reduce computational cost simplifying geometric constraints can be applied to the shape of the canonical gamut. Specifically if the set of canonical descriptors lies on a 2-dimensional plane passing through the origin then this implies that for any unknown illuminant the set of corresponding observations would also lie on a plane (a plane transformed by a linear transform is still a plane). In this framework it is only necessary to characterize the set of mappings which map the image plane onto the canonical plane. This set of mappings can be speedily calculated.

Unfortunately the set of valid maps is quite large—indeed in the absence of other constraints forcing the canonical gamut to a plane is not very useful. However under the conditions where illumination and reflectance belong to 3- and 2-dimensional linear models (the *3–2* case) Maloney [Mal85] has shown that the canonical gamut is exactly a plane and the mapping between image plane and canonical gamut is unique.

Moreover because a plane through the origin is defined by only two observation vectors color constancy is possible at a color edge. That is, it is an unconstrained algorithm which delivers unique color constancy so long as there are at least two distinct surfaces in each scene.

Two criticisms can be made about the Maloney-Wandell approach. First, and this is crucial, reflectance is not 2-dimensional nor even approximately so. Thus in practice the algorithm does not and cannot work. Secondly under the *3–2* conditions it cannot be claimed that the Maloney-Wandell algorithm adheres to the general linear model. Indeed because illumination is 3-dimensional the maps taking observations to descriptors must be parameterized by 3 numbers.

However this observation is balanced by the fact that there are no reports in the literature of a truly general unconstrained linear recovery method of ever having being implemented.

1.6 Outline of Main Results

Chapters 2, 3, and 4 discuss different methods for finding good bases for coefficient color constancy. Chapter 2 itself sets forth 3 methods: *sensor-based sharpening, data-based sharpening and perfect sharpening.* The word sharpening alludes to our expectation that the sensors in the preferred basis should appear visually more narrow-band. Sensor-based sharpening operates independently of any statistical assumption about reflectance or illumination. Rather, it sets out to find sensors which 'look' sharp. Data-based sharpening finds coefficient channels making direct use of observation and descriptor vectors. For any pair of illuminants data-based sharpening returns the optimal coefficient channels. Perfect sharpening is a method for choosing coefficient channels based on linear models of illumination and reflectance. Specifically if illumination is 2-dimensional and reflectance 3-dimensional then there exists a sensor basis, for all sensor sets, for which the coefficient model exactly characterizes illumination change. All three sharpening methods return similar bases. Each basis elevates coefficient color constancy performance to a similar level to that of the general model.

In Chapter 3, I show that when illumination and reflectance are described by 3-

and 2-dimensional linear models (the complement of perfect sharpening assumptions) that there is once again a sensor basis where a coefficient model suffices to explain illumination change. To distinguish the *3–2* derivation from the sharpening transforms I call this method generalized diagonal matrix color constancy. Because many existing theories of color constancy operate under *3–2* restrictions they are all in effect generalized diagonal matrix theories of color constancy. I describe in detail how the Maloney-Wandell algorithm can be elegantly implemented in the generalized diagonal framework.

Chapter 4 is dedicated to two tasks. First our original intuition that narrower is better is re-examined from a different perspective. Specifically I set out to find the sensor basis which samples color signal spectra in a manner most like a trichromatic narrow-band sensor set (where all sets of three narrow-band sensors are initially considered). I derive two bases, corresponding to two different sets of color signal spectra. The *maximum ignorance* set contains all possible spectra and the calibration set various Munsell spectra illuminated by several test illuminants. In both cases the derived bases, which I call maximum ignorance and calibration sensors, improve the performance of coefficient color constancy.

As a second task I set out to compare, theoretically and experimentally, each method for choosing the best coefficient basis. I show that the visually driven sensor-based sharpening is theoretically well founded given the maximum ignorance assumption. I present experimental results which demonstrate that while all derived bases significantly improve the performance of coefficient color constancy, some methods are better than others. As might be expected, the methods can be ranked according to the accuracy of their statistical assumptions. The data-based sharpened sensors are best with the maximum ignorance sensors showing the relatively poorest performance.

In Chapter 5, I set forth a new unconstrained recovery algorithm called *color in perspective* based on Forsyth's CRULE. I begin with a detailed discussion of Forsyth's CRULE algorithm highlighting its strengths and weakness. I show that many of the latter are mitigated by reducing the 3-dimensional descriptor recovery problem to 2-dimensions. Specifically I show that if one sets out to recover only the orientation of descriptors then the shackles of the Mondriaan world can be broken and color

constancy becomes possible even in the presence of shape and specularities and where the illumination power varies across a scene.

The key observation underlying Forsyth's CRULE is that the range of color observations varies with a change in illumination. To this I add another constraint: that the illumination itself can only vary within certain limits. Color in perspective coupled with this illumination constraint can deliver excellent color constancy.

Chapter 6 examines the multi-illuminant color constancy problem. The input consists of the observations of many surfaces under many illuminants. I show that if the coefficient model reasonably accounts for illumination change (with respect to the appropriate basis) then the multi-illuminant color constancy problem is no easier than the single illuminant case. The only exception to this is if illuminations are well described by a 2-dimensional linear model. Under this condition a generalized coefficient scheme suffices for all sensor sets. Color constancy becomes possible because the diagonal matrix which maps between illuminants is unique.

A short conclusion in chapter 7 completes the thesis.

Chapter 2

Spectral sharpening

G.D. Finlayson, M.S. Drew and B.V. Funt. Spectral Sharpening: sensor transformations for improved color constancy. *The Journal of the Optical Society of America A.* 11:1553-1563, 1994.

Part of the work described in this chapter first appeared in: G.D. Finlayson, M.S. Drew and B.V. Funt. Enhancing von Kries Adaptation via Sensor Transformations. In *Proceedings: Human Vision, Visual Processing and Digital Display IV.* SPIE, The International Society for Optical Engineering, Vol 1903: 473-484, 1993.

Abstract

We develop sensor transformations, collectively called spectral sharpening, which convert a given set of sensor sensitivity functions into a new set that will improve the performance of any color constancy algorithm based on an independent adjustment of the sensor response channels. Independent adjustment of multiplicative coefficients corresponds to the application of a diagonal matrix transform (DMT) to the sensor response vector and is a common feature of many theories of color constancy: Land's retinex and von Kries adaptation in particular. We set forth three techniques for spectral sharpening. Sensor-based sharpening focuses on producing new sensors as linear combinations of the given ones such that each new sensor has its spectral sensitivity concentrated as much as possible within a narrow band of wavelengths. Data-based sharpening, on the other hand, extracts new sensors by optimizing the ability of a DMT to account for a given illumination change by examining the sensor response vectors obtained from a set of surfaces under two different illuminants. Finally in Perfect sharpening we demonstrate that, if illumination and surface reflectance are described by 2- and 3-parameter finite-dimensional models, there exists a unique optimal sharpening transform. All three sharpening methods yield very similar results. When sharpened cone sensitivities are used as sensors, a DMT models illumination change extremely well. We present simulation results suggesting that in general non-diagonal transforms can do only marginally better. Our sharpening results correlate well with the psychophysical evidence of spectral sharpening in the human visual system.

2.1 Introduction

The performance of any color constancy algorithm whether implemented biologically or mechanically will be strongly affected by the spectral sensitivities of the sensors providing its input. While in humans the cone sensitivities obviously cannot be changed, we need not assume that they form the only possible input to the color constancy process. New sensor sensitivities can be constructed as linear combinations of the original sensitivities, and in this chapter we explore what the most advantageous such linear transformations might be.

We call the sensor response vector for a surface viewed under an arbitrary test illuminant an *observation*. The response vector for a surface viewed under a fixed canonical light is called a *descriptor*. We will take as the goal of color constancy that of mapping observations to descriptors. Since a descriptor is independent of illumination it encapsulates surface reflectance properties [Bec72].

In discussing a color constancy algorithm there are two separate issues: the type of mechanism or vehicle supporting the transformation from observations to descriptors in general; and the method used to calculate the specific transformation applicable under a particular illumination. In this chapter, we address only the former and therefore are *not* proposing a complete new theory of color constancy.

A diagonal matrix transformation (DMT) has been the transformation vehicle for many color constancy algorithms, in particular von Kries adaptation [WB82], all the retinex/lightness algorithms [LM71, Hor74, Bla85], and more recently Forsyth's gamut-mapping approach [For90]. All these algorithms respond to changing illumination by adjusting the response of each sensor channel independently, although the strategies they use to decide on the actual adjustments they make differ.

DMT support of color constancy is expressed mathematically in equation (2.1). Here $\underline{p}^{i,e}$ denotes an observation (a 3-vector of sensor responses), where e and i index illumination and surface reflectances respectively. The vector $\underline{p}^{i,c}$ represents a descriptor and depends on the *single* canonical illuminant. The diagonal transform $\mathcal{D}^{e,c}$ best maps observations onto descriptors. Throughout, the superscript e will denote dependence on a variable illuminant and the superscript c dependence on the fixed

canonical illuminant. Underscoring indicates vector quantities.

$$\underline{p}^{i,c} \approx \mathcal{D}^{e,c}\underline{p}^{i,e} \tag{2.1}$$

In general, there may be significant error in this approximation. Indeed West and Brill [WB82] and D'Zmura and Lennie [DL86] have shown that a visual system equipped with sensors having the same spectral sensitivity as the human cones can achieve only approximate color constancy via a DMT.

A DMT will work better with some sensor sensitivities than others as can be seen by considering how the illumination, surface reflectance and sensor sensitivities combine in forming an observation. An observation corresponds to:

$$\underline{p} = \int_{\omega} E(\lambda)S(\lambda)\underline{R}(\lambda)d\lambda \tag{2.2}$$

where $E(\lambda)$, $S(\lambda)$, $\underline{R}(\lambda)$ denote illumination, surface reflectance and sensor sensitivities, respectively, and the integral is taken over the visible spectrum ω. For a DMT to suffice in modelling illumination change [WB82] it must be the case for an arbitrary reference reflectance S_r, all reflectances S and illuminants Ei, Ej that

$$\frac{\int_{\omega} E^i(\lambda)S(\lambda)\underline{R}(\lambda)d\lambda}{\int_{\omega} E^i(\lambda)S_r(\lambda)\underline{R}(\lambda)d\lambda} = \frac{\int_{\omega} E^j(\lambda)S(\lambda)\underline{R}(\lambda)d\lambda}{\int_{\omega} E^j(\lambda)S_r(\lambda)\underline{R}(\lambda)d\lambda} \tag{2.3}$$

As others have observed, one way to ensure this condition holds is to use extremely narrow-band sensors, which in the limit leads to sensors sensitive to a single wavelength (Dirac delta functions) [For90]. Our intuition when we began this work was that if we could find a linear combination of sensor sensitivities such that the new sensors would be *sharper* (more narrow-band), then the performance of DMT color constancy algorithms should improve[1] and the error of equation (2.1) would be reduced. With the addition of sharpening equation (2.1) becomes

$$\mathcal{T}\underline{p}^{i,c} \approx \mathcal{D}^{e,c}\mathcal{T}\underline{p}^{i,e} \tag{2.4}$$

[1]It should be noted, however, that equation (2.3) can be satisfied in other ways such as by placing constraints on the space of illuminants or reflectances [WB82].

where $\mathcal{T}$ denotes the *sharpening transform* of the original sensor sensitivities. It is important to note that applying a linear transformation to response vectors has the same effect as applying the transformation to the sensor sensitivity functions.

The sharpening transform effectively generalizes diagonal matrix theories of color constancy. Other authors [WB82, For90, LM71] have also discussed this concept of an intermediate (or sharpening) transform. However, our work appears to be the first to consider the precise form of this transform.

The sharpening transform is a mechanism through which the inherent simplicity of many color constancy algorithms can be maintained. For example, Land's retinex algorithm requires color ratios to be illumination independent (and hence implicitly assumes a diagonal matrix model of color constancy) which, as seen from equation (2.3), they generally will not be. It seems difficult to improve the accuracy of retinex ratioing directly without making the overall algorithm much more complicated [FD88]; however, by applying a simple, fixed sharpening transformation of the sensors as a preprocessing stage, the rest of the retinex process can remain untouched. Similar arguments apply to Forsyth's CRULE [For90] and Brill's [Bri80] volumetric theory.

We initially present two methods for calculating $\mathcal{T}$: sensor-based and data-based sharpening. Sensor-based sharpening is a general technique for determining the linear combination of a given sensor set which is maximally sensitive to sub-intervals of the visible spectrum. This method is founded on the intuition that narrow-band sensors will improve the performance of DMT theories of color constancy. We apply sensor-based sharpening over three different ranges in order to generate three new sharpened sensors that are maximally sensitive in the long, medium, and short wave bands. Figure 2.1 contrasts the cone fundamentals derived by Vos and Walraven [WS82] before and after sharpening. While the new sensitivity functions are "sharper," they are far from meeting the intuitive goal of being very narrow-band (i.e., with strictly zero response in all but a small spectral region); nonetheless, we perform simulations which show that they in fact work much better. Section 2.2 presents the details of sensor-based sharpening.

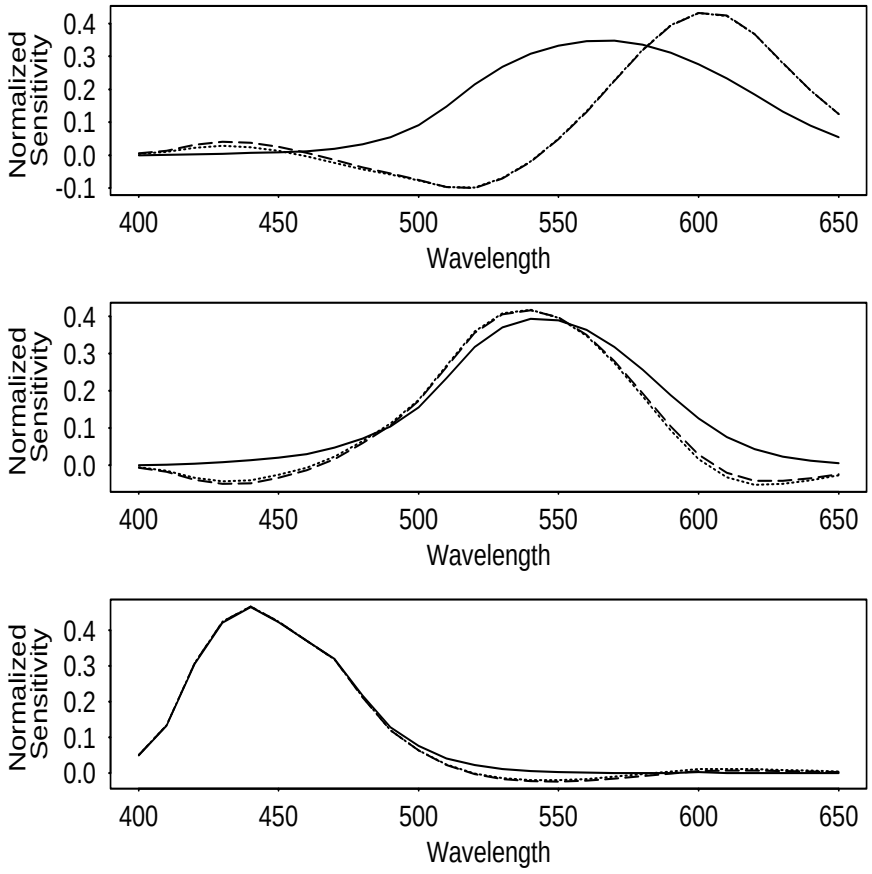

Figure 2.1: Vos Walraven fundamentals (solid line) are contrasted with the sharpened sensitivities derived via sensor-based (dotted line) and data-based (dashed line) sharpening.

Sensor-based sharpening does not take into account the characteristics of the possible illuminants and reflectances but considers only the sharpness of the resulting sensor. Our second sharpening technique, data-based sharpening, is a tool for validating the sensor-based sharpening method. Given observations of real surface reflectances viewed under a test illuminant and their corresponding descriptors, data-based sharpening finds the *best*, subject to a least-squares criterion, sharpening transform $\mathcal{T}$. Interestingly, data-based sharpening yields very stable results for all the test illuminations we tried and in all cases the data-based derived sensors are very similar to the fixed sensor-based sharpened sensors. Data-based sharpening is presented in section 2.3.

In section 2.4 we present simulations evaluating diagonal matrix color constancy for sharpened and un-sharpened sensor sets. Over a wide range of illuminations sensor-based sharpened sensors provide a significant increase in color constancy performance.

Data-based sharpening is related to Brill's [Bri80] volumetric theory of color constancy. This relationship is explored in section 2.5. Through spectral sharpening the volumetric theory is shown to be informationally equivalent to Land's [Lan77] white-patch retinex.

The data-based sharpening technique finds the optimal sharpening transform for a single test illuminant. In section 2.6 we investigate the problem of finding a good sharpening transform relative to multiple illuminants. If surface reflectances are 3-dimensional and illuminants 2-dimensional we show that there exists a sharpening transform with respect to which a diagonal matrix supports perfect color constancy. This analysis constitutes a third technique for deriving the sharpening transform.

Section 2.7 relates this work specifically to theories of human color vision. Sharpened spectral sensitivities have been measured in humans [Fos81, SH71, KH90, HF83, PW90]. We advance the hypothesis that sharpened sensor sensitivities arise as a natural consequence of optimizing the visual system's color constancy abilities through an initial linear transformation of the cone outputs.

2.2 Sensor-Based Sharpening

Sensor-based sharpening is a method of determining the sharpest sensor given an s-dimensional (s is usually 3) sensor basis $\underline{R}(\lambda)$ and wavelength interval $[\lambda_1, \lambda_2]$. The sensor $\underline{R}^t(\lambda)\underline{c}$, where $\underline{c}$ is a coefficient vector, is maximally sensitive in $[\lambda_1, \lambda_2]$ if the percentage of its norm lying in this interval is maximal relative to all other sensors. We can solve for $\underline{c}$ by minimizing:

$$I = \int_\phi [\underline{R}(\lambda)^t \underline{c}]^2 d\lambda + \mu(\int_\omega [\underline{R}(\lambda)^t \underline{c}]^2 d\lambda - 1) \tag{2.5}$$

where ω is the visible spectrum, ϕ denotes wavelengths outside the sharpening interval and μ is a Lagrange multiplier. The Lagrange multiplier guarantees a non-trivial solution for eqn. (2.5)—the norm of the sharpened sensor is equal to 1. Moreover, this constraint ensures that the same sharpened sensor is recovered independent of the initial norms of the basis set $\underline{R}(\lambda)$.

By differentiating with respect to $\underline{c}$ and equating to the zero vector, we find the stationary values of equation (2.5):

$$\frac{1}{2}\frac{\partial I}{\partial \underline{c}} = \int_\phi \underline{R}(\lambda)\underline{R}(\lambda)^t \underline{c} d\lambda + \mu(\int_\omega \underline{R}(\lambda)\underline{R}(\lambda)^t \underline{c} d\lambda) = \underline{0} \tag{2.6}$$

Differentiating with respect to μ yields the constraint equation $\int_\omega [\underline{R}(\lambda)^t \underline{c}]^2 d\lambda = 1$. The solution of eqn. (2.6) can thus be carried out assuming the constraint holds.

Define the $s \times s$ matrix $\Upsilon(\alpha) = \int_\alpha \underline{R}(\lambda)\underline{R}(\lambda)^t d\lambda$ so that equation (2.6) becomes:

$$\Upsilon(\phi)\underline{c} = -\mu\Upsilon(\omega)\underline{c} \tag{2.7}$$

Assuming a nontrivial solution $\underline{c} \neq 0$, $\mu \neq 0$ and rearranging equation (2.7), we see that solving for $\underline{c}$ (and consequently the sharpened sensor) is an eigenvector problem:

$$[\Upsilon(\omega)]^{-1}\Upsilon(\phi)\underline{c} = -\mu\underline{c} \tag{2.8}$$

There are s solutions of equation (2.8), each solution corresponding to a stationary value, so we choose the eigenvector which minimizes $\int_\phi [\underline{R}(\lambda)^t \underline{c}]^2 d\lambda$. It is important that $\underline{c}$ be a real-valued vector as it implies our sharpened sensor is a real-valued

function. That $\underline{c}$ is real-valued follows from the fact that the matrices $[\Upsilon(\omega)]^{-1}$ and $\Upsilon(\phi)$ are positive definite and that eigenvalues of the product of two positive definite matrices are real and non-negative [Wil65]. Solving for $\underline{c}$ for each of 3 wavelength intervals yields a matrix $\mathcal{T}$ for use in equation (2.4). The matrix $\mathcal{T}$ is not dependent on any illuminant and denotes the sensor-based sharpening transform.

We sharpened two sets of sensor sensitivities: the cone absorptance functions measured by Bowmaker [BH80] (BOW) and the cone fundamentals derived by Vos and Walraven [WS82] (VW), which take into account the spectral absorptions of the eye's lens and macular pigment. The BOW functions were sharpened with respect to the wavelength intervals (in nanometers) $[400, 480]$, $[510, 550]$ and $[580, 650]$ and the VW in the intervals $[400, 480]$, $[520, 560]$ and $[580, 650]$[2]. These intervals were chosen to ensure that the whole visible spectrum would be sampled and that the peak sensitivities of the resulting sensors would roughly correlate with those of the cones.

The results for the VW sensors are presented in Figure 2.1—those for the BOW sensors are very similar—where it can be seen that the sharpened curves contain negative sensitivities. These need not cause concern in that they do not represent negative physical sensitivities, but simply negative coefficients in a computational mechanism. Clearly, the sharpening intervals are somewhat arbitrary. They were chosen simply because the resulting sharpened sensors appeared, to the human eye, significantly sharper. Nevertheless, those used are sensible and their suitability is verified by the fact that they provide sharpened sensors which are in close agreement with those derived by data-based sharpening as described in the next section. The actual values of the $\underline{c}$'s in eqn. (2.8) are given below in section 2.6.1.

Figure 2.1 also contrasts the sharpened VW sensor set with the corresponding unsharpened set—the degree of sharpening is quite significant. The peak sensitivities of the new sensors are shifted with respect to the initial sensitivities, which is due both to the choice of sharpening intervals and the shape of the VW sensitivities. The sharpened long-wave mechanism is pushed further to the long-wave end of the spectrum; in contrast the medium wave mechanism is shifted towards the shorter wavelengths;

[2] All spectra used in this chapter are in the range 400nm to 650nm measured at 10nm intervals.

sensors	*% Squared Norm*		
	$[400, 480]$	$[510, 550]$	$[580, 650]$
BOW	98.9	51.4	30.1
BOW sharpened	99.4	66.3	76.2
	$[400, 480]$	$[520, 560]$	$[580, 650]$
VW	97.5	67.6	40.3
VW sharpened	97.8	74.7	89.1

Table 2.1: Percentage of total squared norm in the sharpening intervals

and the short wave mechanism remains essentially the same. Intriguingly, field sensitivities of the human eye measured under white light conditioning with long test flashes [SH71] are sharpened in an analogous manner.

Table 2.1 contrasts the percentage squared norm lying in the sharpening intervals for the original versus the sharpened curves. For both the VW and BOW sensors the degree of sharpening is significant. Furthermore, from Figure 2.1 it is clear that the sharpening effect is not limited to the sharpened interval.

2.3 Data-Based Sharpening

It could be the case that the best sensors for DMT algorithms might vary substantially with the type of illumination change being modelled. If so, sensor-based sharpening, which does not take into account any of the statistical properties of collections of surfaces and illuminants, might perform well in some cases and poorly in others. To test whether or not radically different sharpening transformations are required in different circumstances, we explore a data-based approach to deriving sharpened sensitivities in which the sensors are optimized for DMT algorithms by examining the relationship between observations, obtained under different test illuminants, and their corresponding descriptors.

Let $\mathcal{P}^c$ be a $3 \times n$ matrix of descriptors generated from a set n surfaces observed under a canonical illuminant E^c. Similarly, let $\mathcal{P}^e$ be the matrix of observations of n surfaces imaged under another test illuminant E^e. To the extent that DMT-based algorithms suffice for color constancy, $\mathcal{P}^c$ and $\mathcal{P}^e$ should be approximately equivalent

under a DMT:

$$\mathcal{P}^c \approx \mathcal{D}^{e,c}\mathcal{P}^e \tag{2.9}$$

The diagonal transform is assumed to be an approximate mapping and will have a certain degree of error. The idea of spectral sharpening is that this approximation error can be reduced if $\mathcal{P}^c$ and $\mathcal{P}^e$ are first transformed by a matrix $\mathcal{T}^{e,c}$.

$$\mathcal{T}^{e,c}\mathcal{P}^c \approx \mathcal{D}^{e,c}\mathcal{T}^{e,c}\mathcal{P}^e \tag{2.10}$$

$\mathcal{D}^{e,c}$ will in fact be optimal in the least-squares sense if it is defined by the Moore-Penrose inverse:

$$\mathcal{D}^{e,c} = \mathcal{T}^{e,c}\mathcal{P}^c[\mathcal{T}^{e,c}\mathcal{P}^e]^+ \tag{2.11}$$

where + denotes the Moore-Penrose inverse [3]. Now $[\mathcal{T}^{e,c}]$ must be chosen to ensure that $[\mathcal{D}^{e,c}]$ is diagonal. To see how to do this, carry out some matrix manipulation to yield:

$$[\mathcal{T}^{e,c}]^{-1}\mathcal{D}^{e,c}\mathcal{T}^{e,c} = \mathcal{P}^c[\mathcal{P}^e]^+ \tag{2.12}$$

Since the eigenvector decomposition of the matrix on the right-hand side of equation (2.12) $\mathcal{P}^c[\mathcal{P}^e]^+ = \mathcal{U}^{e,c}\mathcal{D}^{e,c}[\mathcal{U}^{e,c}]^{-1}$ is unique, its similarity to the left-hand side implies $\mathcal{T}^{e,c}$ is unique also, always exists and equals $[\mathcal{U}^{e,c}]^{-1}$, for diagonal $[\mathcal{D}^{e,c}]$.

It is interesting to compare equation (2.12) to that for the problem of finding the *best* general transform $\mathcal{G}^{e,c}$ mapping observations obtained under a test illuminant to their corresponding descriptors.

$$\mathcal{P}^c \approx \mathcal{G}^{e,c}\mathcal{P}^e \tag{2.13}$$

Such *optimal fitting* effectively bounds the possible performance within a linear model of color constancy. When the approximation of equation (2.13) is to be optimized in the least-squares sense, $\mathcal{G}^{e,c}$ is simply

[3] The Moore-Penrose inverse of the matrix A is defined as $A^+ = A^t[AA^t]^{-1}$.

$$\mathcal{G}^{e,c} = \mathcal{P}^c[\mathcal{P}^e]^+ \tag{2.14}$$

Equation (2.12) can be interpreted, therefore, as simply the eigenvector decomposition of the optimal general transform $\mathcal{G}^{e,c}$. Of course it is obvious that the optimal transform could be diagonalized; what is important is that if one knew a sharpening transform $[\mathcal{T}^{e,c}]$ then the best least-squares solution relating $\mathcal{T}^{e,c}\mathcal{P}^c$ and $\mathcal{T}^{e,c}\mathcal{P}^e$ is precisely the diagonal transformation $\mathcal{D}^{e,c}$, that is $\mathcal{T}^{e,c}\mathcal{P}^c[\mathcal{T}^{e,c}\mathcal{P}^e]^+ = \mathcal{D}^{e,c}$. In other words, when using the sharpened sensors the optimal transform is guaranteed to be diagonal, so finding the best diagonal transform after sharpening is equivalent to finding the optimal general transform. Therefore, sharpening allows us to replace the problem of determining the 9 parameters of $\mathcal{G}^{e,c}$ by that of determining only the 3 parameters of $\mathcal{D}^{e,c}$.

Data-based sharpening raises two main questions: Will the resulting sensors be similar to those obtained by sensor-based sharpening and will the derived sensors vary substantially with the illuminant? To answer these questions requires the application of data-based sharpening to response vectors obtained under a single canonical and several test illuminants. For illuminants, we used 5 Judd daylight spectra [JMW64] and CIE standard illuminant A [WS82], and as reflectances the set of 462 Munsell spectra [Coh64]. We arbitrarily chose Judd's D55 (2.55 stands for 5500K) as the canonical illuminant; descriptors are the response vectors for surfaces viewed under D55. For each of the other illuminants, $E^e(\lambda)$, the data-based sharpening transform $\mathcal{T}^{e,c}$ was derived in accordance with equation (2.12), so as in each case to best map via a DMT the observations under $E^e(\lambda)$ to those under D55.

For the VW sensors, Figure 2.2 shows the range of the 5 sets of data-based sharpened sensors obtained for mapping between each of the 5 illuminants and D55. For these 5 illuminants, the results are remarkably stable and hence relatively independent of the particular illuminant, so the mean of these sharpened sensors characterizes the set of them quite well. Referring once again to Figure 2.1, we can see that these mean sensors are very similar to those derived via sensor-based sharpening.

The stability of the results for the 5 cases and their similarity to the sensor-based result is reassuring; nonetheless, it would be nice to find the sharpening transformation

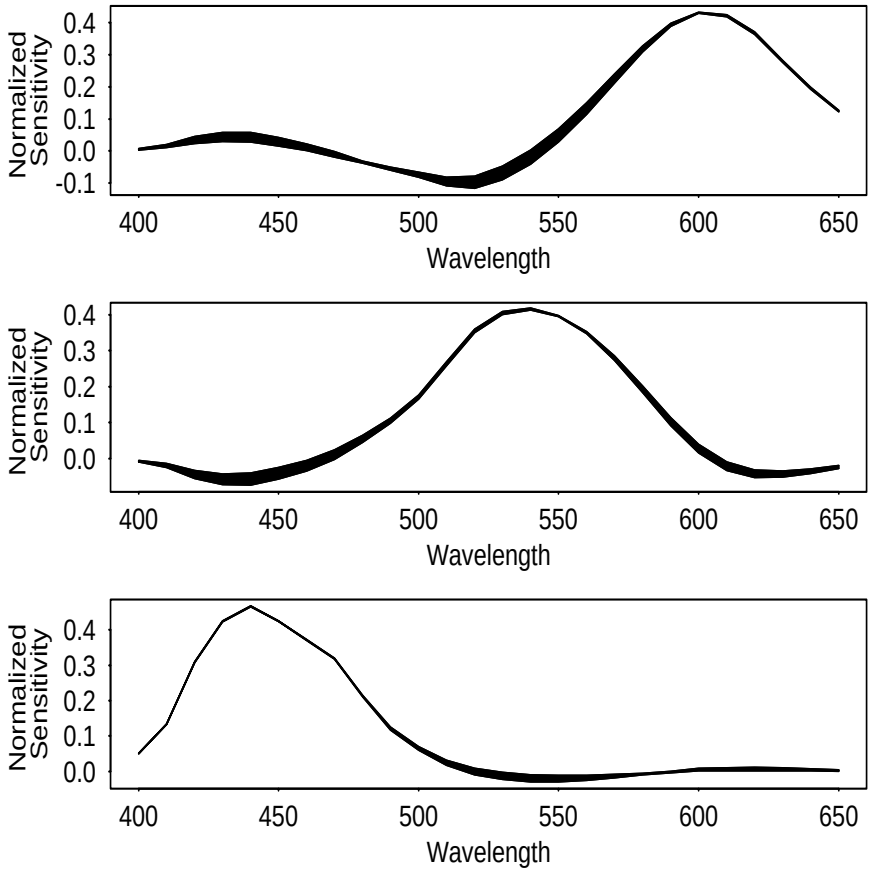

Figure 2.2: For each illuminant data-based sharpening generates different sharpened sensors. The range of sharpened curves over all the test illuminants (CIE A, D48, D65, D75 and D100) is shown for the VW cone mechanisms.

that optimizes over all the illuminants simultaneously. This issue is addressed in section 2.6, where we show that, given illuminant and reflectance spectra which are 2- and 3-dimensional, there exists a unique optimal sharpening transform.

2.4 Evaluating Sharpened Sensors

Since the sensor-based and data-based sharpened sensors are very similar, we will restrict our further attention to the evaluation of sensor-based sharpened sensors alone. As well, the results for VW and BOW sensors are very similar, so we will include figures only for the VW case.

For each illuminant, we generated sensor responses for our 462 test surface reflectances using both the sharpened and unsharpened VW sensors so that we could compare how much sharpening improves DMT performance.

To measure spectral sharpening's effect on DMT mapping of observations between a given test illuminant and the canonical illuminant, we compare *fitted observations* (observations mapped to the canonical illuminant using a diagonal matrix) with corresponding (canonical) descriptors. The Euclidean distance between a fitted observation $\underline{q}^e$ and its descriptor $\underline{p}^c$, normalized with respect to the descriptor's length, provides a good error metric given the definition of color constancy we are using. This percent normalized fitted distance (NFD) metric is defined as:

$$\text{NFD} \;=\; 100 * \frac{\| \underline{p}^c - \underline{q}^e \|}{\| \underline{p}^c \|} \tag{2.15}$$

Let $\mathcal{P}^c$ be a 3×462 matrix of descriptors corresponding to the 462 surfaces viewed under the canonical illuminant. Similarly, let $\mathcal{P}^e$ denote the 3×462 matrix of observations for the 462 surfaces viewed under a test illuminant. Equation (2.9) can then be solved to obtain the best diagonal transformation in the least-squares sense and doing so will be called *simple diagonal fitting*. Since $\mathcal{D}^{e,c}$ is a diagonal matrix, each row of $\mathcal{P}^e$ is fitted independently. The components of $\mathcal{D}^{e,c}$ are derived as follows:

$$\mathcal{D}^{e,c}_{ii} \;=\; \mathcal{P}^c_i[\mathcal{P}^e_i]^+ \;=\; \mathcal{P}^c_i[\mathcal{P}^e_i]^t / \mathcal{P}^e_i[\mathcal{P}^e_i]^t \tag{2.16}$$

where the single subscript i denotes the ith matrix row, the double subscript ii denotes the matrix element at row i column i.

Given a fixed set of sensor functions, equation (2.16) yields the best diagonal transformation that takes observations under the test illuminant onto their corresponding descriptors. Simple diagonal fitting, therefore, does not include sharpening, but rather for a fixed set of sensors finds the best diagonal matrix $\mathcal{D}^{e,c}$ for that set of sensors.

For performance comparisons we will also be interested in the NFD resulting under transformed diagonal fitting. *Transformed diagonal fitting* proceeds in two stages:

1. $\mathcal{T}\mathcal{P}^c \approx \mathcal{D}^{e,c}\mathcal{T}\mathcal{P}^e$, where $\mathcal{T}$ is the fixed sharpening transform and $\mathcal{D}^{e,c}$ is calculated via equation (2.16).

2. $\mathcal{T}^{-1}\mathcal{T}\mathcal{P}^c \approx \mathcal{T}^{-1}\mathcal{D}^{e,c}\mathcal{T}P^{e,c}$.

Applying $\mathcal{T}^{-1}$ transforms the fitted observations back to the original (unsharpened) sensor set so that an appropriate comparison can be made between the fitting errors—the performance of sharpened diagonal matrix constancy is measured relative to the original unsharpened sensors.

Figure 2.3 shows NFD cumulative histograms for:

1. diagonal fitting of VW observations (solid line).

2. diagonal fitting of sharpened VW observations (dotted line).

For each illuminant the sharpened sensors show better performance than the unsharpened ones, as indicated by the fact that in the cumulative NFD histograms the sharpened sensor values are always above those for the unsharpened ones. In general the performance difference increases the more extreme the illuminant color—from D55 to D100 the illuminants become progressively bluer and CIE A is redder.

Figure 2.4 shows the cumulative NFD histograms for

1. diagonal fitting of VW observations (solid line).

2. transformed diagonal fitting of sharpened VW observations (dashed line).

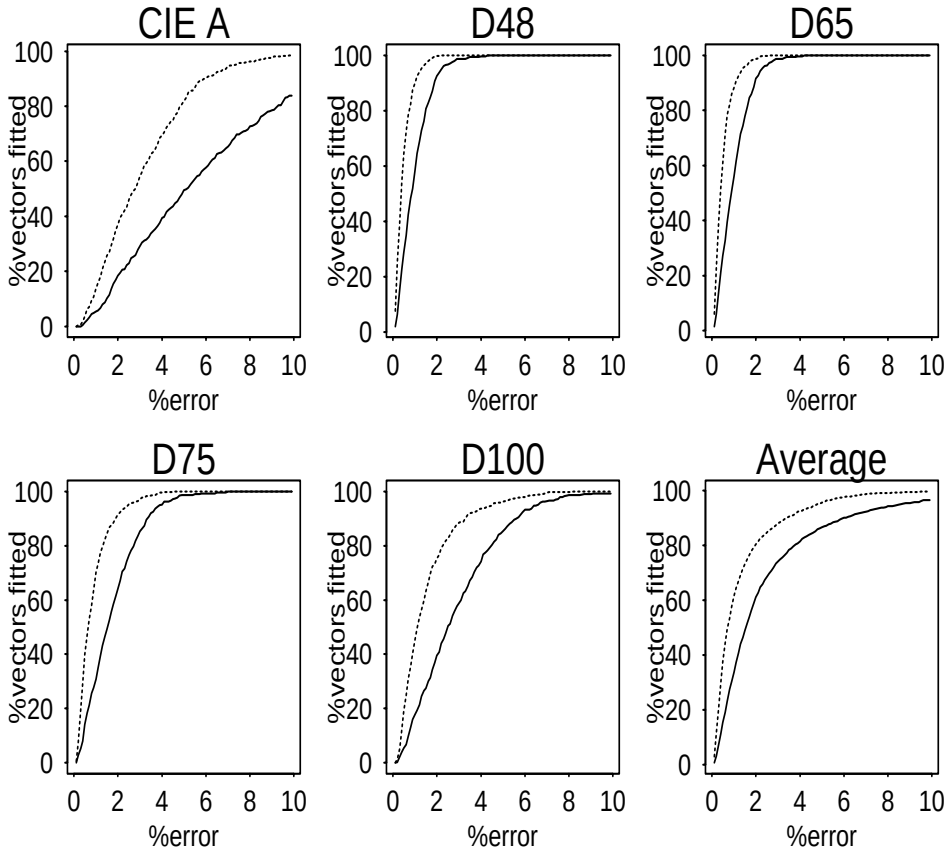

Figure 2.3: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for: diagonal fitting of VW observations (solid line) and diagonal fitting of sensor-based-sharpened VW observations (dotted line). The 6th cumulative NFD histogram shows the average fitting performance.

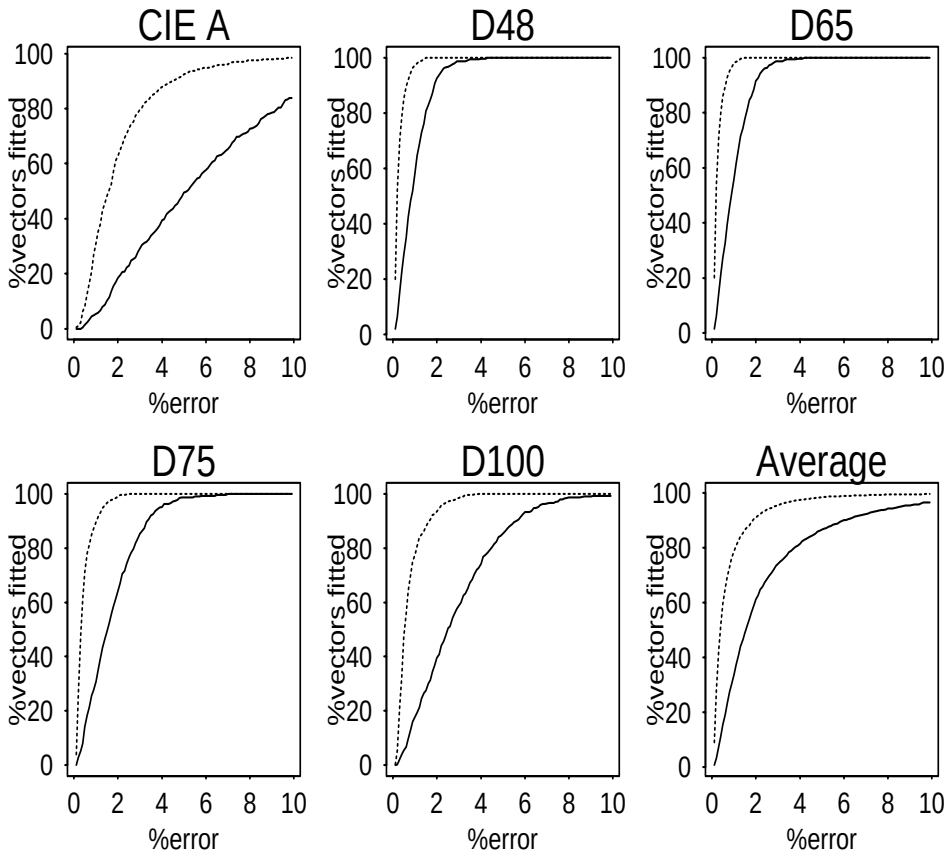

Figure 2.4: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for: diagonal fitting of VW observations (solid line) and transformed diagonal fitting of sensor-based-sharpened VW observations (dotted line). The 6th cumulative NFD histogram shows the average fitting performance.

Once again it is clear that the sharpened sensors perform better; however, the performance difference is greater, which shows that the question of sensor performance is linked to the axes in which color space is described.

Finally, Figure 2.5 contrasts the cumulative NFD histograms for

1. optimal fitting of VW observations, the unrestricted least-squares fit of equation (2.13) (solid line).

2. transformed diagonal fitting of sharpened VW observations (dashed line).

For these cases, with the exception of CIE A, a DMT achieves almost the same level of performance as the best non-diagonal transform.

2.5 Data-Based Sharpening and Volumetric Theory

Data-based sharpening is a useful tool for validating our choice of sensor-based sharpened sensors. However, more than this, data-based sharpening can also be viewed as a generalization of Brill's [Bri80] volumetric theory of color constancy. In that theory, Brill develops a method for generating illuminant invariant descriptors based on two key assumptions:

1. Surface reflectances are well modeled by a 3-dimensional basis set and are thus defined by a surface weight vector $\underline{\sigma}$. E.g., $S(\lambda) = \sum_{i=1}^{3} S_i(\lambda)\sigma_i$.

2. Each image contains 3 known reference reflectances. In the discussion that follows $\mathcal{Q}^e$ will denote the 3×3 matrix of observations for the reference patches seen under $E^e(\lambda)$.

Given the first assumption, observations for surfaces viewed under $E^e(\lambda)$ are generated by applying a *lighting matrix* [4] to surface weight vectors:

[4] This term was first used by Maloney [Mal85].

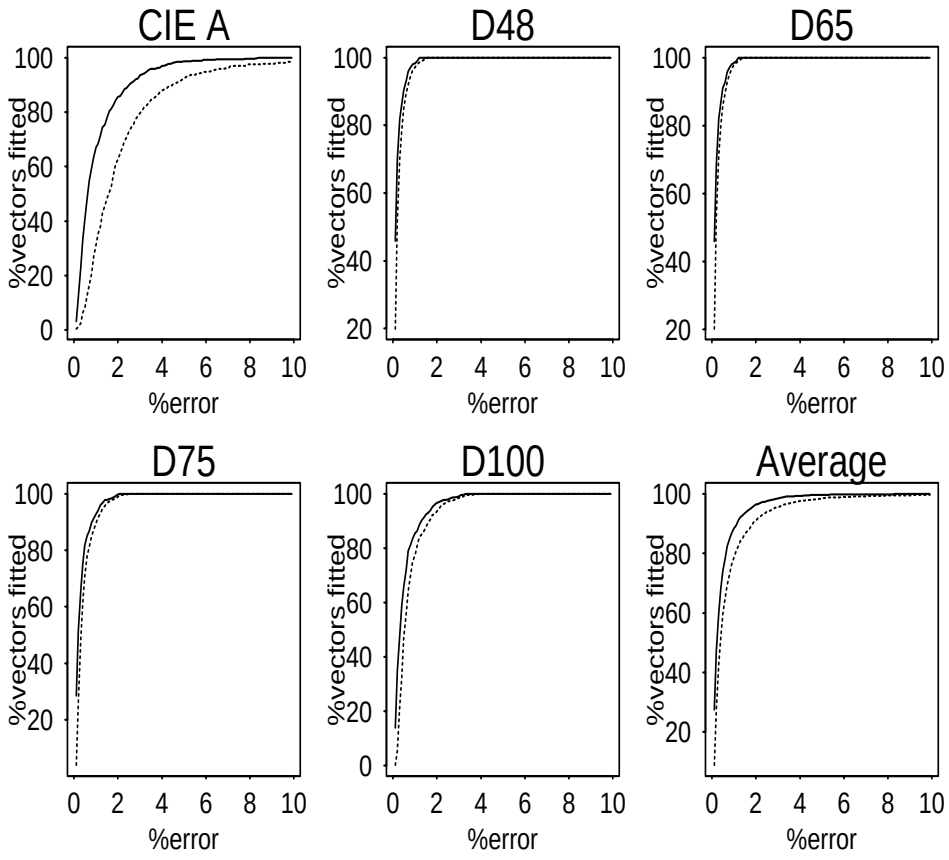

Figure 2.5: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for: optimal fitting of VW observations (solid line) and transformed diagonal fitting of sensor-based-sharpened VW observations (dotted line). The 6th cumulative NFD histogram shows the average fitting performance.

$$\underline{p^e} = \Lambda(E^e(\lambda))\underline{\sigma} \tag{2.17}$$

where the ijth entry of $\Lambda(E^e(\lambda))$ is equal to $\int R_i(\lambda)E^e(\lambda)S_j(\lambda)d\lambda$. It follows immediately that $\mathcal{Q}^e$ is a *fixed* linear transform of the lighting matrix:

$$\mathcal{Q}^e = \Lambda(E^e(\lambda))\mathcal{A} \tag{2.18}$$

where the columns of $\mathcal{A}$ correspond to the surface weight vectors of the reference reflectances and are independent of the illuminant. Now, given any arbitrary response vector p^e, an illuminant invariant descriptor is easily generated by premultiplying with $[Q^e]^{-1}$:

$$[\mathcal{Q}^e]^{-1}\underline{p^e} = \mathcal{A}^{-1}[\Lambda(E^e(\lambda))]^{-1}\Lambda(E^e(\lambda))\underline{\sigma} = \mathcal{A}^{-1}\underline{\sigma} \tag{2.19}$$

The color constancy performance of Brill's volumetric theory is directly linked to the dimensionality of surface reflectances. Real reflectance spectra are generally not 3-dimensional (Maloney [Mal86] suggests that a basis set of between 3 and 6 functions is required) and this leads to inaccuracies in the calculated descriptors.

Data-based sharpening, like volumetric theory, aims to generate illuminant invariant descriptors by applying a linear transform. If we impose the very strong constraint that all surfaces reflectances appear in each image, then data-based sharpening can be viewed as an algorithm for color constancy. As a color constancy algorithm, data-based sharpening has a distinct advantage over volumetric theory in that it is optimal with respect to the least-squares criterion and consequently is guaranteed to outperform the volumetric method in this sense. Unfortunately this performance increase is gained at the expense of the extremely strong requirement that all surface reflectances must appear in each image.

In practice, we can weaken this constraint and assume that there are k known reference patches per image, where k is small. Novak and Shafer [NS90] develop a similar theory called *supervised color constancy* based on the assumption that there are 24 known reference patches in each image; however, unlike data-based sharpening, their constancy transform is derived by examining the relationship between measured

responses and finite-dimensional models of reflectance and illumination. Certainly for the reference patches themselves, the data-based sharpening method will outperform Novak's supervised color constancy since, for these patches, data-based sharpening finds the optimal least-squares transform. However, further study is required to compare overall color constancy performance. In the limiting case, where $k = 3$, data-based sharpening reduces to the volumetric theory.

Volumetric theory requires 3 reference patches in order to recover the 9 parameters of $[\mathcal{Q}^e]^{-1}$ and thereby achieve color constancy. As shown by the performance tests of the preceding section, however, after a fixed sharpening transformation a DMT models illumination change almost as well as a non-diagonal matrix. Since only 3 parameters instead of 9 need to be determined to specify the diagonal matrix when sharpened sensors are being used, only 1 reference patch is required instead of 3 to achieve color constancy. This follows because a single response vector seen under a test illuminant $E^e(\lambda)$ can be mapped to its canonical appearance by a single diagonal matrix:

$$\underline{p}^c = \mathcal{D}^{e,c}\underline{p}^e \tag{2.20}$$

$$\mathcal{D}^{e,c}_{ii} = \frac{p^c_i}{p^e_i} \tag{2.21}$$

If we choose our reference patch to be a white reflectance then through sharpening volumetric theory reduces to Land's white-patch retinex [Lan77]. Similarly West and Brill [WB82] consider white-patch normalization to be consistent with von Kries adaptation.

We performed a simulation, called *transformed white-patch normalization*, to evaluate the quality of color constancy obtainable using a single reference patch. For each illumination (CIE A, D48, D55, D65, D75 and D100) we,

1. generated a matrix $\mathcal{P}^e$ of observations of Munsell patches for VW sensors.

2. transformed observations to the (sensor-based) sharpened sensors: $\mathcal{T}\mathcal{P}^e$.

3. calculated $\mathcal{D}^{e,c}\mathcal{T}\mathcal{P}^{e}$, where $\mathcal{D}^{e,c}_{ii} = \frac{1}{p_i^{e,w}}$ (the reciprocal of the ith sharpened sensor's response to the white patch[5]).

4. transformed white-patch normalized observations to VW sensors, $\mathcal{T}^{-1}\mathcal{D}^{e,c}\mathcal{T}\mathcal{P}^{e}$.

Again D55 was the canonical light. Thus constancy was evaluated by calculating the NFD between the white-patch-corrected observations seen under D55 (the descriptors) with the white-patch-corrected observations under each other illuminant. In Figure 2.6 we contrast the cumulative NFD histograms for white-patch normalization (dashed lines) with the optimal fitting performance (solid lines). White-patch normalization yields very good constancy results which are generally comparable to the optimal fitting performance.

2.6 Sharpening Relative to Multiple Illuminants

Data-based sharpening was introduced primarily to validate the idea of sensor-based sharpening and to ensure that our particular choice of sharpening parameters led to reasonable results. Figure 2.2 shows that the optimal sensors, as determined via data-based sharpening for each of the illuminants, closely resemble one another and furthermore they resemble the unique set of sensor-based sharpened sensors as well. While the sensors are all similar, the question remains as to whether or not there might be an optimal sharpening transform for the entire illuminant set.

2.6.1 Perfect DMT Color Constancy

In [FDF93b] we derive conditions for *perfect* DMT color constancy using sharpened sensors. Because the sharpening transform applies to a whole space of illuminants, it in essence is a type of global data-based sharpening.

The theoretical result is based on finite-dimensional approximations of surface reflectance and illumination, and what is shown is that if surface reflectances are

[5]The Munsell reflectance which is closest to the uniform-white, in the least-squares sense, was chosen as the white reference patch.

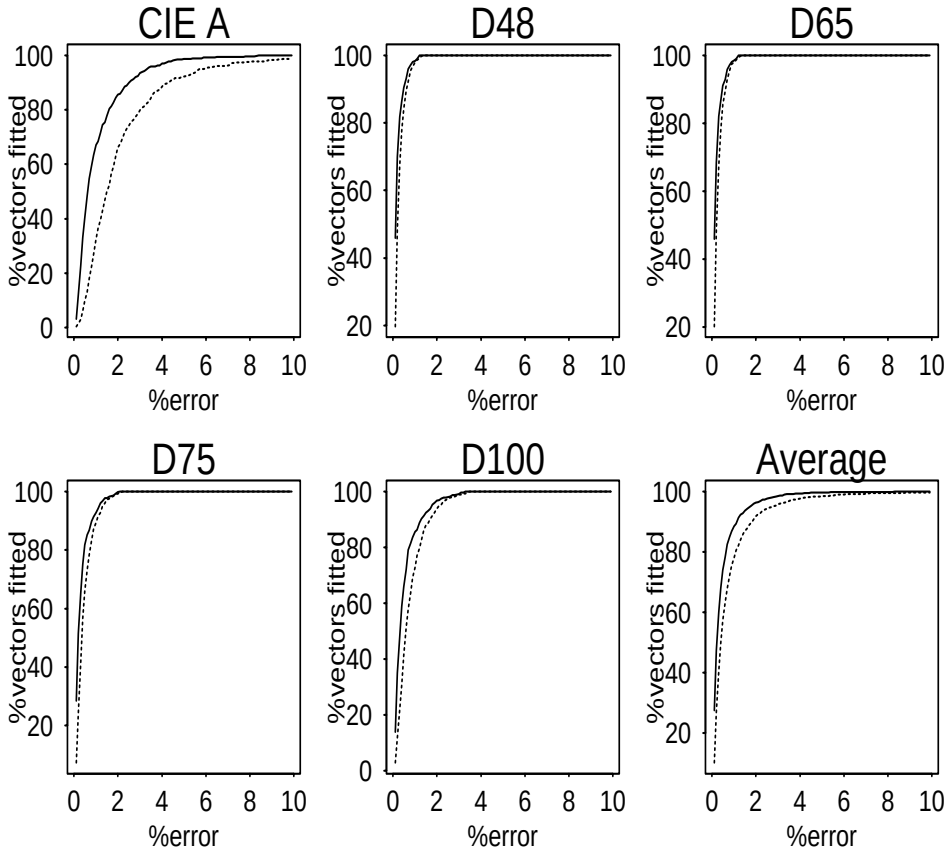

Figure 2.6: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for: an optimal fitting of VW observations (solid line) and a transformed white-patch normalization of VW observations (dotted line). The 6th cumulative NFD histogram shows average color constancy performance.

well-modelled by 3 basis functions and illuminants by 2 basis functions then there exists a set of new sensors for which a DMT can yield perfect color constancy. These restrictions are quite strong; nonetheless, statistical studies have shown that a 3-dimensional basis set provides a fair approximation to real surface reflectance[Mal86] and a 2-dimensional basis set describes daylight illumination[JMW64] reasonably well. Moreover Marimont and Wandell [MW92] have developed a method for deriving basis functions which is dependent on how reflectance, illuminant and sensors interact to form sensor responses (i.e. eqn. (2.2) is at the heart of their method). A 3-dimensional model of reflectance and a 2-dimensional model of illumination are shown to provide very good models of actual response vectors.

Given these dimensionality restrictions on reflectance and illumination, cone response vectors of surfaces viewed under a canonical illuminant, that is descriptors, can be written as:

$$\underline{p}^c = \Lambda(E^c(\lambda))\underline{\sigma} \tag{2.22}$$

where the superscript c denotes the canonical illuminant. Since illumination is 2-dimensional there is necessarily a second illuminant $E_2(\lambda)$ linearly independent with $E^c(\lambda)$ (together they form the span). Associated with this second illuminant is a second linearly independent lighting matrix $\Lambda(E_2(\lambda))$. It follows immediately that $\Lambda(E_2(\lambda))$ is some linear transform $\mathcal{M}$ away from $\Lambda(E^c(\lambda))$:

$$\Lambda(E_2(\lambda)) = \mathcal{M}\Lambda(E^c(\lambda)) \tag{2.23}$$

$$\mathcal{M} = \Lambda(E_2(\lambda))[\Lambda(E^c(\lambda))]^{-1} \tag{2.24}$$

Since every illuminant is a linear combination of $E^c(\lambda)$ and $E_2(\lambda)$, lighting matrices in turn are linear combinations of $\Lambda(E^c(\lambda))$ and $\mathcal{M}\Lambda(E^c(\lambda))$, as a result of equation (2.24). Consequently an observation vector obtained for any surface under an illuminant $E^e(\lambda) = \alpha E^c(\lambda) + \beta E_2(\lambda)$ can be expressed as a linear transform of its descriptor vector:

$$\underline{p}^e = [\alpha\mathcal{I} + \beta\mathcal{M}]\Lambda^c\underline{\sigma} = [\alpha\mathcal{I} + \beta\mathcal{M}]\underline{p}^c \tag{2.25}$$

where $\mathcal{I}$ is the identity matrix. Calculating the eigenvector decomposition of $\mathcal{M}$

$$\mathcal{M} = \mathcal{T}^{-1}\mathcal{D}\mathcal{T} \tag{2.26}$$

and expressing the identity matrix $\mathcal{I}$ in terms of $\mathcal{T}$

$$\mathcal{I} = \mathcal{T}^{-1}\mathcal{I}\mathcal{T} \tag{2.27}$$

enables us to rewrite the relationship between an observation and descriptor, equation (2.25), as a diagonal transform:

$$\mathcal{T}\underline{p}^e = [\alpha\mathcal{I} + \beta\mathcal{D}]\mathcal{T}\underline{p}^c \tag{2.28}$$

Writing $\underline{p}^c$ in terms of $\underline{p}^e$ leads directly to

$$\mathcal{T}\underline{p}^c = [\alpha\mathcal{I} + \beta\mathcal{D}]^{-1}\mathcal{T}\underline{p}^e \tag{2.29}$$

The import of this last equation is that when the appropriate initial sharpening transformation $\mathcal{T}$ is applied, a diagonal transform supports perfect color constancy, subject of course to the restrictions imposed on illumination and reflectance.

These restrictions compare favorably with those employed by D'Zmura[D'Z92], who showed that given 3-dimensional reflectances and 3-dimensional illuminants, perfect color constancy can be obtained given two images of three color patches under two different illuminants, using a non-diagonal transform.

From the Munsell reflectance spectra and our six test illuminants we used principal component analysis to derive the basis vectors for reflectance and illumination. Using these vectors, lighting matrices were constructed and the sharpening transform then calculated via eqn. (2.26). The formulae for the perfect sharpened sensors are given in equations (2.30), (2.31) and (2.32) where they are contrasted with the corresponding formulae obtained with respect to sensor-based and data-based sharpening.

The symbols R, G and B denote the Vos Walraven (red, green and blue) cone mechanisms, scaled to have unit norms, and the superscripts p, s and d denotes perfect, sensor-based and data-based sharpening.

$$\begin{aligned} R^p &= 2.44R - 1.93G + 0.110B \\ R^s &= 2.46R - 1.97G + 0.075B \\ R^d &= 2.46R - 1.98G + 0.100B \end{aligned} \tag{2.30}$$

$$\begin{aligned} G^p &= 1.55G - 0.63R - 0.16B \\ G^s &= 1.58G - 0.66R - 0.12B \\ G^d &= 1.52G - 0.58R - 0.14B \end{aligned} \tag{2.31}$$

$$\begin{aligned} B^p &= 1.0B - 0.13G + 0.08R \\ B^s &= 1.0B - 0.14G + 0.09R \\ B^d &= 1.0B - 0.13G + 0.07R \end{aligned} \tag{2.32}$$

It is reassuring that the perfect sharpened sensors are almost identical to those derived via sensor-based and data-based sharpening. Therefore, even though neither the sensor-based sharpened sensors nor the data-based sharpened ones are optimized relative to a whole set of illuminants, sharpening in all cases generates sensors which are similar to those that work perfectly for a large, although restricted, class of illuminants. This theoretical result provides strong support for the hypothesis, already confirmed in part by the consistency of the data-based sharpening results, that a single sharpening transformation will work well for a reasonable range of illuminants.

2.7 Spectral Sharpening and the Human Visual System

If the human visual system employs a DMT for color constancy, our results show that we should expect it to use sharpened sensors since doing so would optimize its performance. In this section we briefly examine some of the psychophysical evidence for sharpened sensitivities.

2.7.1 Psychophysical Evidence

Sharpened sensitivities have been detected both in field- and test-sensitivity experiments (for a review of these terms see Foster [Fos84]). Sperling and Harwerth [SH71] measured the test spectral sensitivities of human subjects conditioned to a large white background and found, consistent with our sharpened sensors, sharpened peaks at 530nm and 610nm with no sharpening of the blue mechanism.

These authors propose that a linear combination of the cone responses accounts for the sharpening. They found that the sharpened red sensor can be modelled as the red cone minus a fraction of the green, and the sharpened green as the green cone minus a fraction of the red. This corresponds well with our theoretical results in that our sharpening transformations, equations (2.30), (2.31) and (2.32), involve basically red minus green and green minus red, with only a slight contribution from the blue.

More recently, Foster [Fos81] observed that field- and test-sensitivity spectra show sharpened peaks when derived in the presence of a small monochromatic auxiliary field coincident with the test field. Foster [Sne83] extended this work by performing a hybrid experiment with a white, spatially-coincident auxiliary field; and sharpened sensitivities again were found. In both cases, these experimentally determined, sharpened sensitivities agree with our theoretical results. Like Sperling, Foster [FS83] verified that the sharpened sensitivities were a linear combination of the cone sensitivities.

Krastel [HKB83] has measured spectral field sensitivities under changing illumination where, like Sperling, a white conditioning field is employed. The illumination color was changed by placing colored filters in front of the eye. The same test spectral sensitivity curve is measured under both a reddish and bluish illuminant. This suggests that the eye's sharpened mechanisms are unaffected by illumination. More recently Kalloniatis [KH90] has measured cone spectral sensitivities under white adapting fields of different intensity and found the sharpened sensitivities to be independent of the intensity of the adapting field.

Poirson and Wandell [PW90] have developed techniques for measuring the spectral sensitivity of the eye with respect to the task of color discrimination. For color

discrimination among briefly presented targets, the spectral sensitivity curve has relatively sharp peaks at 530nm and 610nm.

Although the general correspondence between our sharpened sensors and the above psychophysical results does not imply that sharpening in humans exists for the purpose of color constancy, at least the evidence that a linear combination of the cone responses is employed somewhere in the visual system lends plausibility to the idea that sharpening might be used in human color constancy processing. Conversely, since we show that sharpening could improve the performance of some color constancy methods, our results suggest a reason as to why spectral sharpening might be found in humans.

2.8 Conclusion

Spectral sharpening generates sensors which improve the performance of color constancy theories (von Kries adaptation, Land's retinex, etc.) employing diagonal matrix transformations. Data-based sharpening finds sensors which are optimal with respect to a given set of surface reflectances and illuminants. Sensor-based sharpening finds the most narrow-band sensors that can be created as a linear combination of a given set of sensors. Finally for restricted classes of illuminants and reflectance (they are constrained to be 2- and 3-dimensional) we have shown that there exists a sharpening transform with respect to which a diagonal matrix will support perfect color constancy. The sharpening transform derived via this analysis is in close agreement with the sensor-based sharpening and data-based-sharpening transforms. In all three cases sharpened sensors substantially improve the accuracy with which a DMT can model changes in illumination. The sensor-based and data-based sharpening techniques are quite general and can be applied to visual systems with greater than 3 sensors.

As with the cone sensitivities, sharpening a color camera's sensitivities can also have a significant effect. Using overlapping, broad-band filters such as Wratten #66, #52 and #38 [Eas81] could be advantageous since from an exposure standpoint they

filter out less light, but disadvantageous from a color constancy standpoint. Sharpening such filters as shown in Figure 2.7 can provide a good compromise between the competing requirements.

Spectral sharpening is not, in itself, a theory of color constancy in that it makes no statement about how to choose the coefficients of the diagonal matrix. Instead, we propose sharpening as a mechanism for improving the *theoretical* performance of DMT algorithms of color constancy regardless of how any particular algorithm might calculate the diagonal matrix coefficients to use in adjusting for an illumination change.

Since the performance of DMT algorithms increases quite significantly when sharpened sensitivities are employed, and furthermore since it then compares favorably with that of the best possible non-diagonal transform, our results suggest that if a linear transform is a central mechanism of human color constancy then after an appropriate, fixed sharpening transformation of the sensors, there is little to be gained through the use of anything more general than a DMT.

Our results lend support to DMT-based theories of color constancy in general. In addition, since spectral sharpening aids color constancy, we have advanced the hypothesis that sharpening might provide a motivation for the psychophysical finding of sharpening in the human visual system.

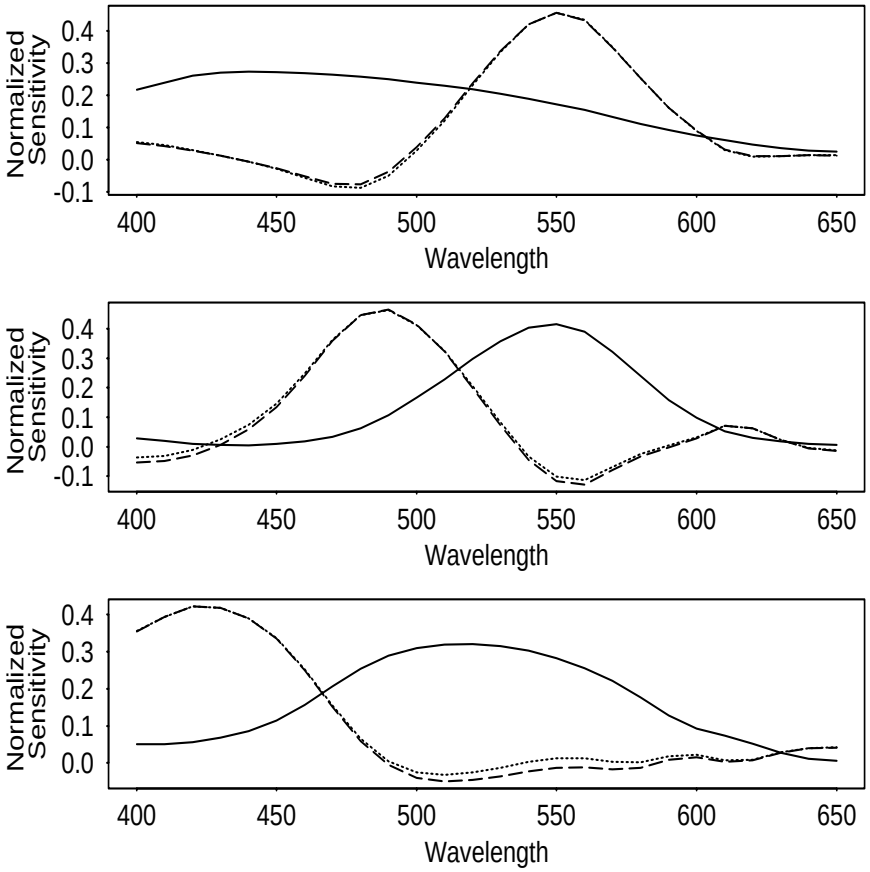

Figure 2.7: Solid lines are Kodak Wratten filters #66, #52 and #38. Dotted lines show the results of sensor-based sharpening. Dashed lines show the mean of the data-based sharpened sensors obtained for the 5 test illuminants (CIE A, D48, D65, D75, D100).

Chapter 3

Generalized Diagonal Color Constancy

G.D. Finlayson, M.S. Drew and B.V. Funt. Color Constancy: Generalized Diagonal Transforms Suffice. *The Journal of the Optical Society of America A* 11:3011-3020, 1994.

Part of the work described in this chapter first appeared in: G.D. Finlayson, M.S. Drew and B.V. Funt. Diagonal Transforms Suffice for Color Constancy. In *Proceedings of the Fourth International Conference on Computer Vision.* IEEE Computer Society & European Vision Society: 164-171, 1993.

Abstract

This chapter's main result is to show that under the conditions imposed by the Maloney-Wandell color constancy algorithm, illuminants are 3-dimensional and reflectances 2-dimensional (the 3-2 world), color constancy can be expressed in terms of a simple independent adjustment of the sensor responses—in other words as a von Kries adaptation type of coefficient rule algorithm—so long as the sensor space is first transformed to a new basis. A consequence of this result is that any color constancy algorithm which makes 3-2 assumptions, these include the Maloney-Wandell subspace algorithm, Forsyth's MWEXT and Funt and Drew's Lightness algorithm, must effectively calculate a simple von Kries type scaling of sensor responses; that is a diagonal matrix. Our results are strong in the sense that no constraint is placed on the initial spectral sensitivities of the sensors. In addition to purely theoretical arguments, the chapter contains results from simulations of von Kries type color constancy in which the spectra of real illuminants and reflectances along with the human cone sensitivity functions are used. The simulations demonstrate that when the cone sensor space is transformed to its new basis in the appropriate manner, a diagonal matrix supports close to optimal color constancy.

3.1 Introduction

We present a theoretical analysis connecting several color constancy theories—von Kries adaptation [WB82, FDF93b] [1] , Land's retinex [LM71], the Maloney-Wandell algorithm [MW86], Funt and Drew's lightness algorithm [FD88], Forsyth's MWEXT and CRULE [For90]—in which we prove that, if illuminants and reflectances are well-approximated by finite-dimensional models of low dimension, then under an appropriate change of basis for the sensor space, every one of these methods effectively calculates a simple *independent* adjustment of coefficients in this new space.

We caution the reader that we will use the term von Kries adaptation in a somewhat more general sense than is customary. Von Kries adaptation in the broad sense will be taken to apply to any sensor basis and not be restricted solely to the cone sensitivity functions. Specifically von Kries adaptation, with respect to any sensor, is a simple scaling; each scaling component is the reciprocal of the sensor response induced from a reference patch (usually white). For example if the sensor response for some surface reflectance is x and that of a reference patch y then the von Kries adapted response is x/y.

Generally color cameras, like the human eye, are trichromatic; hence in a color image each pixel is a 3-vector, one component per sensor channel. A color constancy algorithm maps each color vector $\underline{p}^e$ to a descriptor vector $\underline{d}$ which is independent of the illuminant. This mapping is usually considered linear—a matrix transform is applied to color vectors. Indeed, under Forsyth's formulation [For90] of the color constancy problem, the transform *must* be linear. In this chapter we provide a theoretical analysis along with simulation results demonstrating that if the transform is linear, then it need only be diagonal. In other words, a diagonal matrix transform suffices as a vehicle for color constancy. Our results are strong in the sense that they place no constraints on the initial spectral sensitivities of the visual system.

[1] Adaptation made using linear combinations of the adapted cone functions is sometimes referred to as second-site adaptation, for example see D'Zmura and Lennie [DL86]. This can be confusing, however, since second-site adaptation implies a second adaptation stage; whereas, we will use only a single adaptation stage with the difference being that the adaptation is applied to sensors derived as linear transformations of the cone sensitivity functions.

The various computational schemes for simulating color constancy apply different structural constraints to the form of the matrix transform. Many authors assume that the transform is a diagonal matrix, and in the model of Maloney and Wandell [MW86] the transform is a 2×3 projection. Only Forsyth's MWEXT [For90] algorithm places no constraints on the form of the transform. In studying color constancy algorithms, therefore, we must ask two questions:

1. Independent of the computational scheme for computing the matrix, how well in principle can a particular matrix form discount the effect of the illuminant?

2. How successful is a given color constancy algorithm in solving for the correct (or best) transform?

Our main focus in this chapter is on the first of these questions.

A diagonal matrix has long been proposed as a viable mechanism for color constancy (e.g. [WS82, LM71, For90]). However, West and Brill [WB82] and D'Zmura and Lennie [DL86] cast doubt on the suitability of diagonal matrix theories by demonstrating that for a given set of sensor sensitivities a diagonal matrix supports color constancy only for a *restricted* set of reflectance and illuminant spectra. With respect to the human cone sensors the restricted set of reflectance and illuminant spectra are statistically very different from actual measured illuminants and reflectances. Consequently the majority of recent color constancy theories discard the computational simplicity of the diagonal matrix transform for more complex matrix forms which *supposedly* can model illuminant change better.

In contrast to this trend Finlayson et al [FDF93b, FDF94b] have recently proved that diagonal matrix transforms can support perfect color constancy under small-dimensional model constraints—the illuminant space linearly spanned by a *2*-dimensional basis and the reflectance space by a *3*-dimensional basis. We term this set of constraints a *2-3* model. That analysis employs a generalization, which we will use here, of the concept of a diagonal matrix transform in which a *fixed sensor transformation* $\mathcal{T}$ is allowed prior to the application of a diagonal matrix:

$$\underline{d} = \mathcal{D}\underline{p} \qquad \text{(1: simple diagonal constancy)}$$

$$\mathcal{T}\underline{d} = \mathcal{D}\mathcal{T}\underline{p} \qquad \text{(2: generalized diagonal constancy)}$$

In the *2-3* case, given a known reference patch in each scene, the correct diagonal matrix transform can be computed to yield perfect color constancy. The elegant color constancy algorithm of Maloney et. al does not require a reference patch, but it operates under a different set of restrictions. For a trichromatic visual system these restrictions, which we will call the *3-2* restrictions, require a 3-dimensional illuminant space and a 2-dimensional reflectance space.

The main result of this chapter is to show that, in a world in which illuminants and reflectances are governed by Maloney's *3-2* restrictions, color constancy can *always be formulated as a generalized diagonal matrix transform independent of the spectral characteristics of the sensors.* In a world in which these restrictions hold only approximately, a diagonal matrix transform theory of color constancy will still do a good job.

The ramifications of this result for theories of color constancy are widespread. The most immediate implication is that the *3-2* version of Maloney's theory of color constancy is effectively a diagonal-matrix-based theory of color constancy. Specifically, in the *3-2* world, the color vectors of all surfaces viewed under any illuminant will always be a diagonal matrix transform from the color vectors of the same surfaces viewed under a fixed canonical illuminant. Finite-dimensional restrictions are also at the foundation of Funt and Drew's [FD88] color constancy algorithm. Their computational method simplifies, via our analysis, to diagonal matrix operations in the *3-2* case and as such reduces to Blake's version of the Lightness algorithm [Bla85]. Finally, our work plays a unifying role in connecting the theories of Maloney and Forsyth.

Forsyth's work on color constancy consists of two algorithms: MWEXT and the simpler CRULE. In MWEXT, color constancy proceeds by parameterizing all the possible matrices mapping the gamut of image colors into the gamut of descriptors. The more colorful the image, the smaller the set of possible mappings becomes. Unfortunately this algorithm is extraordinarily complex and, as Forsyth suggests, may not be suitable for machine vision. Restricting color constancy transforms to diagonal matrices results in Forsyth's simpler CRULE algorithm. This algorithm can be

efficiently implemented and is a suitable candidate for a machine vision implementation of color constancy. Forsyth [For90] proposes his MWEXT algorithm to solve for color constancy under the *3-2* restrictions. Our results prove that his simpler CRULE algorithm is adequate for this task.

Generalized diagonal matrix transforms also relate to the problems of color correction and color balancing. White-point mapping, a very common scheme for color balancing, is based on a von Kries style adjustment of the three sensor channels aimed at making a white patch in a scene appear white in an image. Simply adjusting white to look white does not guarantee, however, that the other colors will be correctly reproduced. By using a generalized diagonal matrix transformation instead of a simple diagonal matrix transformation much better results should be obtained.

In section 3.2 we provide the necessary definitions required to develop a mathematical model for color image formation and color constancy. In section 3.3 we develop techniques for finding the sensor transform $\mathcal{T}$ which affords perfect diagonal matrix color constancy under *3-2* restrictions. It should be noted that this analysis does not place restrictions on the possible form of the initial set of sensors. In section 3.4 we formally connect our results with other computational theories of color constancy. Finally in section 3.5 we present simulation results which evaluate the performance of generalized diagonal matrix color constancy.

3.2 The Model

The light reflected from a surface depends not only on the spectral properties of illumination and surface reflectance, but also on other confounding factors such as specularities and mutual illumination. To simplify our analysis we will, in line with many other authors, develop our theory for the simplified Mondriaan world; a Mondriaan is a planar surface composed of several, overlapping, matte (Lambertian) patches. We assume that the light striking the Mondriaan is of uniform intensity and is spectrally unchanging. In this world the only factor confounding the retrieval of surface descriptors is illumination.

Light reflected from a Mondriaan falls onto a planar array of sensors and at each

location X in the sensor array there are three different classes of sensors. The value registered by the kth sensor, p_k^X (a scalar), is equal to the integral of its response function multiplied by the incoming color signal. For convenience, we arrange the index X such that each p_k^X corresponds to a unique surface reflectance:

$$p_k^X = \int_\omega C^X(\lambda)\, R_k(\lambda)\, d\lambda \qquad \text{(3.3 Color observation)}$$

where λ is wavelength, $R_k(\lambda)$ is the response function of the kth sensor, $C^X(\lambda)$ is the color signal at X and the integral is taken over the visible spectrum ω. The color signal is the product of a single surface reflectance $S(\lambda)$ multiplied by the ambient illumination $E(\lambda)$: $C(\lambda) = E(\lambda)S(\lambda)$. Henceforth we drop the index X.

3.2.1 Finite-Dimensional Models

Illuminant spectral power distribution functions and surface spectral reflectance functions are well described by finite-dimensional models. A surface reflectance vector $S(\lambda)$ can be approximated as:

$$S(\lambda) \approx \sum_{i=1}^{d_S} S_i(\lambda)\sigma_i \tag{3.4}$$

where $S_i(\lambda)$ is a basis function and $\underline{\sigma}$ is a d_S-component column vector of weights. Maloney [Mal85] and Parkkinen et al [PHJ89b] presents evidence which suggests surface reflectances can be well modelled by a set of between 3 and 8 basis vectors. Similarly we can model illuminants with a low-dimension basis set:

$$E(\lambda) \approx \sum_{j=1}^{d_E} E_j(\lambda)\epsilon_j \tag{3.5}$$

$E_j(\lambda)$ is a basis function and $\underline{\epsilon}$ is a d_E dimensional vector of weights. Judd [JMW64] measured 605 daylight illuminants and showed they are well modelled by a set of 3 basis functions.

Basis functions are generally chosen by performing a principal component analysis of each data set (reflectances and illuminants) in isolation [Coh64, NNJ43, Mal86]. This type of analysis is weak in the sense that it does not take into account how

illuminant, reflectance and sensor interact in forming a color vector (eqn. (3.3)). Recently Marimont and Wandell [MW92] developed a method for deriving reflectance and illuminant basis functions which best model color observations —Eqn. (3.3) is the foundation for their method. They conclude that a 3-dimensional basis set for surface reflectance and a 3-dimensional basis set for illumination is sufficient to model the color observations of the 462 Munsell chips [NNJ43] under a wide range of black-body radiator illuminants.

3.2.2 Lighting and Surface Matrices

Given finite-dimensional approximations to surface reflectance, a color observation eqn. (3.3) can be rewritten as a matrix transform. A *lighting matrix* $\Lambda(\underline{\epsilon})$ maps reflectances, defined by the $\underline{\sigma}$ vector, onto a corresponding color vector:

$$\underline{p} = \Lambda(\underline{\epsilon})\underline{\sigma} \tag{3.6}$$

where $\Lambda(\underline{\epsilon})_{ij} = \int_\omega R_i(\lambda)E(\lambda)S_j(\lambda)d\lambda$. The lighting matrix is dependent on the illuminant weighting vector $\underline{\epsilon}$, with $E(\lambda)$ given by eqn. (3.5). The roles of illumination and reflectance are symmetric; we can write a color observation as a *surface matrix* transforming an epsilon vector:

$$\underline{p} = \Omega(\underline{\sigma})\underline{\epsilon} \tag{3.7}$$

where $\Omega(\underline{\sigma})_{ij} = \int_\omega R_i(\lambda)E_j(\lambda)S(\lambda)d\lambda$, with $S(\lambda)$ defined in eqn. (3.4). This symmetry is a key part of the analysis presented in section 3.3.

3.2.3 The Color Constancy Problem

The aim of any color constancy algorithm is to transform the color observation vector $\underline{p}$ to its corresponding illuminant-independent descriptor $\underline{d}$.

$$\underline{d} = \mathcal{Q}\underline{p} \tag{3.8}$$

where $\mathcal{Q}$ is a linear transform. However, there is no consistent definition for a descriptor. For example Maloney [MW86] uses the surface weight vector $\underline{\sigma}$ for the descriptor

(eqn. (3.9)); in contrast Forsyth defines a descriptor to be the observation of a surface seen under a canonical illuminant, defined by the weight vector $\underline{c}$ (eqn. (3.10)).

$$\underline{d}^M = [\Lambda(\underline{\epsilon})]^{-1}\Lambda(\underline{\epsilon})\underline{\sigma} \qquad \text{(3.9 Maloney's descriptor)}$$

$$\underline{d}^F = \Lambda(\underline{c})[\Lambda(\underline{\epsilon})]^{-1}\Lambda(\underline{\epsilon})\underline{\sigma} \qquad \text{(3.10 Forsyth's descriptor)}$$

Because each color constancy algorithm applies a linear transform to color vectors, different descriptor definitions differ only by a fixed linear transform, for example $\underline{d}^F = \Lambda(\underline{c})\underline{d}^M$. Therefore, demonstrating the adequacy of a diagonal matrix for one descriptor form demonstrates its adequacy for color constancy in general. If a diagonal matrix is the vehicle for color constancy for any given descriptor then we say that color constancy is in general a diagonal matrix problem. For example we will show that under the Maloney-Wandell *3-2* conditions a diagonal matrix is the vehicle for color constancy given Forsyth descriptors. Thus, even although the Maloney-Wandell algorithm does not explicitly calculate a diagonal matrix, by the equivalence of descriptor forms, we say that it is effectively a diagonal matrix theory of color constancy. In the analysis of section 3.3 we use Forsyth's descriptor form.

3.2.4 Illuminant Invariance

Color constancy seeks illuminant-invariant color descriptors. A closely related problem is to find illuminant-invariant *relationships* between color vectors instead. One candidate relationship is the diagonal matrix mapping between the color vectors of the two surfaces:

$$\mathcal{D}^{i,j}\underline{p}^{i,x} = \underline{p}^{j,x} \quad . \qquad (3.11)$$

Here i and j index two different surface reflectances, x indexes an illuminant, and $\mathcal{D}^{i,j}$ is a diagonal matrix. It is important to note that that $\mathcal{D}^{i,j}$ means the entire 3×3 diagonal matrix relating $\underline{p}^{i,x}$ and $\underline{p}^{j,x}$, not the ij component of a matrix $\mathcal{D}$. We refer to eqn. (3.11) as *diagonal invariance*. Diagonal invariance is sometimes referred to as *ratio invariance*, because the diagonal elements of $\mathcal{D}^{i,j}$ equal the ratios of the components of $\underline{p}^{j,x}$ over $\underline{p}^{i,x}$.

Diagonal invariance will be said to hold if for *all* illuminants x, a fixed diagonal matrix $\mathcal{D}^{i,j}$ maps the color vector for surface i to the color vector for surface j. Diagonal invariance plays a key role in the lightness computations of Horn [Hor74] and Blake [Bla85], the image segmentation work of Hurlbert [Hur86] and in the object recognition work of Funt and Finlayson [FF95]. Brill [Bri78] develops a more general theory of illuminant invariance, where the relationship between surfaces can be a general linear transform.

3.3 Diagonal Transforms and the *3-2* Case

Finlayson et al [FDF93b] proved that assuming illumination is 2-dimensional and reflectance 3-dimensional (the *2-3* case), there exists a transformed sensor basis in which a diagonal matrix supports perfect color constancy. In this section we prove the equivalent result for the *3-2* case.

Theorem 3.1 If illumination is 3-dimensional and surface reflectance 2-dimensional then there exists a sensor transform $\mathcal{T}$ for which a diagonal matrix supports perfect color constancy.

We prove Theorem 3.1 in two stages. First we demonstrate a symmetry between diagonal invariance and diagonal matrix color constancy. Then we prove the existence of a sensor transform which supports diagonal invariance.

Lemma 3.1 A diagonal matrix supports perfect color constancy *if and only if* there is diagonal invariance.

Proof. When a diagonal matrix supports perfect color constancy, illumination change is exactly modelled by a diagonal matrix.

$$\underline{p}^{i,c} = \mathcal{D}^{e,c}\underline{p}^{i,e} \; ; \; \underline{p}^{j,c} = \mathcal{D}^{e,c}\underline{p}^{j,e} \qquad (3.12: \text{ Same surfaces})$$

where i, j index surface reflectance and the diagonal matrix $\mathcal{D}^{e,c}$ maps the observation of surfaces under an arbitrary illuminant e to their observation with respect to the canonical illuminant c. Clearly we can map $\underline{p}^{i,e}$ to $\underline{p}^{j,e}$ by applying a diagonal matrix.

$$\underline{p}^{i,e} = \mathcal{D}^{i,j}\underline{p}^{j,e} \qquad (3.13)$$

Applying the color constancy transform $\mathcal{D}^{e,c}$ to both sides of equation (3.13) we see that:

$$\mathcal{D}^{e,c}\underline{p}^{i,e} \;=\; \mathcal{D}^{e,c}\mathcal{D}^{i,j}\underline{p}^{j,e} \tag{3.14}$$

Because transformation by diagonal matrices is commutative we can rewrite equation (3.14) as

$$\mathcal{D}^{e,c}\underline{p}^{i,e} \;=\; \mathcal{D}^{i,j}\mathcal{D}^{e,c}\underline{p}^{j,e} \tag{3.15}$$

Substituting equations (3.12) into equation (3.15) we see that

$$\underline{p}^{i,c} \;=\; \mathcal{D}^{i,j}\underline{p}^{j,c} \tag{3.16}$$

Equation (3.16) is a statement of diagonal invariance. The above argument is clearly symmetric—given diagonal invariance, diagonal matrix color constancy must follow. For the proof, we need only change the meaning of the superscripts in equations (3.12)-(3.16) so the first indexes the illuminant and the second reflectance ($\mathcal{D}^{e,c}$ becomes a diagonal invariant and $\mathcal{D}^{i,j}$ a color constancy transform).

Lemma 3.2 Given *3-2* restrictions, there exists a transformation of the sensor response functions for which, independent of the illuminant, color vectors are diagonally invariant.

Proof. Under the *3-2* restrictions the color observation of a reflectance $\underline{\sigma}$ under an illuminant $\underline{\epsilon}$ can be written in terms of two surface matrices. To see this, first note that matrix $\Omega(\underline{\sigma})$ in eqn. (3.7) can be decomposed into two parts. If the 2–vector $\underline{\sigma}$ has components $(\sigma_1, \sigma_2)^T$, then defining two special Ω matrices associated with the two basis directions in σ–space,

$$\Omega(1) \;\leftrightarrow\; (1,0)^T\,, \;\; \Omega(2) \;\leftrightarrow\; (0,1)^T\,,$$

we have

$$\Omega(\underline{\sigma}) \;=\; \sigma_1\Omega(1) \;+\; \sigma_2\Omega(2)\;.$$

Therefore eqn. (3.7) becomes

$$\underline{p} \;=\; \sigma_1\Omega(1)\underline{\epsilon} + \sigma_2\Omega(2)\underline{\epsilon} \tag{3.17}$$

Let us define a *canonical* surface reflectance, $\underline{s}$, and examine its relationship to the color observation of other surfaces. Without loss of generality we choose the first surface basis function as the canonical surface. The observation of the second surface basis function is an illuminant-independent, linear transform from the color observation of the canonical surface:

$$\Omega(2)\underline{\epsilon} \;=\; \mathcal{M}\Omega(1)\underline{\epsilon} \tag{3.18}$$

$$\mathcal{M} = \Omega(2)[\Omega(1)]^{-1} \tag{3.19}$$

Now we can rewrite eqn. (3.17), the general observation of arbitrary surfaces, as a fixed transform from the observation of the canonical surface.

$$\underline{p} \;=\; [\sigma_1\mathcal{I} + \sigma_2\mathcal{M}]\Omega(1)\underline{\epsilon} \tag{3.20}$$

where $\mathcal{I}$ is the identity matrix. Therefore we have shown that the observation of the canonical surface can be mapped to the observation of any other surface reflectance by applying a linear combination of the identity matrix $\mathcal{I}$ and the matrix $\mathcal{M}$. We define a *generalized diagonal transform* as a basis transformation followed by a diagonal matrix transform. That there exists a generalized diagonal transform mapping the observation of the canonical surface follows from the eigenvector decomposition of $\mathcal{M}$:

$$\mathcal{M} \;=\; \mathcal{T}^{-1}\mathcal{D}\mathcal{T} \tag{3.21}$$

We can also express the identity matrix $\mathcal{I}$ in terms of the eigenvectors of $\mathcal{M}$:

$$\mathcal{I} \;=\; \mathcal{T}^{-1}\mathcal{I}\mathcal{T} \tag{3.22}$$

Consequently we can rewrite eqn. (3.17) as a generalized diagonal matrix transform.

$$\mathcal{T}\underline{p} \;=\; [\sigma_1\mathcal{I} + \sigma_2\mathcal{D}]\mathcal{T}\Omega(1)\underline{\epsilon} \tag{3.23}$$

Equation (3.23) states that diagonal invariance holds between the *canonical* surface and all other surfaces given the fixed sensor transformation $\mathcal{T}$. In fact eqn. (3.23) implies that diagonal invariance holds between *any* two surfaces. Let i and j index two arbitrary surfaces described by 2–vectors $\underline{\sigma}^i$ and $\underline{\sigma}^j$. From eqn. (3.23), under

any illuminant, we can write $\mathcal{T}\underline{p}^i$ and $\mathcal{T}\underline{p}^j$ as *fixed* diagonal transforms of $\mathcal{T}\underline{p}^s$ (the observation of the canonical surface):

$$\mathcal{T}\underline{p}^i = [\sigma_1^i\mathcal{I} + \sigma_2^i\mathcal{D}]\mathcal{T}\underline{p}^s \tag{3.24}$$

$$\mathcal{T}\underline{p}^j = [\sigma_1^j\mathcal{I} + \sigma_2^j\mathcal{D}]\mathcal{T}\underline{p}^s \tag{3.25}$$

Clearly we can write $\mathcal{T}\underline{p}^i$ as a diagonal matrix premultiplying $\mathcal{T}\underline{p}^j$:

$$\mathcal{T}\underline{p}^i = \mathcal{D}^{i,j}\mathcal{T}\underline{p}^j \tag{3.26}$$

where

$$\mathcal{D}^{i,j} = [\sigma_1^i\mathcal{I} + \sigma_2^i\mathcal{D}][\sigma_1^j\mathcal{I} + \sigma_2^j\mathcal{D}]^{-1} \tag{3.27}$$

This completes the proof of Lemma 3.2. In the *3-2* case there exists a sensor transformation $\mathcal{T}$ with respect to which there is diagonal invariance and this invariance implies that a diagonal matrix is sufficient to support perfect color constancy (Lemma 3.1). Therefore, this also completes the proof of Theorem 3.1.

■

The crucial step in the above derivation is the eigenvector decomposition of the transform matrix $\mathcal{M}$. To relate this analysis to traditional theories of diagonal matrix color constancy we would like the eigenvalues of $\mathcal{M}$ to be real-valued. However, whether or not they are depends on the form of the surface matrices (and hence the initial sensor spectral sensitivities).

On first consideration complex eigenvalues appear problematic—e.g., transforming the sensors by a complex matrix of eigenvectors does not have a plausible physical interpretation. The problem lies in the fact that the new sensors would be partly imaginary; however, we show in the Appendix at the end of this chapter that complex eigenvalues fit seamlessly into our generalized theory of diagonal matrix color constancy.

3.4 Implications for Other Theories of Color Constancy

Under the *3-2* conditions the lighting matrices $\Lambda(\underline{\epsilon})$ are 3×2 injective maps—color vectors are linear combinations of the two column vectors of $\Lambda(\underline{\epsilon})$—and surfaces seen under a single illuminant span a plane in the 3-dimensional receptor space. Maloney and Wandell [MW86] exploit this plane constraint in their algorithm for color constancy. Maloney [Mal85] proves that each illuminant corresponds to a unique plane of response vectors. Given uniqueness Maloney and Wandell [MW86] present an algorithm which can determine the illuminant weight vector $\underline{\epsilon}$ and hence the pseudo-inverse $[\Lambda(\underline{\epsilon})]^{-1}$. Consequently the surface weight vector (or Maloney descriptor) can be recovered via equation (3.9).

We present an alternative diagonal color constancy algorithm for the *3-2* world. We solve for color constancy in terms of Forsyth descriptors, equation (3.10), and therefore explicitly solve for the diagonal matrix mapping the gamut of observed responses into the gamut of canonical responses.

3.4.1 Diagonal Color Constancy

In the *3-2* world the response vectors for surface reflectances under the canonical illuminant lie on the 'canonical plane' $\mathcal{P}^c$ (the canonical gamut). The span of the canonical plane is defined by the column vectors, $\underline{v}_1^c$ and $\underline{v}_2^c$, of a 3×2 spanning matrix V^c, and is calculated prior to the color constancy computation.

Under each other illuminant, response vectors for surfaces lie on the observed plane $\mathcal{P}^o$ (the image gamut). Our goal is to solve for the diagonal matrix mapping $\mathcal{P}^o$ to $\mathcal{P}^c$. If this mapping is unique then there is a single solution to the color constancy problem. The diagonal mapping is unique if we assume the following:

Assumption 3.1 There are 2 linearly independent surfaces in our image.

Assumption 3.2 None of the components of the surface normals $\underline{n}^c$, $\underline{n}^o$ (of planes $\mathcal{P}^c$ and $\mathcal{P}^o$) are equal to zero: $n_i^c \neq 0$ and $n_j^o \neq 0$ $(i, j = 1, 2, 3)$.

If Assumption 3.1 holds then we can solve for the spanning matrix V^o—the

columns of V^o, $\underline{v}_1^o$ and $\underline{v}_2^o$ are simply observation vectors of any two distinct surfaces. Assumption 3.2 states that the normal of the canonical or observed planes cannot lie on the x-y, x-z or y-z planes. If all planes are equally likely the probability that a plane normal has a zero component is vanishingly small. Maloney [Mal85] sets out "metameric black" and "unique light" conditions to characterize when the Maloney-Wandell [MW86] subspace algorithm can solve the color constancy problem. These conditions are captured by Assumption 3.2.

Theorem 3.2 (Uniqueness Theorem) Given Assumptions 3.1 and 3.2 the diagonal transform mapping $\mathcal{P}^o$ onto $\mathcal{P}^c$ is unique.

Lemma 3.3 The only diagonal matrices mapping the planes $\mathcal{P}^c$ and $\mathcal{P}^o$ onto themselves are the identity matrix $\mathcal{I}$ and scalar multiples of the identity matrix $\gamma\mathcal{I}$.

Proof of Lemma 3.3. Let V be a 3×2 matrix defining the span of a plane (either $\mathcal{P}^c$ or $\mathcal{P}^o$), $\underline{n}$ denote its plane normal, and $\mathcal{D}$ is a diagonal matrix. Writing Lemma 3.3 in mathematical notation, we would like to determine the conditions on $\mathcal{D}$ where:

$$[\mathcal{D}\underline{n}]^t V = [\ 0 \quad 0\ ] \tag{3.28}$$

Without loss of generality let us write the columns $\underline{v}_1$ and $\underline{v}_2$ of the spanning matrix V in terms of the surface normal $\underline{n}$:

$$\underline{v}_1 = [\ n_2 \quad -n_1 \quad 0\ ]^t \tag{3.29}$$

$$\underline{v}_2 = [\ 0 \quad n_3 \quad -n_2\ ]^t \tag{3.30}$$

It is easy to verify that $\underline{v}_1$ and $\underline{v}_2$ are linearly independent and orthogonal to $\underline{n}$. If $\mathcal{D}$ has a zero component, $\mathcal{D}_{ii} = 0$ for some i, then all vectors in the plane with spanning matrix $\mathcal{D}V$ must have a zero as their ith component. In this case a vector which has a non-zero ith component and is zero elsewhere is normal to all vectors in the span of $\mathcal{D}V$. If $\mathcal{D}$ satisfies equation (3.28) then the matrices V and $\mathcal{D}V$ span the same plane and the normal of $\mathcal{P}^c$ or $\mathcal{P}^o$ has a zero component. By Assumption 3.2 this is not the case. Consequently all components of the diagonal matrix $\mathcal{D}$ are non-zero. Writing equation (3.28) in full:

$$\begin{aligned} \mathcal{D}_{11} n_1 n_2 - \mathcal{D}_{22} n_2 n_1 &= 0 \\ \mathcal{D}_{22} n_2 n_3 - \mathcal{D}_{33} n_3 n_2 &= 0 \end{aligned} \tag{3.31}$$

Because none of the variables ($\mathcal{D}_{ii}$, n_j $i,j = 1,2,3$) are equal to zero, these equations are satisfied if and only if $\mathcal{D} = \gamma\mathcal{I}$.

■

Proof of Theorem 3.2. Let us assume that there are two diagonal matrices, $\mathcal{D}^1$ and $\mathcal{D}^2$, which differ by more than a simple scaling, mapping $\mathcal{P}^o$ onto $\mathcal{P}^c$:

$$\mathcal{D}^1 V^o = V^c A_1 \tag{3.32}$$

$$\mathcal{D}^2 V^o = V^c A_2 \tag{3.33}$$

where A_1 and A_2 are 2×2 matrices (of full rank) transforming the span V^c. Both the matrices $\mathcal{D}^1$ and $\mathcal{D}^2$ must have full rank, otherwise the normal to V^c contains a zero component. Solving for V^o in eqn. (3.32) and substituting into eqn. (3.33) we see that

$$\mathcal{D}^2 [\mathcal{D}^1]^{-1} V^c A_1 = V^c A_2 \tag{3.34}$$

By Lemma 3.3, only the identity matrix maps the canonical plane onto itself. Hence $\mathcal{D}^1 = \gamma\mathcal{D}^2$ (where γ is a scalar), which contradicts our initial assumption; thus Theorem 3.2 follows.

■

The truth of Theorem 3.2 depends on Assumption 3.2 holding. Let us assume, without loss of generality, that the first component of the surface normal is zero $\underline{n} = [\ 0 \quad n_2 \quad n_3\ ]^t$ (Assumption 3.2 is violated). If the vector $\underline{v}$ lies on the plane orthogonal to $\underline{n}$ then its dot-product with $\underline{v}$ equals zero: $v_2 n_2 + v_3 n_3 = 0$. The vector $\underline{v}$ transformed by a diagonal matrix with components $\mathcal{D}_{11} = a$ and $\mathcal{D}_{22} = \mathcal{D}_{33} = k$ is also orthogonal to $\underline{n}$: $k v_2 n_2 + k v_3 n_3 = 0$. In this case it follows that Lemma 3.3 and Theorem 3.2 are no longer true.

If the diagonal matrix mapping $\mathcal{P}^o$ to $\mathcal{P}^c$ is unique it is easily determined. The first spanning vector of V^o, $\underline{v}_1^o$, can be mapped onto $\mathcal{P}^c$ by applying a linear combination of two diagonal matrices

$$[\alpha\mathcal{D}^{11} + \beta\mathcal{D}^{12}]\underline{v}_1^o = \alpha\underline{v}_1^c + \beta\underline{v}_2^c \tag{3.35}$$

Similarly $\underline{v}_2^o$ can be mapped onto $\mathcal{P}^c$ by applying linear combinations of the diagonal matrices $\mathcal{D}^{21}$ and $\mathcal{D}^{22}$. Because the diagonal matrix mapping $\mathcal{P}^o$ to $\mathcal{P}^c$ is unique, the set of diagonal matrices spanned by $\mathcal{D}^{11}$ and $\mathcal{D}^{12}$ must intersect those spanned by $\mathcal{D}^{21}$ and $\mathcal{D}^{22}$ in a unique diagonal matrix.

We can use this property to develop a simple algorithm, called S-CRULE (*simplified* CRULE), for color constancy. The algorithm requires two distinct colors in the image. It proceeds in 3 stages:

1. Find the set D_1 of diagonal matrices mapping the first image color to the set of all canonical colors.

2. Find the set D_2 of diagonal matrices mapping a second image color to the set of all canonical colors.

3. The unique diagonal matrix mapping all image colors to their observation under the canonical illuminant is equal to $D_1 \bigcap D_2$.

This algorithm is closely related to Forsyth's CRULE [For90]. One difference, however, is that through our analysis we can solve for the unique diagonal matrix by examining the color observations of two surfaces. In contrast CRULE would examine all observed response vectors. If we relax the *3-2* model restrictions both Maloney's algorithm and S-CRULE cannot solve the color constancy problem.

However Forsyth's general CRULE can achieve color constancy even when the *3-2* conditions are relaxed. For example if a vision system has very narrow-band sensors then a diagonal matrix is always, without any restriction on illuminant or reflectance spectra, a perfect vehicle for color constancy [For90]. Therefore, for trichromatic color

constancy Maloney's theory is a sub-theory of Forsyth's CRULE. Previously Forsyth had proposed that his more complex MWEXT algorithm would be required to solve for color constancy under *3-2* conditions.

3.4.2 Other Theories

The color constancy problem is made more difficult if the illuminant intensity varies across the image. Horn [Hor74] presented an algorithm for removing intensity gradients from images of a Mondriaan world. Unfortunately his approach imposed strong constraints on the form of the Mondriaan boundary. Later Blake [Bla85] extended this algorithm to allow less restrictive boundary constraints. Key to their algorithms is diagonal invariance, and hence diagonal matrix color constancy. Therefore lightness recovery can be improved by the addition of a sensor transformation.

Funt and Drew [FD88] presented a non-diagonal lightness algorithm for illuminants and reflectances that are well-approximated by finite-dimensional models. Their method is independent of the sensor spectral sensitivities; however, we showed in section 3.3 that diagonal invariance holds for arbitrary spectral sensitivity functions under an appropriate sensor transformation. Our analysis therefore circumvents the need for a non-diagonal lightness theory—Funt and Drew's algorithm reduces to Blake's algorithm under a sensor transformation in the case of *3-2* world conditions.

Land's Retinex theory [Lan77] and its precursor, von Kries adaptation [WB82], assume that color constancy is achieved if each image contains a known reference patch. By assuming diagonal invariance between the observations of arbitrary surfaces with the observation of the reference patch, implies that a diagonal matrix supports color constancy. Diagonal invariance holds for all sensor sets given *3-2* (and *2-3* [FDF93b, FDF94b]) constraints.

Video cameras cannot account for changing illumination. Consequently, images taken under different illuminants must be *balanced* before display to a human observer. This balancing usually takes the form of a simple scaling in each color channel—the color video image is transformed by a diagonal matrix. To ensure illumination

change is successfully corrected, video cameras are normally equipped with narrow-band sensors. The results in this chapter indicate that a diagonal matrix transform is a suitable balancing technique independent of the sensor sensitivities used—broad-band sensors are as suitable a choice as narrow-band sensors.

3.5 Experimental Results

Real illuminants are not 3-dimensional and real surfaces are not 2-dimensional—the *3-2* conditions only provide an approximation of actual color observations—and hence a diagonal matrix can achieve only approximate color constancy. Here we perform simulations, using measured surface reflectances and measured illuminants, comparing the performance of diagonal matrix and generalized diagonal matrix color constancy. We employ the analysis of section 3.3 to derive the generalized diagonal matrix transform but test its efficacy using a von Kries type algorithm. The von Kries algorithm has the advantage that it does not restrict response vectors to lie on a plane.

The color observations of surfaces viewed under different illuminants are generated using eqn. (3.3). The human cone responses measured by Vos and Walraven [VW71] are used as our sensors, the 462 Munsell Spectra [NNJ43] for surfaces and the 5 Judd Daylight phases [JMW64](D48, D55, D65, D75 and D100) and CIE A [WS82] for illuminants. All spectra are sampled at 10nm (nanometer) intervals from 400 to 650nm. Consequently the integral of eqn. (3.3) is approximated as a summation.

The sensor transformation $\mathcal{T}$ was calculated via the technique outlined in section 3.3. Singular value decompositions of the Munsell and illuminant spectra were performed to derive the required surface and illuminant basis functions. Figure 3.1 displays the cone functions before and after the fixed sensor transformation $\mathcal{T}$. Notice that the transformed sensors appear more narrow-band—this is consistent with the pragmatic observation that narrow-band sensors afford better diagonal matrix color constancy. A similar narrowing has been observed in various psychophysical experiments [FS83, HKB83, SH71, KH90, PW90, FDF94b] involving the human visual system.

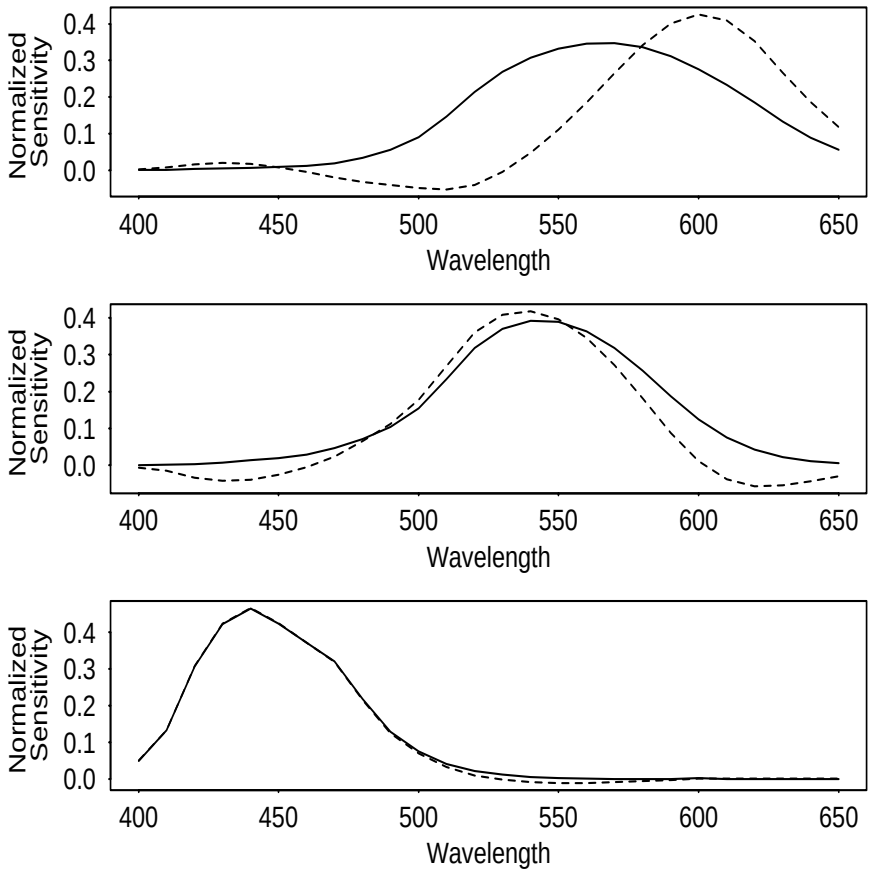

Figure 3.1: Result of sensor transformation $\mathcal{T}$. Solid lines: Vos–Walraven cone fundamentals; dashed lines: transformed sensors.

In Figure 3.2, we contrast the transformed sensors derived assuming *3-2* conditions with those assuming *2-3* conditions [FDF93b]. Both these sensor sets are remarkably similar to each other. This similarity is not altogether surprising—both the illuminant and reflectance basis sets are statistically quite similar.

There are many algorithms for diagonal matrix color constancy; each differs in its strategy for determining the diagonal matrix. Here we present simulation results for von Kries adaptation—or *white patch normalization*. The starting point for that algorithm is diagonal invariance. A color vector $\underline{p}_i$ is assumed to be diagonally invariant to the observation of a white patch $\underline{p}_w$.

$$\underline{p}^i = \mathcal{D}^{i,w}\underline{p}^w \tag{3.36}$$

Hence it is the diagonal matrix $\mathcal{D}^{i,w}$ which is independent of the illuminant, and consequently can be used as a descriptor. Usually $\mathcal{D}^{i,w}$ is written in vector (or descriptor) form $\underline{d}^{iw}$ where $d_k^{iw} = \frac{p_k^i}{p_k^w}$. By the symmetry between diagonal matrix color constancy and diagonal invariance we can rewrite eqn. (3.36) as a color constancy transform.

$$\underline{d}^{iw} = [diag(\underline{p}^w)]^{-1}\underline{p}^i \tag{3.37}$$

where the function *diag* converts the vector $\underline{p}^w$ to a diagonal matrix (diagonal elements correspond to the rows of $\underline{p}^w$). Arbitrarily we chose the white patch descriptor vectors calculated for D55 as the *canonical descriptor vectors*—these provide a reference for determining color constancy performance. Under each of the other 5 illuminants we calculate white patch descriptors. The Euclidean distance between these descriptors and their canonical counterparts, normalized with respect to the canonical descriptor's length, provides a measurement of constancy performance. The percent normalized fitted distance (NFD) metric is defined as:

$$\text{NFD} = 100 * \frac{\| \underline{d}^{iw,e} - \underline{d}^{iw,c} \|}{\| \underline{d}^{iw,c} \|} \tag{3.38}$$

where $\underline{d}^{iw,c}$ denotes a canonical descriptor and $\underline{d}^{iw,e}$ a descriptor for some other illuminant e. For each illuminant we calculated the following 3 cumulative NFD histograms:

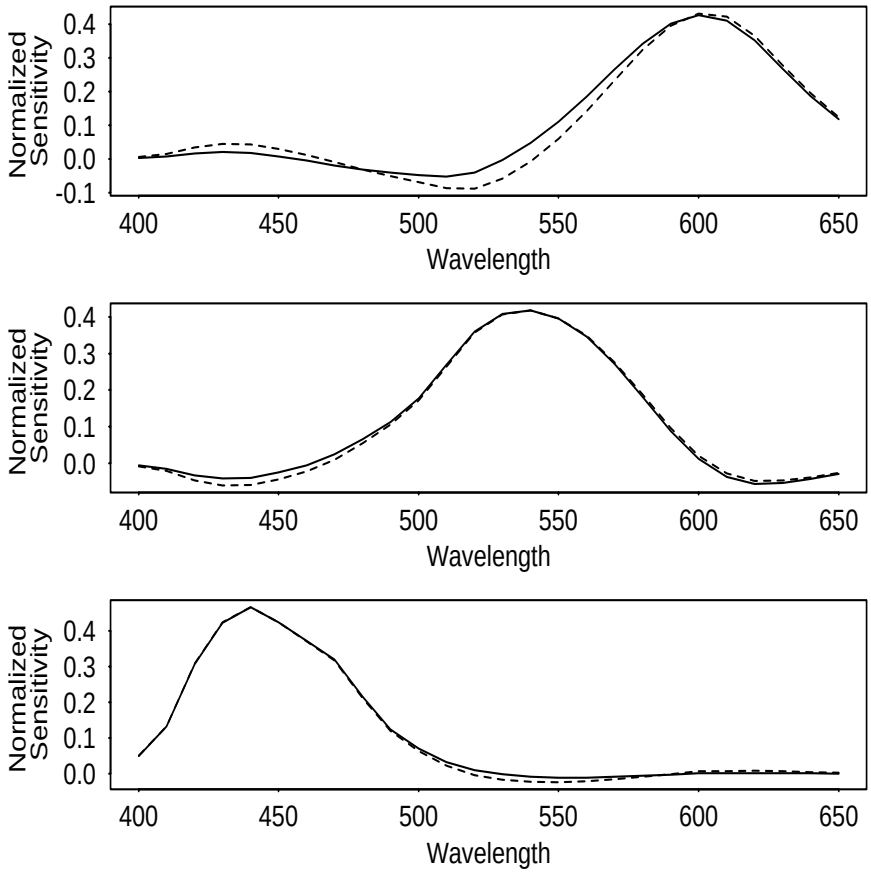

Figure 3.2: Comparison of transformed sensors derived under *3-2* and *2-3* model assumptions. Solid lines: sensors derived assuming a *3-2* world; dashed lines: sensors derived assuming *2-3* world.

1. the NFD error of white patch normalized responses for the cone functions.

2. the NFD error of generalized white patch normalized responses, generalized in the sense of $\underline{d}^{iw,e} = \mathcal{T}^{-1}[diag(\mathcal{T}\underline{p}^{w,e})]^{-1}\mathcal{T}\underline{p}^{i,e}$. The reason we apply $\mathcal{T}^{-1}$ after applying the diagonal matrix is to ensure that all our comparisons are with respect to the same sensor basis.

3. the optimal color constancy performance for a general linear transform.

We define optimal color constancy performance to be a least-squares fit relating the observations of all surfaces under an illuminant e to their corresponding observations under the canonical illuminant c. This optimal case serves as a control for evaluating the color constancy performance afforded by a diagonal matrix.

Figure 3.3 displays these 3 cumulative histograms for the test illuminants CIE A, D48, D65, D75 and D100 (dashed lines for simple white patch normalization, dotted lines for generalized white patch normalization and solid lines for the optimal constancy performance). In all cases generalized diagonal matrix color constancy outperforms, by a large margin, simple diagonal matrix constancy. Generalized diagonal matrix constancy also compares favorably with optimal color constancy. Only for the extremes in test illuminants, CIE A and to a lesser extent D100, is there a significant performance difference.

Vrhel and Trussell [VT93b] have considered the suitability of diagonal and non-diagonal matrices operating on cone responses as vehicles for color balancing. They concluded that while a non-diagonal matrix performed well a diagonal matrix (or a *white point mapping*) was inadequate for color balancing. Our results refute this conclusion.

3.6 Conclusion

A diagonal matrix is the simplest possible vehicle for color constancy. Indeed, it is its inherent simplicity which has motivated research into more complex matrix forms—if a diagonal matrix can give good color constancy a non-diagonal matrix, which has

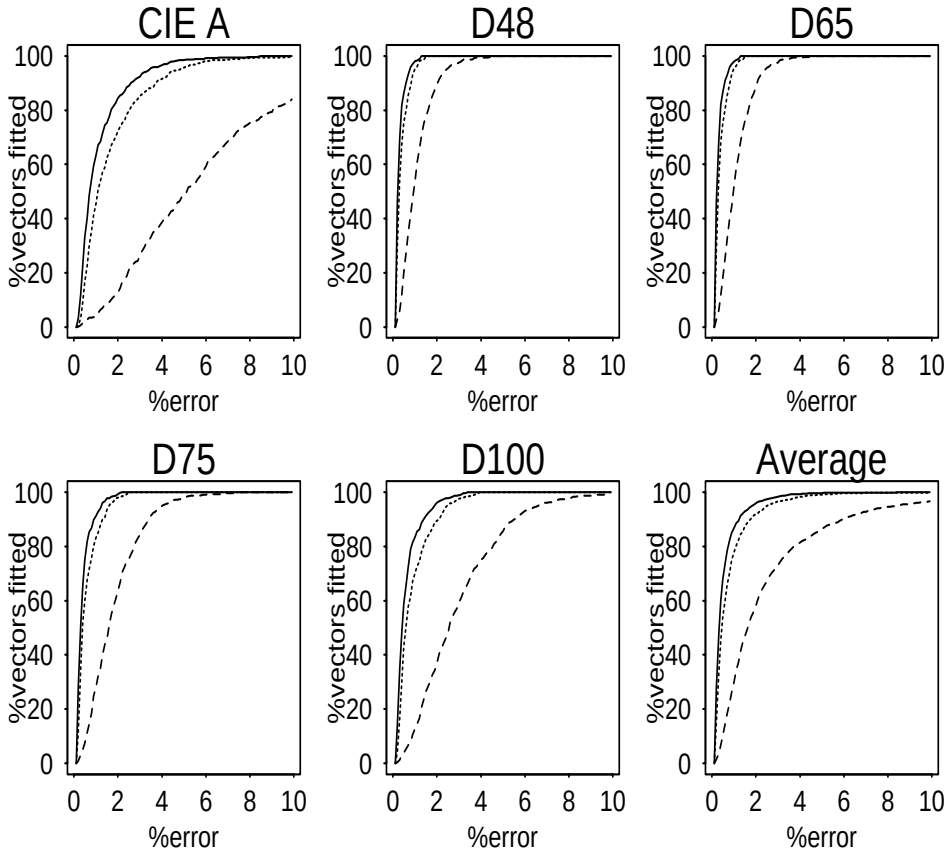

Figure 3.3: Cumulative histograms showing improved performance of generalized diagonal color constancy. Dashed lines: simple diagonal color constancy; dotted lines: generalized diagonal color constancy; solid lines: optimal (non-diagonal) color constancy.

9 instead of 3 parameters, must be able to support better color constancy, or so the reasoning goes. The analysis presented in this chapter concludes that this is in fact **not** the case. Under the Maloney-Wandell world constraints a diagonal matrix, in conjunction with an appropriate fixed transformation of the sensor basis, has been shown to suffice for the support of perfect color constancy. This result is strong in the sense that no constraints are placed on the spectral sensitivities of the sensors.

Our simulation studies investigated whether the optimal sensors as expressed in the new sensor basis derived for the *3-2* world would continue to support good color constancy when the *3-2* restrictions were relaxed. For many real reflectances imaged under real illuminants, a diagonal matrix continued to give close to optimal color constancy.

Our analysis establishes a relationship among several theories of color constancy. For a world where illumination is 3-dimensional and surface reflectance 2-dimensional, the Maloney-Wandell [MW86] algorithm, Forsyth's MWEXT [For90] and the lightness theory of Funt and Drew [FD88] are all effectively diagonal matrix theories of color constancy, since a diagonal matrix is always the vehicle for color constancy with respect to the Foryth descriptor. Moreover diagonal transforms are already at the heart of Forsyth's CRULE and von Kries adaptation. We contend these non-diagonal algorithms are more complex than necessary and can all be simplified by a fixed transformation of the sensor basis.

Appendix: Complex Eigenvalues

Complex eigenvalues may arise in the eigenvector decomposition of the transform matrix $\mathcal{M}$, but as we will show, they do not present a serious problem.

In traditional theories of diagonal matrix color constancy it is clear that each diagonal constancy transform can be expressed as the sum of three basis transforms. Indeed it is this condition which makes diagonal matrix color constancy so appealing. For example, suppose we observe the color vector $\underline{p}$ and this corresponds to the descriptor $\underline{d}$. This information is sufficient to solve for the constancy transform:

$$\underline{d} = \mathcal{D}\underline{p} \ , \ \mathcal{D}_{kk} \ = \ \frac{d_k}{p_k} \tag{3.39}$$

This same *uniqueness condition* is clearly true in generalized diagonal matrix color constancy if the sensor transformation $\mathcal{T}$ is real–valued. In fact, the uniqueness condition also holds in the general case where the elements of $\mathcal{T}$ can have complex terms.

Theorem 3.3 Under any sensor transformation $\mathcal{T}$ (where $\mathcal{T}$ can have complex elements) there are exactly 3 linearly independent diagonal matrices consistent with generalized diagonal matrix color constancy. Consequently the mapping between a color vector and its descriptor is unique.

Proof. Our original statement of diagonal matrix color constancy, eqn. (3.2), can be written in the following mathematically equivalent form:

$$\underline{d} \ = \ \mathcal{T}^{-1}\mathcal{D}\mathcal{T}\underline{p} \tag{3.40}$$

Both $\underline{d}$ and $\underline{p}$ are real-valued vectors and hence $\mathcal{T}^{-1}\mathcal{D}\mathcal{T}$ must be a real-valued matrix. Theorem 3.3 follows if we can demonstrate that there exist only 3 linearly independent, real-valued matrices with the same eigenvectors—the columns of $\mathcal{T}^{-1}$.

A diagonal matrix $\mathcal{D}$ has 6 variable components, 3 reals and 3 imaginary numbers. Consequently there are in general 6 linearly independent matrices sharing the same eigenvectors. The matrices $\mathcal{T}^{-1}\mathcal{I}\mathcal{T}$, $\mathcal{T}^{-1}\mathcal{D}\mathcal{T}$ and $\mathcal{T}^{-1}\mathcal{D}^{-1}\mathcal{T}$ are all linearly independent, real-valued matrices. Similarly $\mathcal{T}^{-1}\mathcal{I}j\mathcal{T}$, $\mathcal{T}^{-1}\mathcal{D}j\mathcal{T}$ and $\mathcal{T}^{-1}\mathcal{D}^{-1}j\mathcal{T}$ are all linearly independent, purely imaginary matrices (j is the square root of -1). The sum of imaginary numbers is always imaginary and conversely the sum of real numbers is always real; hence these 6 matrices span the set of all matrices with eigenvectors $\mathcal{T}^{-1}$. Including complex numbers in the field over which we form a span, this means that only 3 matrices form a basis for the span of all real valued matrices with eigenvectors $\mathcal{T}^{-1}$. This completes the proof for Theorem 3.3.

■

Theorem 3.3 states that generalized diagonal matrix constancy holds equally well even when the sensor transformation is complex. For any sensor transformation the

diagonal color constancy transform can be expressed as the sum of three diagonal basis matrices $\mathcal{D}$, $\mathcal{D}^{-1}$ and $\mathcal{I}$. The mapping $\mathcal{D}^{ij}$, in equation (3.13), taking $\underline{p}^{i,e}$ to $\underline{p}^{j,e}$ is still unique and is independent of the illuminant.

Chapter 4

Generalized Coefficient Color Constancy

Part of the work described in this chapter first appeared in: G.D. Finlayson and B.V. Funt. Coefficient Channels for Colour Constancy. *Proceedings of the John Dalton Conference on Colour Vision.* Manchester, Sep 1994.

Part of the work described in this chapter first appeared in: G.D. Finlayson and B.V. Funt. Optimal Spectral Sharpening. *The First Pan-Chromatic Conference.* ISCC, the Inter-Society Color Council, Williamsburg VA, February 1995.

Abstract

By hypothesis we state that there exists a set of three color channels, linear combinations of the cones, which behave exactly like a set of narrow-band sensors. If this is true then it follows that relative to the coefficient channels a von Kries type scaling (or coefficient rule) is an adequate vehicle for color constancy. We develop a computational method for finding the three color channels most consistent with the hypothesis—we call these the coefficient color channels. Simulations demonstrate that a von Kries type scaling is an excellent vehicle for color constancy relative to the new coefficient channels.

4.1 Introduction

The mechanisms for achieving color constancy are the source of much research and debate [FDF93a, DL86, MW86, For90, FDF94b, WB82, Lan77, LM71]. Here we concentrate on the *coefficient* approach to color constancy. Advocates of coefficient solutions believe that sufficient color constancy can be achieved by applying simple scaling coefficients to the sensor responses. At the outset we ally ourselves with this view and then proceed to consider the consequences. The coefficient approach in mathematical notation is as follows:

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{4.1}$$

here p_i denotes the response of the ith cone class for some surface viewed under an unknown illuminant. Each p_i is scaled by a coefficient c_i to discount the effect of the illuminant. The vector of components d_i, which we call a descriptor, describes illuminant-independent properties of the surface.

It has already been shown [DL86, WB82] that scaling coefficients applied to cone responses cannot account for illuminant change very accurately; thus equation (4.1) is a poor model for color constancy. In this chapter we ask if this failure can be mitigated by first transforming the cone responses to a new basis before applying the scaling coefficients. We call the coupling of a change of sensor basis with equation (4.1) the *generalized coefficient model of color constancy*:

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{4.2}$$

Computational methods are developed to find the *best* linear combination of the cones, the *coefficient channels*, for use in equation (4.2). We present experimental simulations which demonstrate that simple scalings applied to the coefficient channels provide excellent color constancy.

In section 4.2 we formalize the derivation of coefficient channels as a combinatorial optimization problem. The experimental performance of the derived channels is examined in section 4.3. We relate our derived coefficient channels to other theoretical studies in section 4.4. Experimental performance of the competing methods is considered in 4.5. Some conclusions are given in section 4.6.

4.2 Deriving Coefficient Channels

Our goal is to find linear combinations of the cone responses, the coefficient channels, which are *optimal* in the sense that equation (4.2) is a good model of illuminant change. It is well known [For90] that without restricting the spectral characteristics of either reflectance or illuminant spectra that equation (4.2) is a perfect model if and only if the transformed cone basis consists of 3 narrow-band sensors. Of course there does not exist a linear combination of the cones which are narrow-band. Therefore to derive coefficient channels we simply set out to find the linear combination of cones which behaves most like a set of narrow-band sensors.

4.2.1 Color response

The response of the cone with spectral sensitivity function $R_k(\lambda)$ to a reflectance $S(\lambda)$ illuminated by a spectral power distribution $E(\lambda)$ is equal to:

$$p_k = \int_\omega E(\lambda)S(\lambda)R_k(\lambda)\,d\lambda \quad (k = 1, 2, 3) \qquad \text{(4.3 Color observation)}$$

where the integral is taken over the visible spectrum ω. We call the 3-vector of cone responses, denoted $\underline{p}$, a *color observation*. The product of reflectance multiplied by the illuminant, $E(\lambda)S(\lambda)$, is called a color signal. The response of a narrow-band sensor set, $\alpha_k\delta(\lambda - \lambda_k)$ $(k = 1, 2, 3)$, will be called a *narrow observation*:

$$q_k = \int_\omega E(\lambda)S(\lambda)\alpha_k\delta(\lambda - \lambda_k)\,d\lambda = \alpha_k E(\lambda_k)S(\lambda_k) \quad (k = 1, 2, 3)$$
$$\text{(4.4 Narrow observation)}$$

The scalars α_k mediate the sensitivity of the narrow-band sensors.

Let us assume that the visible spectrum can be represented adequately by samples taken 10nm apart over the range 400 to 650nm[1]. By adopting this convention the integrals in equations (4.3) and (4.4) can be replaced by summations. Representing the response functions of the eye and narrow-band sensor sets as 26×3 matrices $\mathcal{R}$ and $\mathcal{N}$ and a color signal by a 26-vector $\underline{c}$ we can rewrite equations (4.3) and (4.4) as:

$$\underline{p} = \mathcal{R}^t \underline{c} \tag{4.5}$$

$$\underline{q} = \mathcal{N}^t \underline{c} \tag{4.6}$$

where ^{t} is the the transpose operation. Let the columns of the $26 \times n$ matrix $\mathcal{C}$ denote a set of n color signal spectra. The eye and narrow-band response to $\mathcal{C}$ are captured by the $3 \times n$ observation matrices $\mathcal{P}$ and $\mathcal{Q}$:

$$\mathcal{P} = \mathcal{R}^t \mathcal{C} \tag{4.7}$$

$$\mathcal{Q} = \mathcal{N}^t \mathcal{C} \tag{4.8}$$

4.2.2 Optimal coefficient channels

Our goal is to find the linear transform of the cones, the coefficient channels, which behaves most like a narrow-band sensor set. However, we must take care in constructing our measure of similarity. For example suppose we derived the coefficient channels which minimized the difference (maximized the similarity) defined in equation (4.9).

$$||\mathcal{T}_{i,j,k}\mathcal{P} - \mathcal{Q}_{i,j,k}||_F \quad (i,j,k \in 1,\cdots,26)\ i \neq j\ ,\ i \neq k\ ,\ j \neq k$$
$$(4.9: \quad \text{naive formulation})$$

Here (and henceforth) the subscript i,j,k denotes wavelengths i, j and k; $\mathcal{Q}_{i,j,k}$ is the observation matrix for the narrow-band sensor set with delta functions placed at i, j and k. $||.||_F$ in (4.9) denotes the Frobenius norm—the square root of the sum of squares difference between $\mathcal{T}_{i,j,k}\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$; $\mathcal{T}_{i,j,k}$ is a 3×3 matrix mapping the

[1]Sampling assumptions are routine and form the basis for the linear systems approach to color vision.

observations to the coefficient channels. Let $\mathcal{Q}_{a,b,c}$ and $\mathcal{Q}_{u,v,w}$ denote the observation matrices for two different sets of narrow-band sensors. Let us assume that:

$$||\mathcal{T}_{a,b,c}\mathcal{P} - \mathcal{Q}_{a,b,c}||_F > ||\mathcal{T}_{u,v,w}\mathcal{P} - \mathcal{Q}_{u,v,w}||_F \tag{4.10}$$

Trivially there exists a scalar γ, $0 < \gamma < 1$ such that:

$$||\mathcal{T}_{a,b,c}\mathcal{P} - \mathcal{Q}_{a,b,c}||_F < ||\gamma\mathcal{T}_{u,v,w}\mathcal{P} - \gamma\mathcal{Q}_{u,v,w}||_F \tag{4.11}$$

A more serious problem with the measure defined in equation (4.9) lies in maintaining the full dimensionality of the data set under the transform $\mathcal{T}_{i,j,k}$. Let us suppose that the rows in an observation matrix differ only by a scaling so that $\mathcal{Q}_{i,j,k}$ has a rank of only 1. In this case the matrix $\mathcal{T}_{i,j,k}$ minimizing equation (4.9) will also have rank 1 and the 3-dimensional cone responses are mapped onto a 1-dimensional subspace.

Both the problem of trivial scaling and rank deficiency result from treating the row vectors of $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ as if they were fixed quantities. Rather they are simply the sensor responses for *a particular pair of sensor bases.* A more informed measure of similarity treats all possible vectors in the row spaces of $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ as equal. A useful tool in developing this notion of similarity is the projection operator π. The projection of a matrix $\mathcal{X}$ is defined as:

$$\pi(\mathcal{X}) \;=\; \mathcal{X}^t[\mathcal{X}\mathcal{X}^t]^{-1}\mathcal{X} \tag{4.12}$$

For an arbitrary $n \times 1$ vector $\underline{v}$, the $n \times 1$ vector $\pi(\mathcal{P})\underline{v}$ is the vector in the row space of $\mathcal{P}$ which is closest in the least-squares sense to $\underline{v}$. The projection of a matrix is independent of basis: $\pi(\mathcal{X}) = \pi(\mathcal{M}\mathcal{X})$ for arbitrary matrix $\mathcal{M}$ of full rank. This is a useful property given the definition of similarity. We say that the row spaces of matrices $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ are similar if $\pi(\mathcal{P})\underline{v}$ and $\pi(\mathcal{Q}_{i,j,k})\underline{v}$ are *expected* to be similar.

Let $\mathcal{Z}$ denote an n-row matrix where each n-vector occurs with equal probability. Under the equi-probability assumption, the autocorrelation matrix of $\mathcal{Z}\mathcal{Z}^t$ is equal to the $n \times n$ identity matrix $\alpha\mathcal{I}$ [VT93a] (where α is a scalar). The difference $d(\mathcal{P}, \mathcal{Q}_{i,j,k})$ between the row spaces of $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ is defined as:

$$d(\mathcal{P},\mathcal{Q}) \;=\; ||[\pi(\mathcal{P})-\pi(\mathcal{Q})]\mathcal{Z}||_F \tag{4.13}$$

(the expected distance between $\pi(\mathcal{P})\underline{v}$ and $\pi(\mathcal{Q}_{i,j,k})\underline{v}$). Our goal is to find the three narrow-band sensors which minimize this difference. In fact equation (4.13) can be rewritten independent of $\mathcal{Z}$, using the well known identity

$$||\mathcal{M}||_F^2 \;=\; trace(\mathcal{M}\mathcal{M}^t) \tag{4.14}$$

yielding

$$d(\mathcal{P},\mathcal{Q})^2 \;=\; trace([\pi(\mathcal{P})-\pi(\mathcal{Q})]\mathcal{Z}\mathcal{Z}^t[\pi(\mathcal{P})-\pi(\mathcal{Q})]^t) \tag{4.15}$$

Since by assumption $\mathcal{Z}\mathcal{Z}^t=\mathcal{I}$, (4.15) can be rewritten as:

$$d(\mathcal{P},\mathcal{Q})^2 \;=\; trace([\pi(\mathcal{P})-\pi(\mathcal{Q})][\pi(\mathcal{P})-\pi(\mathcal{Q})]^t) \tag{4.16}$$

Using the identity in equation(4.14) in the reverse direction:

$$d(\mathcal{P},\mathcal{Q}) \;=\; ||[\pi(\mathcal{P})-\pi(\mathcal{Q})]||_F \tag{4.17}$$

For a given color signal set $\mathcal{C}$ and a given set of narrow-band sensors it is easy to calculate $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ (equations (4.7) and (4.8)). Consequently the difference (or gap [Cha93]) between the row spaces of $\mathcal{P}$ and $\mathcal{Q}$ defined in equation (4.17) can be computed. To find the optimal triplet of narrow-band sensors we must find the i, j and k which minimize

$$d(\mathcal{P},\mathcal{Q}_{i,j,k})\ (i,j,k\in 1,\cdots,26)\ i\neq j\ ,\ i\neq k\ ,\ j\neq k \qquad (4.18 \quad \text{the optimization})$$

Both matrices $\mathcal{P}$ and $\mathcal{Q}_{i,j,k}$ are $n\times n$ making equation (4.17) and the whole optimization appear computationally expensive. However we show, in the following section, that (4.17) can be evaluated by simple operations on a pair of 3×3 matrices. Thus even though there are $O(n^3)$ (where n is the number of sample wavelengths) distinct narrow band-sensor sets the optimization can be solved. When there are 26 sample points there are $\left(\begin{array}{c}26\\3\end{array}\right)=2600$ combinations.

While the minimization based on equation (4.13) does not suffer from the trivial scaling or rank deficiency problem, it does not define a cone basis which will meet our goal of behaving like a set of narrow-band sensors. Such a basis is easily determined, however. If $\mathcal{N}_{x,y,z}$ is the optimal set of narrow-band sensors then the least-squares fit mapping the cone observations to narrow observations defines the desired basis:

$$\mathcal{T}_{x,y,z}\mathcal{R}^t\mathcal{C} \approx \mathcal{N}^t_{x,y,z}\mathcal{C} \tag{4.19}$$

$$\mathcal{T}_{x,y,z} = \mathcal{N}^t_{x,y,z}\mathcal{C}\mathcal{C}^t\mathcal{R}[\mathcal{R}^t\mathcal{C}\mathcal{C}^t\mathcal{R}]^{-1} \tag{4.20}$$

where $\mathcal{C}^t\mathcal{R}[\mathcal{R}^t\mathcal{C}\mathcal{C}^t\mathcal{R}]^{-1}$ is the Moore-Penrose inverse of $\mathcal{R}^t\mathcal{C}$. The cones transformed by $\mathcal{T}_{x,y,z}$ are the *coefficient channels*.

4.2.3 Solving the optimization: the general case

The difference function $d(\mathcal{P},\mathcal{Q})$ is the square root of the sum of squared differences between the projection matrices of $\pi(\mathcal{P})$ and $\pi(\mathcal{Q})$. In examining the structure of this computation we can, without loss of generality, assume that both $\mathcal{P}$ and $\mathcal{Q}$ are orthonormal since this assumption does not alter the row-space of either matrix:

$$\mathcal{P}\mathcal{P}^t = \mathcal{I} \tag{4.21}$$

$$\mathcal{Q}\mathcal{Q}^t = \mathcal{I} \tag{4.22}$$

We will denote the sensor bases satisfying (4.21) and (4.22) as $\hat{\mathcal{R}}$ and $\hat{\mathcal{N}}$. If $\mathcal{X}$ is an orthonormal matrix its projection is written as:

$$\pi(\mathcal{X}) = \mathcal{X}^t\mathcal{X} \tag{4.23}$$

since $\mathcal{X}\mathcal{X}^t = \mathcal{I}$ in (4.12). Given orthonormality and making the roles of the color signal set $\mathcal{C}$ and sensor sets $\hat{\mathcal{R}}$ and $\hat{\mathcal{N}}$ explicit, equation (4.17) can be rewritten as:

$$d(\mathcal{P},\mathcal{Q}) = |||[\mathcal{C}^t\hat{\mathcal{R}}\hat{\mathcal{R}}^t\mathcal{C} - \mathcal{C}^t\hat{\mathcal{N}}\hat{\mathcal{N}}^t\mathcal{C}]|||_F \tag{4.24}$$

The following identities (and equation (4.14)) are useful in simplifying equation (4.24):

$$\pi(\mathcal{X})\pi(\mathcal{X}) \;=\; \pi(\mathcal{X}) \tag{4.25}$$

$$trace(\mathcal{A}\mathcal{B}^t) \;=\; trace(\mathcal{A}^t\mathcal{B}) \tag{4.26}$$

By algebraic manipulation equation (4.24) can be rewritten as:

$$d(\mathcal{P},\mathcal{Q})^2 \;=\; 6 - 2trace(\hat{\mathcal{N}}^t\mathcal{C}\mathcal{C}^t\hat{\mathcal{R}}\hat{\mathcal{R}}^t\mathcal{C}\mathcal{C}^t\hat{\mathcal{N}}) \tag{4.27}$$

The matrix $\mathcal{C}\mathcal{C}^t\hat{\mathcal{R}}\hat{\mathcal{R}}^t\mathcal{C}\mathcal{C}^t$ depends only on the sensitivities of the eye and the autocorrelation of the calibration set and can be precomputed prior to the optimization. For a particular triplet of narrow-band sensors $\mathcal{N}_{i,j,k}$, delta functions at i, j and k, we must find a 3×3 transform $\mathcal{W}_{i,j,k}$, such that the orthonormality condition of equation (4.22) is satisfied; that is, we must find $\hat{\mathcal{N}}_{i,j,k} = \mathcal{N}_{i,j,k}\mathcal{W}_{i,j,k}$

$$\mathcal{W}^t_{i,j,k}[\mathcal{N}_{i,j,k}]^t\mathcal{C}\mathcal{C}^t\mathcal{N}_{i,j,k}\mathcal{W}_{i,j,k} \;=\; \mathcal{I} \tag{4.28}$$

Because the 3×3 matrix $[\mathcal{N}_{i,j,k}]^t\mathcal{C}\mathcal{C}^t\mathcal{N}_{i,j,k}$ is symmetric positive definite it can be rewritten as [GvL83]:

$$[\mathcal{N}_{i,j,k}]^t\mathcal{C}\mathcal{C}^t\mathcal{N}_{i,j,k} \;=\; \mathcal{V}_{i,j,k}\mathcal{D}_{i,j,k}\mathcal{V}^t_{i,j,k} \tag{4.29}$$

where $\mathcal{V}_{i,j,k}$ is a 3×3 orthonormal matrix and $\mathcal{D}_{i,j,k}$ is a diagonal matrix with real diagonal components which are strictly greater than zero. Setting $\mathcal{W}_{i,j,k} = \mathcal{V}_{i,j,k}\sqrt{[\mathcal{D}_{i,j,k}]^{-1}}$ satisfies equation (4.28). The sensor set $\hat{\mathcal{N}}_{i,j,k}$ has non-zero entries at rows i, j and k but is zero elsewhere. It follows that $\hat{\mathcal{N}}_{i,j,k}$ filters out columns i, j and k of $\hat{\mathcal{R}}^t\mathcal{C}\mathcal{C}^t$ and $\hat{\mathcal{N}}^t_{i,j,k}$ filters out rows i, j and k of $\mathcal{C}\mathcal{C}^t\hat{\mathcal{R}}$. The resulting correlation structure $\hat{\mathcal{N}}^t_{i,j,k}\mathcal{C}\mathcal{C}^t\hat{\mathcal{R}}\hat{\mathcal{R}}^t\mathcal{C}\mathcal{C}^t\hat{\mathcal{N}}_{i,j,k}$ is zero everywhere except rows i, j and k at columns i, j and k. Under this filtering equation (4.27) simplifies:

$$d(\mathcal{P},\mathcal{Q})^2 \;=\; 6 - 2trace(\mathcal{W}^t_{i,j,k}\langle\mathcal{C}\mathcal{C}^t\hat{\mathcal{R}}\hat{\mathcal{R}}^t\mathcal{C}\mathcal{C}^t\rangle_{i,j,k}\mathcal{W}_{i,j,k}) \tag{4.30}$$

where $\langle \mathcal{M} \rangle_{i,j,k}$ denotes the 3×3 matrix containing the i, j and kth rows of $\mathcal{M}$ at columns i, j and k.

The cost of computing (4.24) is small—the eigenvector decomposition of (4.29) and the two matrix multiplications in (4.30)—and is independent of the size of the calibration set (save the need to calculate its autocorrelation). The optimization in equation (4.18) is tractable.

4.2.4 Solving the optimization: the maximum ignorance case

The general optimization suffers from two problems. Firstly the optimal narrow-band sensor set and coefficient channels (equation (4.19)) are dependent on the set of color signal spectra used; different sets of color signal spectra will result in different coefficient channels. Secondly, while the optimization is certainly computable it is an expensive procedure; changing the calibration set incurs a high overhead.

Let us assume that all color signal spectra, with positive and negative power, occur with equal probability—the *maximum ignorance* case. The autocorrelation of the maximum ignorance calibration set is equal to the identity matrix [VT93a]:

$$\mathcal{C}\mathcal{C}^t = \alpha\mathcal{I} \tag{4.31}$$

Where α is a scalar. For each $\mathcal{N}_{i,j,k}$ we must find a basis transform $\mathcal{W}_{i,j,k}$ such that the orthonormality condition of equation (4.28) is satisfied:

$$\mathcal{Q}\mathcal{Q}^t = \mathcal{W}_{i,j,k}^t[\mathcal{N}_{i,j,k}]^t\mathcal{C}\mathcal{C}^t\mathcal{N}_{i,j,k}\mathcal{W}_{i,j,k} = \mathcal{W}_{i,j,k}^t[\mathcal{N}_{i,j,k}]^t\mathcal{N}_{i,j,k}\mathcal{W}_{i,j,k} = \mathcal{I} \tag{4.32}$$

With a maximum ignorance calibration set $\mathcal{W}_{i,j,k} = \mathcal{I}$ satisfies equation (4.32) (for all triplets i, j and k ($[\mathcal{N}_{i,j,k}]^t\mathcal{N}_{i,j,k}$ always equals the identity matrix $\mathcal{I}$). Equation (4.30) simplifies to:

$$d(\mathcal{P}, \mathcal{Q}_{i,j,k})^2 = 6 - 2trace(\langle \hat{\mathcal{R}}\hat{\mathcal{R}}^t \rangle_{i,j,k}) = 6 - 2trace(\langle \pi(\mathcal{R}) \rangle_{i,j,k}) \tag{4.33}$$

The optimization in equation (4.18) reduces to finding the 3 largest diagonal elements of the projection matrix of the cone sensors. As a consequence the general optimization, which was combinatorial in nature, reduces to finding the 3 largest diagonal elements of the projection matrix of the cone sensors. These are easily found by sweeping through the diagonal three times. In Figure 4.1 we plot the square-root of the diagonal elements of the projection matrix against sample wavelength; this function is sometimes called the Rmag function.

Though there are three clear peaks in the Rmag function, at 450nm, 540nm and 610nm; the 3 maximum values actually lie at 440nm, 450nm and 460nm. Intuitively we know that by choosing 440nm, 450nm and 460nm as positions for our narrow-band sensors we run the risk of a dimension loss in our data—we might expect (for a set of real color signal spectra) that the transformed cone response matrix, equation (4.19), to be rank reduced. Thus, while the maximum ignorance structure is useful for analysis it does not really reflect the true covariance structure of real color signal spectra.

We can circumvent this problem by defining a minimum spacing between narrow-band sensors. Setting this minimum at 20nm instead of 10nm and the three optimal peaks lie at 440nm, 540nm and 600nm.

The minimum spacing can be theoretically justified. Let us first write the $C(\lambda)$ in terms of its Fourier series decomposition [KKOP66]:

$$C(\lambda) = a_0/2 + \sum_{k=1}^{\infty}(a_k \cos(k\frac{x-400}{250}2\pi) + b_k \sin(k\frac{650-x}{250}2\pi) \qquad (4.34)$$

where x is in nanometers in the range 400-650nm (400nm is the bottom of the visible spectrum which for our purposes is 250nm wide). Each k defines how many times the sin or cos function repeats within the visible spectrum. The set of coefficients a_i and b_j are uniquely defined for each $C(\lambda)$.

As k becomes large the sin and cos functions repeat many times over the visible spectrum. Because the cones are broad-band functions, very fast variation in a color signal is invisible. Buchsbaum [BG84] and Barlow [Bar82] both estimate that the eye cannot see variation greater than 6 cycles over the visible spectrum (the eye acts

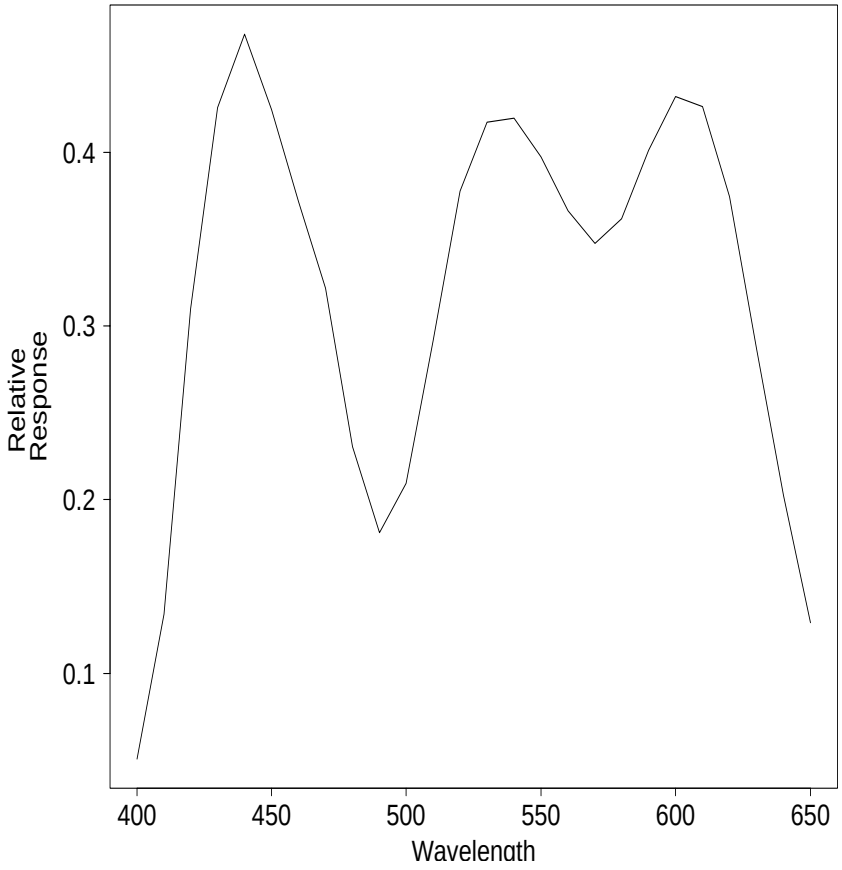

Figure 4.1: The Rmag function. If $\mathcal{R}$ denotes the 26×3 matrix of cone sensitivities and $\pi(\mathcal{R})$ its projection matrix then Rmag is the square root of the diagonal elements of this projection. Wavelengths where the values of Rmag are large are suitable places for placing narrow-band sensors—see text.

as a low-pass filter). It follows that, for the purposes of analysis, we can substitute a low-frequency counterpart of $C(\lambda)$, which we denote $C'(\lambda)$, since the eye cannot discriminate between them:

$$C'(\lambda) = a_0/2 + \sum_{k=1}^{6}(a_k \cos(k\frac{x-400}{250}2\pi) + b_k \sin(k\frac{650-x}{250}2\pi) \quad (4.35)$$

There are only 13 coefficients in equation (4.35) and these can be found if we sample $C'(\lambda)$ at 13 sample points (this observation is an incarnation of the Nyquist sampling theorem).

Because the eye acts as a low-pass filter, our maximum ignorance assumption is more general than it need be—we need only assume that all color signal spectra with a band-limit of less than or equal to 6 cycles over the visible spectrum occur with equal likelihood. In this framework any subset of 13 sample points is as good as any other in the sense it will capture all the signal information. Our intuitive notion of minimum spacing simply rules out a few of these possible subsets.

4.3 Performance of coefficient channels

Maximum ignorance and *calibration* coefficient channels were derived from the Vos and Walraven cone fundamentals [VW71] and two color signal sets: the first contains all possible color signal spectra and the second the 426 Munsell spectra [NNJ43], multiplied by 6 exemplar illuminants (the 5 Judd [JMW64] daylight phases D48, D55, D65, D75 and D100 and CIE standard illuminant A). The maximum ignorance and calibration coefficient channels are contrasted with the cones in Figure 4.2. The optimal narrow-band sensors for the maximum ignorance assumption are positioned at 440nm, 540nm and 600nm. The best wavelengths for the general optimization are positioned at: 450nm, 530nm and 610nm.

We now test the derived coefficient channels to see if they support improved color constancy. Let us suppose that the goal of color constancy is to map color observations viewed under an unknown illuminant to their corresponding descriptors observed under a fixed canonical illuminant. We will consider 3 candidate mappings: the simple

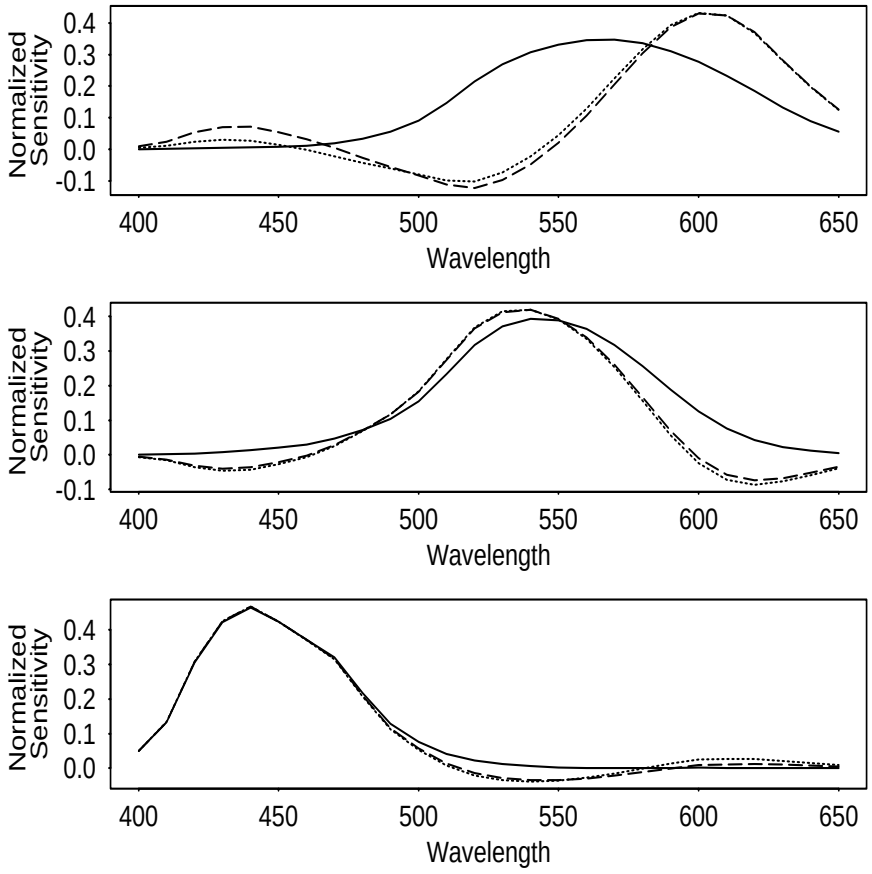

Figure 4.2: The coefficient channels derived under maximum ignorance assumptions (dotted lines) and for a calibration set of real color signal spectra (dashed lines) are contrasted with the Vos and Walraven cone fundamentals (solid lines).

coefficient model (equation (4.1)), the generalized coefficient model (equation (4.2)) and the general linear model shown below:

$$\underline{d} = \mathcal{G}\underline{p} \tag{4.36}$$

where $\mathcal{G}$ is *any* 3×3 bijective map. The general linear model must perform better than either the simple or generalized coefficient models and serves as a control for our simulations.

Under each of the 6 illuminants the cone observations of the 462 Munsell reflectances [NNJ43] are calculated. Arbitrarily we choose D55 as the canonical illuminant. For each test illuminant we find the mappings minimizing the error norms:

$$||\mathcal{P}^{\mathrm{D55}} - \mathcal{D}^{\mathrm{test,D55}}\mathcal{P}^{\mathrm{test}}||_F \qquad (4.37 \quad \textit{simple coefficient})$$

$$||\mathcal{T}\mathcal{P}^{\mathrm{D55}} - \mathcal{D}^{\mathrm{test,D55}}\mathcal{T}\mathcal{P}^{\mathrm{test}}||_F \qquad (4.38 \quad \textit{generalized coefficient}))$$

$$||\mathcal{P}^{\mathrm{D55}} - \mathcal{G}^{\mathrm{test,D55}}\mathcal{P}^{\mathrm{test}}||_F \qquad (4.39 \quad \textit{general linear})$$

where $\mathcal{P}$ denotes a 3×462 matrix of cone responses and the superscript denotes the illuminant. The superscript *test* is one of D48, D65, D75, D100 and CIE A. The diagonal matrix $\mathcal{D}^{\mathrm{test,D55}}$ in (4.37) takes cone observations to descriptors with minimum least-square error; (4.37) effectively bounds bounds the performance of a von Kries type mapping operating on cone responses. The diagonal matrix $\mathcal{D}^{\mathrm{test,D55}}$ in (4.38) takes coefficient channel observations to coefficient descriptors with minimum squared error. Because the coefficient channels and the cones are not the same it is not a fair comparison to compare the error in (4.38) with the error in (4.37). To account for this after solving for $\mathcal{D}^{\mathrm{test,D55}}$ in (4.38) we map coefficient channel descriptors and mapped coefficient observations back to the cone basis:

$$||\mathcal{T}^{-1}\mathcal{T}\mathcal{P}^{\mathrm{D55}} - \mathcal{T}^{-1}\mathcal{D}^{\mathrm{test,D55}}\mathcal{T}\mathcal{P}^{\mathrm{test}}||_F \tag{4.40}$$

Finding the best diagonal transform with respect to one sensor basis and then transforming the result to another basis for comparison has previously been called

transformed diagonal fitting [FDF94b]. Finally the general linear $\mathcal{M}$ in equation (4.39) is the 3×3 matrix taking cone observations to descriptors with least error. Equation (4.39) bounds the performance which is possible assuming a linear model of color constancy. To compare the performance of each of the three mappings (4.37), (4.39) and (4.40) we calculate the Euclidean distance between each fitted observation and its descriptor normalized with respect to the descriptors length. The percent normalized fitted distance metric is defined as:

$$\text{NFD} \;=\; 100 * \frac{||\underline{p}^{\text{D55}} - \mathcal{M}\underline{p}^{\text{test}}||}{||\underline{p}^{\text{D55}}||} \tag{4.41}$$

where $\mathcal{M}$ is either the diagonal matrix defined in (4.37), the transformed diagonal matrix defined in (4.38) and (4.40) or the general linear transform defined in (4.39).

Figure 4.3 contrasts NFD histograms obtained for the simple and generalized coefficient mappings (maximum ignorance case) with the cumulative histogram obtained for the general linear case. It is clear that a general linear transform, with 9 free parameters, greatly outperforms the simple, 3-parameter, coefficient model. The generalized coefficient model fares much better providing performance of the same order as the general case.

In Figure 4.4 we compare the generalized coefficient cumulative NFD histogram calculated for the calibration set with the simple coefficient and general linear cases. As before the 3-parameter generalized coefficient model provides close to general linear performance. While both the maximum ignorance and calibration coefficient channels enhance coefficient color constancy it is not clear how they compare to each other. In Figure 4.5 the cumulative histograms for both sets of coefficient channels are compared.

As we might expect the calibration coefficient channels generally perform better; though not by much. Indeed performance is so similar that the maximum ignorance optimization might be preferable. The reasons for this are twofold: first the computation is simple (linear in the number of sensors); and second, the answer is unique since they are derived assuming that all spectra are equally likely.

The simulations suggest that a generalized coefficient model is almost as expressive

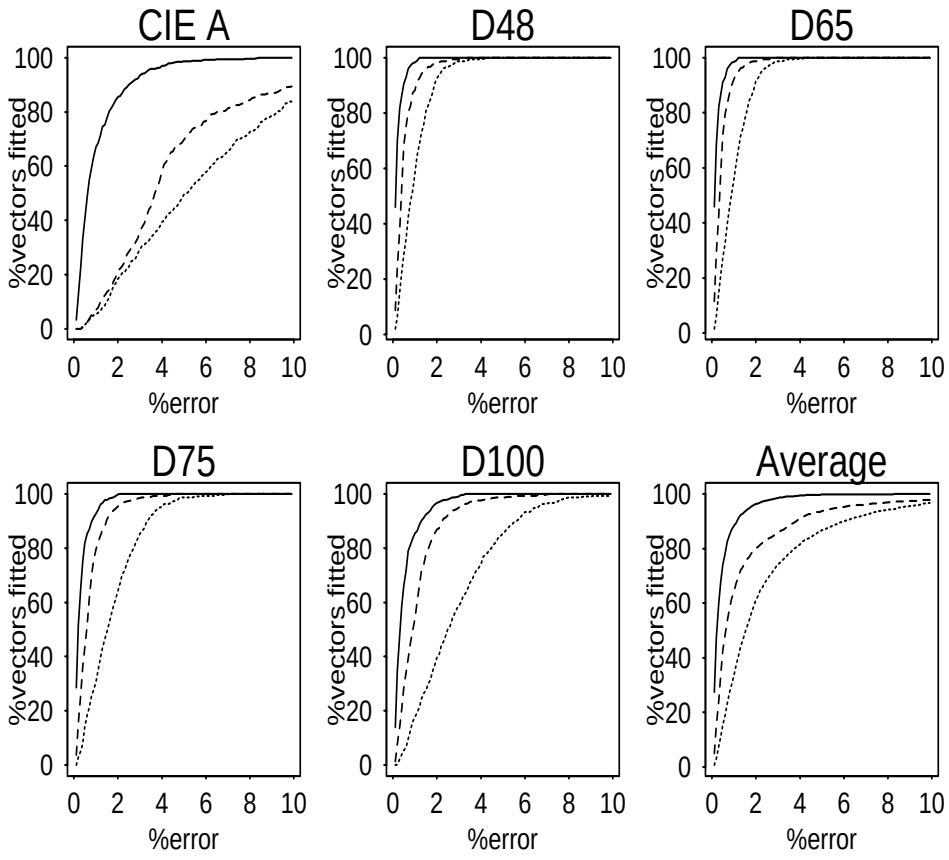

Figure 4.3: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for simple coefficient color constancy (dotted lines), generalized coefficient color constancy using maximum ignorance sensors (dashed lines) and for the general linear model (solid lines).

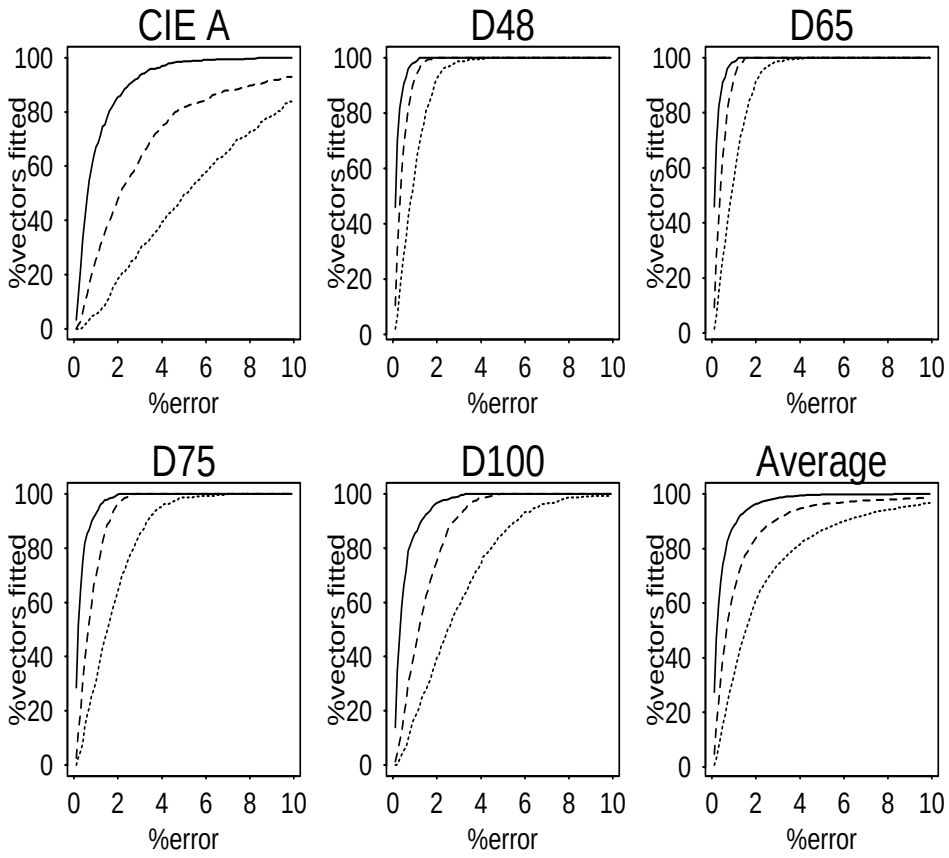

Figure 4.4: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for simple coefficient color constancy (dotted lines), generalized coefficient color constancy using calibration sensors (dashed lines) and for the general linear model (solid lines).

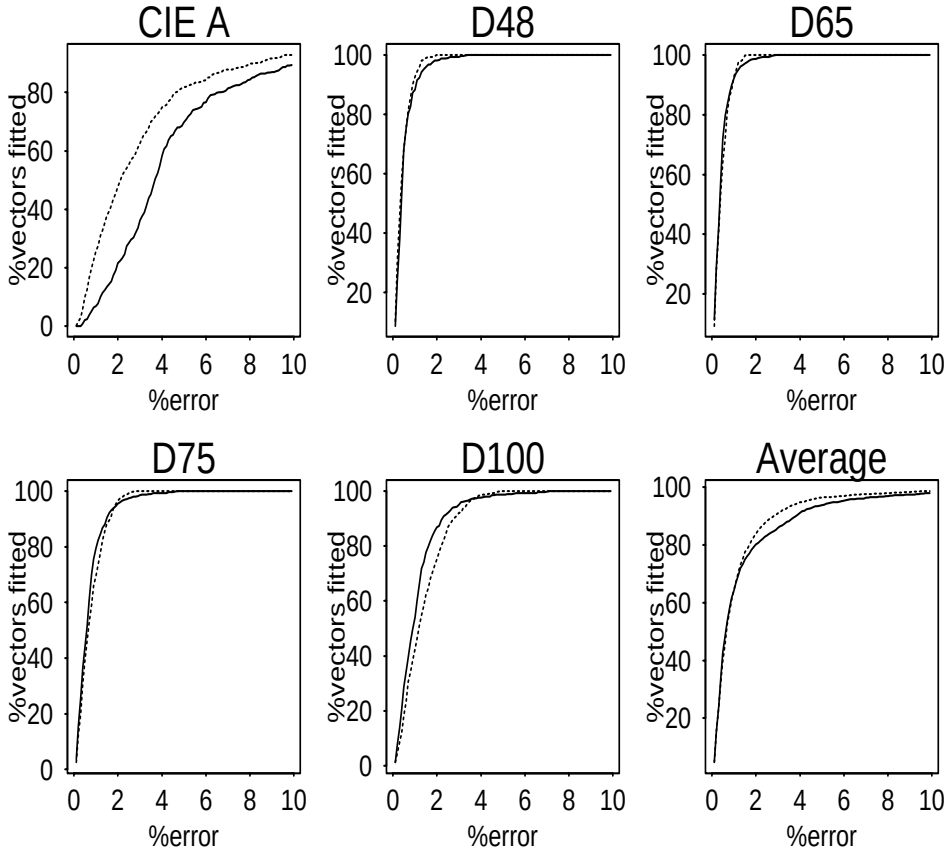

Figure 4.5: The cumulative NFD histograms obtained with each test illuminant (CIE A, D48, D65, D75 and D100) for generalized coefficient color constancy using calibration (dotted lines) and maximum ignorance (solid lines) sensors.

as the general linear model. It is parameterized by only three unknowns—the three scaling coefficients—and as consequence color constancy remains a 3-dimensional problem. In contrast color constancy, in the general linear case, is a 9-dimensional problem—since there are 9 free parameters to be determined.

4.4 Relationship to other theoretical Studies

The methods presented in section 4.2 are closely related to other theoretical work on the coefficient model. In particular sensor-based spectral sharpening [FDF94b] at a single wavelength, where the three globally sharpest sensors are sought, is equivalent to deriving coefficient channels under the maximum ignorance assumption. The derivation of coefficient channels in light of a set of real color signal spectra is closely related to the data-based sharpening method. The latter method finds the optimal coefficient channels for a particular set of reflectances viewed under a pair of illuminants.

4.4.1 Sensor-based Sharpening

Finlayson et al [FDF94b] present alternative methods for finding coefficient type channels. The starting point for their analysis is the intuition that the narrower your sensors the more appropriate the coefficient model of color constancy becomes. Sensor-based sharpening is an analytic tool for finding the linear combinations of the cones (or any sensor set) which is most sensitive to a given interval $[\lambda_1, \lambda_2]$ in the visible spectrum. Techniques are set out for maximizing:

$$||\mathcal{R}^{[\lambda_1,\lambda_2]}\underline{c}|| \quad (||\mathcal{R}\underline{c}|| = 1) \tag{4.42}$$

where $\mathcal{R}^{[\lambda_1,\lambda_2]}$ denotes the rows of $\mathcal{R}$ from sampling wavelengths λ_1 through λ_2. The vector $\underline{c}$ which maximizes $||\mathcal{R}^{[\lambda_1,\lambda_2]}\underline{c}||$ and satisfies the constraint that the resulting new sensor has unit length.

No method is given, however, for choosing intervals for sharpened sensors. Some

intervals will lead to fairly narrow sensors while others will be less successful. Under the special case of sensor-based sharpening at a single wavelength the techniques presented in this chapter provide a method for choosing sharpening intervals. Without loss of generality let us transform the cone response matrix $\mathcal{R}$ to an equivalent orthonormal set $\mathcal{R}_\perp$. We can write the projection matrix of $\mathcal{R}$, $\pi(\mathcal{R})$ (using identity (4.23)) as:

$$\pi(\mathcal{R}) \;=\; \mathcal{R}_\perp \mathcal{R}_\perp^t \tag{4.43}$$

denoting the ith row of $\mathcal{R}_\perp$ as $\underline{r}_i^t$, then the ith diagonal element of $\pi(\mathcal{R})$ is equal to:

$$\underline{r}_i^t \underline{r}_i \;=\; ||\underline{r}_i||^2 \tag{4.44}$$

Because $\mathcal{R}_\perp$ is orthonormal the constraint term in the sensor-based sharpening formulation is satisfied if and only if $\underline{c}$ is a unit vector. Under a rotation $\mathcal{O}$ the diagonal elements of $\pi(\mathcal{R})$ are unchanged and:

$$\underline{r}_i^t \mathcal{O}^t \mathcal{O} \underline{r}_i \;=\; \underline{r}_i^t \underline{r}_i \;=\; ||\underline{r}_i||^2 \tag{4.45}$$

Each column of $\mathcal{R}_\perp$ or $\mathcal{R}_\perp \mathcal{O}$ denotes a distinct sensor. At wavelength i we are creating the sharpest sensor if all of the length $||\underline{r}_i||^2$ is captured in a single column of $\mathcal{R}_\perp O$ (that is in a single sensor). This is the case when the vector $\frac{\underline{r}}{||\underline{r}||}$ appears as one of the columns of $\mathcal{O}$. The diagonal of $\pi(\mathcal{R})$ therefore is a measure of the maximum percentage squared length that any linear combination of the cones can have at a particular wavelength subject to the constraint that the new sensor over all wavelengths has unit length.

Choosing the three largest elements of the diagonal of $\pi(\mathcal{R})$ is equivalent to choosing the three wavelengths where sensor-based sharpening will be most successful. That is, our maximum ignorance optimization (eqn. (4.33)) is equivalent to finding the 3 sharpest sensors. More than simply providing candidate regions of the visible spectrum where sharpening is beneficial, our optimization work gives a theoretical underpinning to the intuition of sharpening. If we know nothing about the statistics

of real color signal spectra then the sharpest sensors are a good choice for coefficient channels.

4.4.2 Data-based sharpening

Data-based sharpening is a method for finding the optimal coefficient channels given only the canonical and a **single** test illuminant. For each test illuminant Finlayson et al [FDF94b] show that there exists a set of coefficient channels formed by a basis change $\mathcal{T}^{\text{test,D55}}$ such that the general linear matrix $\mathcal{G}^{\text{test,D55}}$ minimizing (4.39) is diagonal: $\mathcal{M}^{\text{test,D55}} = \mathcal{D}^{\text{test,D55}}$.

$$||\mathcal{T}^{\text{test,D55}}\mathcal{P}^{\text{D55}} - \mathcal{D}^{\text{test,D55}}\mathcal{T}^{\text{test,D55}}\mathcal{P}^{\text{test}}||_F \;=\; ||\mathcal{T}^{\text{test,D55}}\mathcal{P}^{\text{D55}} - \mathcal{G}^{\text{test,D55}}\mathcal{T}^{\text{test,D55}}\mathcal{P}^{\text{test}}||_F \quad (4.46)$$

The sharpening transform depends on the test illuminant; different illuminants result in different transforms. However Finlayson et al [FDF93b, FDF94b] showed that if illuminant and reflectance spectra are exactly characterized by 2 and 3 parameter finite-dimensional models then the data-based sharpening transform is the same for all test illuminants. The same was shown true for the complementary case of 2-dimensional reflectances and 3-dimensional illuminants [FDF93a] (the restrictions assumed by the trichromatic Maloney-Wandell subspace algorithm for color constancy [MW86]).

The general optimization presented here is a method for finding a single fixed data-based sharpening transform for many test illuminants without placing any restriction on the dimensionality of reflectance and illuminant spectra. The solution for the multi-test-illuminant case is however bought at a cost: the requirement that coefficient channels behave exactly like narrow-band sensors. While we can certainly argue that this requirement is reasonable, it is not assumed in the pairwise data-based sharpening method.

4.4.3 Perfect Sharpening

Perfect sharpening is a method for choosing coefficient channels based on finite-dimensional descriptions of illumination and reflectance. Specifically, Finlayson et. al have shown that if illumination and reflectance are precisely described by 2- and 3-dimensional linear models [FDF94b, FDF93b] respectively (or the reverse [FDF93a, FDF94a]) then the generalized coefficient model supports perfect color constancy. Moreover this results holds for all sensors sets and is not contingent on any special property of the cone basis.

The central argument in perfect sharpening is algebraic in nature and does not hinge directly or indirectly on the observation that narrow-band sensors support perfect color constancy. Indeed any choice of sensor basis (e.g the cones) can be coefficient channels when appropriate illumination and reflectance models [WB82] are chosen. As such, perfect sharpening relates to the current study only insofar as it returns coefficient channels similar to those derived here.

4.4.4 Experimental Comparison

Generalized coefficient color constancy, generalized diagonal color constancy [FDF94a], spectral sharpening, data-based sharpening and perfect sharpening [FDF94b] are all methods for choosing a sensor basis with respect to which simple scaling coefficients take observations between illuminants. Which of these methods is best? The answer to this question clearly depends on the statistical properties of the reflectances, illuminants and sensors.

In Figure 4.6 we compare the average mapping performance, calculated for the 462 Munsell chips over the 5 test illuminants, for each method. We begin, as before, by plotting the NFD histogram curve obtained when scaling coefficients are applied to the cone basis. This curve is (labelled by a 1), as expected, the closest over all other curves to the error axis. There are seven other curves plotted in Figure 4.6, alternately in solid and dotted lines; each is further from the error axis indicating improved performance. The first six of these (labelled with numbers 2 through 7) in order of decreasing error correspond to a transformed diagonal fitting using coefficient

channels derived:

2. under the maximum ignorance assumption (solid line) [this chapter].

3. with respect to a calibration set of real spectra (dotted line) [this chapter].

4. via spectral sharpening (solid line) [FDF94b].

5. via perfect sharpening, for 2- and 3-dimensional models of reflectance and illumination [FDF94a].

6. via perfect sharpening, for 3- and 2-dimensional models of reflectance and illumination [FDF94b].

7. via data-based sharpening [FDF94b].

Finally the last solid curve (labelled 8) furthest from the error axis corresponds to the performance afforded by a general linear transform.

Both of the coefficient transforms derived in this chapter take the simple coefficient histogram curve and move it on average about half way toward the general linear transform curve. That these channels do no better than that is in large part due to their poor performance with respect to the red CIE A illuminant—see Figure 4.5. Spectral sharpening, which finds narrow sensors by looking at wider wavelength intervals, fares much better raising coefficient performance to roughly four fifths of what it could be. Factoring statistical knowledge into the mechanisms for choosing coefficient channels (without any a priori requirement that they should behave like narrow-band sensors) improves coefficient performance still further. As we move from *2–3* to *3–2* finite-dimensional models of reflectance and illumination to the data-driven data-based sharpening the coefficient performance moves steadily closer to the general histogram curve. This is in large part to be expected since each method is based on more and more realistic statistical assumptions about color observations. Data-based sharpening gives performance close to the best that can be achieved by a linear model of color constancy.

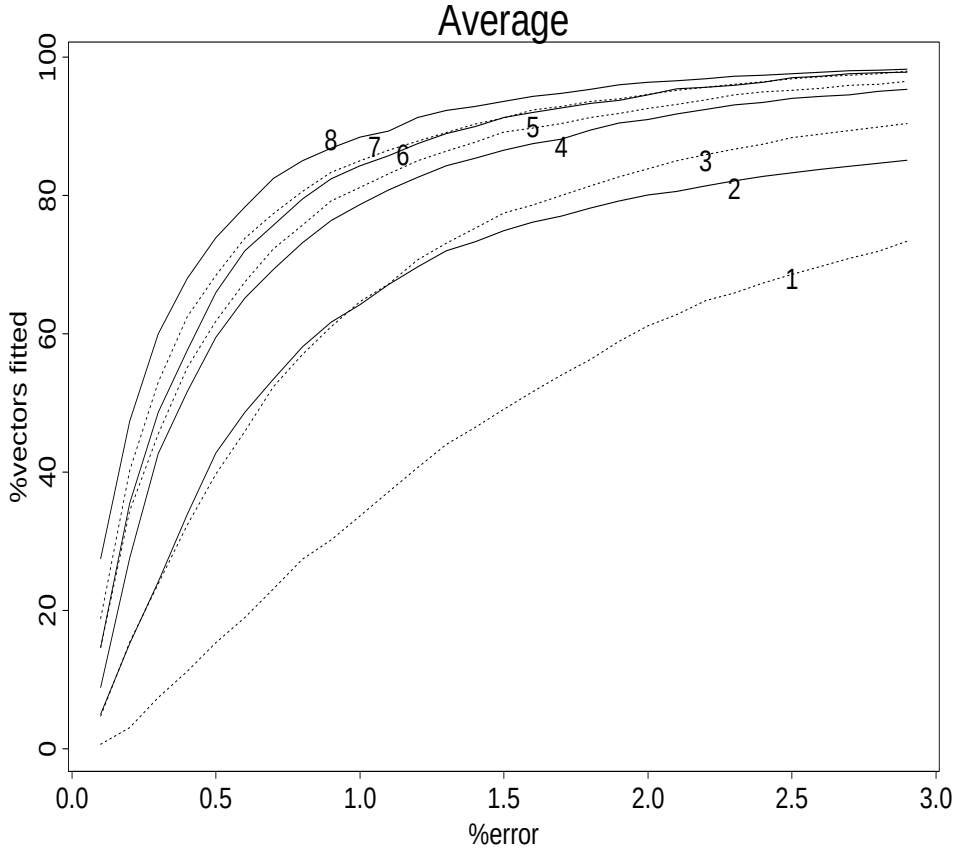

Figure 4.6: The average cumulative NFD histograms obtained for different sets of coefficient channels and the general image model. See text for description of labels.

The coefficient channels returned by all of the above methods are very similar to each other. In Figure 4.7 the maximum ignorance coefficient channels derived earlier are contrasted with the data-based and sensor-based sharpened curves. The most noticeable difference is that the databased-sharpened green sensor curve has a significantly shallower negative lobe in the red end of the spectrum. The corresponding lobe for the green sensor-based sharpened sensor is more pronounced though still shallower than the maximum ignorance curve.

4.5 Conclusion

We began with the hypothesis that there exists a linear combination of the cones—the coefficient channels—with respect to which a coefficient model is sufficient for color constancy. Methods were developed to solve for the cone combinations most consistent with this hypothesis. Simulation experiments indicated that derived coefficient channels support excellent color constancy.

The techniques developed here are closely related to other theoretical studies. We provide a theoretical justification of the spectral sharpening method of choosing coefficient channels. In particular we show that if we know nothing about the statistics of the color signal spectra entering the eye then sensor-based spectral sharpening is the appropriate tool. If the statistics of color signal spectra are known our methods are closely related to data-based sharpening [FDF94b].

We compare the coefficient channels derived in this study with those derived previously. While finding cone combinations which behave like narrow-band sensors certainly enhances coefficient color constancy it does not perform best overall. For the Munsell spectra viewed under daylight phases D48, D55, D65, D75, D100 and CIE A, the data-driven data-based sharpening method shows the best performance.

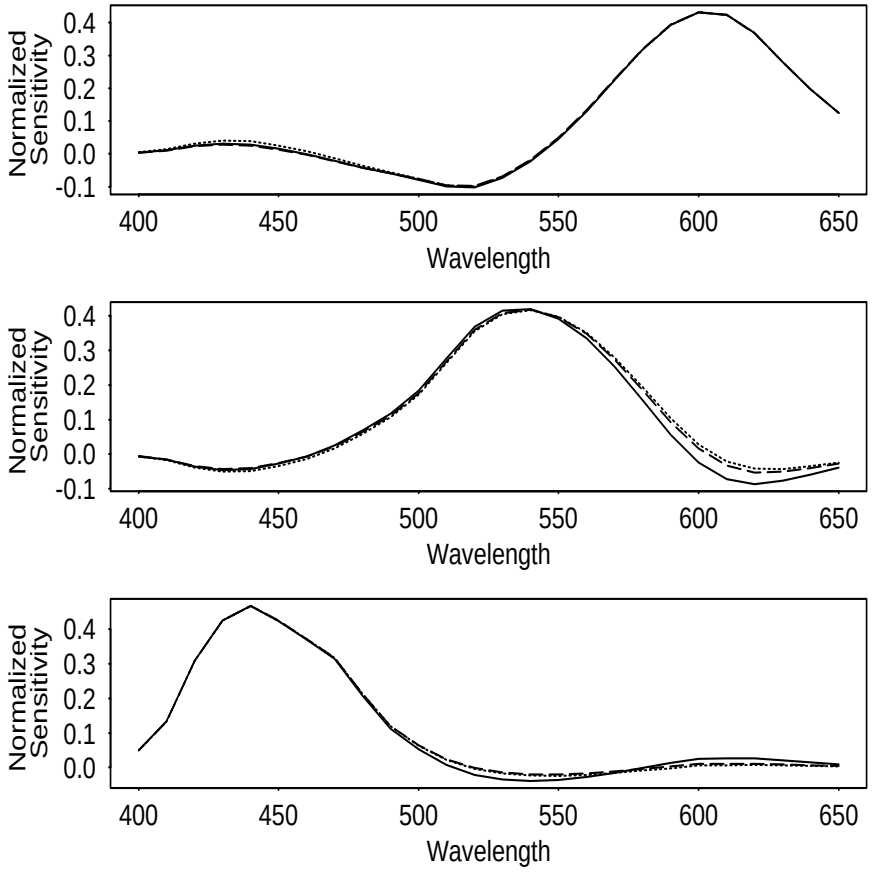

Figure 4.7: The maximum ignorance coefficient channels (solid lines) are contrasted with data-based sharpened curves (dotted lines) and sensor-based spectral sharpened curves (dashed lines).

Chapter 5

Color in Perspective

The work described in this chapter appears in: G.D. Finlayson. Color constancy in diagonal chromaticity space. *Proceedings of the 5th International Conference on Computer Vision.* IEEE Computer Society. June 1995.

Abstract

Simple constraints on the sets of possible surface reflectances and illuminants are exploited in a new color constancy algorithm that builds upon Forsyth's[*For*90] *theory of color constancy. The goal defined for a color constancy algorithm is to discount variations in the color and intensity of the incident illumination and thereby extract illumination-independent descriptors of surface colors from images. Forsyth's method invokes the constraint that the surface colors under a canonical illuminant all fall within an established maximal convex gamut of possible colors. This gamut constraint turns out to be very effective; however, other strong assumptions about the scenes are required for the method to work. In particular the illumination must be uniform, the surfaces must be planar, and there can be no specularities. To overcome these restrictions, the algorithm is first modified to work with the colors under a perspective projection, in essence in a chromaticity space. A crucial observation is that the convexity of the 3-dimensional gamut constraint is preserved by the perspective projection because gamut convexity is also required in the 2-dimensional chromaticity space if the gamut constraint is to be exploited efficiently. The new algorithm working in perspective is simpler than Forsyth's (its computational complexity is reduced) and more importantly the restrictions on the illuminant, surface shape and specularities can be relaxed. The algorithm is then extended to include a maximal gamut constraint on the set of illuminants that is analogous to the gamut constraint on surface colors. Using a perspective chromaticity space facilitates the expression of the illumination constraint in the algorithm. Tests on real images show that the algorithm provides good color constancy.*

5.1 Introduction

Color constancy is an important problem for any vision system requiring stable perception of the world. Common illuminants such as daylight and tungsten light or even just clear sky versus cloudy sky differ markedly in their spectral properties, and since the spectrum of the light impinging on an eye or camera is the product of the incident illumination and the percent spectral reflectance of the surface, the illumination must be accounted for and discounted if there is to be stable perception of the surface color. Despite extensive research into color constancy, for example [For90, FDH91, D'Z92, DL86, MW86, Lan77], there still does not exist a color constancy algorithm that performs sufficiently well that it either matches human color perception or that it provides a robot with adequate color recognition.

One of the most interesting color constancy theories to date is Forsyth's [For90] and since the new algorithm described in this chapter follows from it, his algorithm will be summarized first. Forsyth's algorithm assumes that scenes are populated only by flat, matte surfaces and that the (otherwise unknown) incident illumination is spatially constant in color and intensity. Given a 3-band *rgb* image of the scene, the defined task is to recover the *rgb* descriptor of each different surface in the scene as it would be seen by the camera under a standard canonical illuminant.

Consider the set of *rgb* response vectors obtained by imaging a maximal set of reflectances, in the sense that they are representative of all real surfaces, under the canonical illuminant; these *rgb* vectors will occupy a convex region of space. Forsyth calls this convex region the *canonical gamut*. The canonical gamut embodies a physical realizability constraint since an *rgb* response outside the gamut cannot be induced by a real surface. The same set of surfaces viewed under a second unknown illuminant leads to a second convex gamut which differs in shape and position from the canonical gamut. However the difference is not arbitrary; illumination change, and by extension the change in shape and position of gamuts, is well characterized by a linear transform [MW92]. In principal we can solve for the linear transform taking this gamut back to the canonical gamut thereby solving the color constancy problem. Of course it is unlikely that any real scene will contain all possible surface reflectances

and as a consequence the mapping back to the canonical illuminant will not be unique. Forsyth's MWEXT algorithm [For90] characterizes the set of all possible maps:

$$\forall \underline{p} \in \Gamma(\mathcal{I}) \ , \ \mathcal{M}\underline{p} \in \Gamma(\mathcal{C}) \qquad (5.1: \text{ MWEXT formulation})$$

The 3×3 map $\mathcal{M}$ is a possible solution to the color constancy problem if and only if each rgb response vector $\underline{p}$ in the image gamut, $\Gamma(\mathcal{I})$, is mapped inside the canonical gamut $\Gamma(\mathcal{C})$.

Characterizing the solutions to equation (5.1) is problematic in two respects. First for general linear maps there are 9 unknowns to be determined; Forsyth [For90] demonstrated that this is computationally expensive. Second, not all linear maps correspond to plausible changes in illumination. To address these problems Forsyth [For90] proposes that illuminant maps should be restricted to 3-parameter diagonal matrices.

Forsyth's restriction is valid as diagonal matrices provide a good model of illumination change for visual systems with relatively narrow-band spectral sensitivity functions. Such sensors are typical in color cameras. Even for sensors such as the human cones which are not narrow-band, a diagonal matrix remains a good vehicle for modelling illumination change so long as the sensitivity functions are first transformed to an appropriate sensor basis [FDF94b]. Forsyth's CRULE algorithm solves for all those diagonal matrices $\mathcal{D}$ which take the image gamut into the canonical gamut.

$$\forall \underline{p} \in \Gamma(\mathcal{I}) \ , \ \mathcal{D}\underline{p} \in \Gamma(\mathcal{C}) \qquad (5.2: \text{ CRULE formulation})$$

With the diagonal restriction, color constancy is simpler (has a lower computational cost) than the general linear case. However, it is still computationally expensive and only an approximate solution to equation (5.2) is actually calculated [For90]. Forsyth tested his CRULE algorithm on images of various planar patchworks of matte reflectances viewed under several different everywhere uniform illuminations. If an image contains a diverse set of colors then there are few diagonal matrices which can take the image gamut into the canonical gamut. For such diverse scenes CRULE was capable of delivering color constancy.

Real-world scenes are not populated only by flat matte surfaces illuminated by spatially constant illumination. In practice, specularities, varying illumination and

changes in shape, all affect the *rgb* responses such that a single linear transform will not suffice for mapping the image gamut to the canonical gamut; simply, in real scenes Forsyth's CRULE cannot work. In principle a program might preprocess the image and account for *rgb* variations due to shape, specularities and a varying illumination. However, one suspects that such preprocessing is a more difficult problem than color constancy.

Part of the failing of MWEXT and CRULE is that they set out to retrieve 3-dimensional color descriptors (eqns. (5.1) and (5.2)) when in fact this may not be the correct goal. What if we decide to settle for 2-dimensional color descriptors instead of 3-dimensional ones? Specifically suppose that the goal of color constancy is to retrieve only the orientation of *rgb* vectors under the canonical illuminant with color intensity not being retrieved. Such an approach has been taken by various other authors [MW86, D'Z92]—they argue that it is difficult to distinguish between a very reflective surface viewed under a dim illuminant from a dull surface viewed under a bright light. This argument can be extended to specularities, varying illumination and changes in shape. While each of these confound against intensity recovery, it is in principle still possible to retrieve canonical color orientation.

In abandoning intensity recovery the color constancy problem reduces to 2 dimensions; it makes sense therefore, to design a 2-dimensional solution method. To do this we first factor intensity out of the problem formulation by using a perspective projection. A perspective projection of color space results in what is commonly known as a chromaticity space. A typical chromaticity space is defined by

$$r' = \frac{r}{b} \ , \ g' = \frac{g}{b} \ , \ 1 = \frac{b}{b} \tag{5.3}$$

The vector $(r', g', 1)^t$ has the same orientation as $(r, g, b)^t$ but a different length or intensity. Because the third component $\frac{b}{b}$ always equals 1, perspective vectors can be represented as 2-vectors $(r', g')^t$.

Equation (5.3) is one of an infinite number of candidate perspective projections. It is special (though not unique) in the sense that the 3-dimensional canonical-gamut constraint can be easily maintained in the 2-dimensional projection; that is, the mappings from one illuminant to another in the 2-dimensional perspective space are linear

(we defer the proof of this to section 5.3). Relative to the perspective space defined in equation (5.3) our new color constancy algorithm is analogous to the 3-dimensional CRULE algorithm. The fundamental difference is that the algorithm now solves for the 2-dimensional linear maps transforming the 2-dimensional image gamut to the 2-dimensional canonical gamut. The shift from 3 to 2 dimensions greatly simplifies the gamut mapping algorithm. Indeed the computation becomes so much simpler that while in the 3-D case it is only practical to search for approximate solution, in the 2-dimensional case exact solutions can be found.

In either the 3-dimensional or 2-dimensional cases the algorithm computes a set of possible solutions to the color constancy problem, represented as set of mappings between the unknown scene illuminant and the canonical illuminant. If there are only a few colors in the image then the image gamut may provide very little constraint, and the set of possible mappings will be quite large. The solutions can be further constrained, however, if the gamut of possible illuminants is taken into account. Unfortunately, constraints on the illuminant and reflectance gamuts cannot both be represented linearly at the same time. Representing the illumination constraint non-linearly increases the computational cost of our algorithm. However, the overall time complexity is unchanged, the cost increase is linear, and the algorithm runs in a timely manner. In the 2-dimensional case the illumination constraint significantly improves the algorithm's color constancy.

In section 5.2, we review the fundamentals of Forsyth's gamut mapping approach to color constancy. Practical difficulties with Forsyth's CRULE algorithm are discussed in section 5.3. Addressing these leads to the perspective color constancy presented in section 5.4. In section 5.5 a physical realizability constraint on the illuminant is factored into our algorithm. Color constancy results for real camera images are given in section 5.6. Good color constancy is possible. Moreover physical realizability on the illuminant is shown to be a powerful, and necessary, constraint. Finally we present conclusions in section 5.7.

5.2 The Gamut Mapping Method

The gamut mapping method of estimating the spectral properties of the illumination begins with the idea of a canonical gamut consisting of the set of sensor response vectors obtained from viewing all physically realizable surfaces under a canonical illuminant. To construct an approximation to the full canonical gamut, consider an image of a planar patchwork of n surfaces under the canonical illuminant.

$$C = \{\underline{p}^{1,c}, \underline{p}^{2,c}, ..., \underline{p}^{n,c}\} \tag{5.4}$$

Where the superscript j, c denotes the jth surface viewed under the canonical illuminant c. Each $\underline{p}^{i,c}$ is a 3-vector, the rgb camera response to the ith reflectance. Because each rgb arises from a physically realizable surface, any convex combination $\underline{p}^{x,c}$ within the set $\mathcal{C}$ is also physically realizable [For90]:

$$\underline{p}^{x,c} = \sum_{j=1}^{n} \omega_j \underline{p}^{j,c} \quad , \quad \forall(j)\, \omega_j > 0 \; , \; \sum_{j=1}^{n} \omega_j = 1 \tag{5.5}$$

Equation (5.5) can be interpreted in the following way. Imagine that we take a small piece of surface that maps to a single pixel in the viewing camera and cover it with each of the n surfaces in proportion to the weights ω_i. In this case, (5.5) exactly characterizes the camera response for that pixel. The set of all convex combinations of C is denoted $\Gamma(C)$ and will be used as the canonical gamut.

Although $\Gamma(C)$ is an infinite set, the computational cost of gamut mapping depends only on the number of points in C. This cost can be reduced by removing points from C which are redundant in constructing the set of convex combinations. The convex hull [PS85] of C, denoted $\mathcal{C}$, is the minimal cardinality subset of C such that $\Gamma(C) = \Gamma(\mathcal{C})$.

Let $\mathcal{I}$ be the convex hull points for a set of rgb's obtained under an unknown illuminant e. Any mapping which takes the image gamut $\Gamma(\mathcal{I})$ into the canonical gamut is a candidate solution to the color constancy problem. The uth point $\underline{p}^{u,e}$ (unless zero) in $\mathcal{I}$ can be mapped to the vth point $\underline{p}^{v,c}$ of $\mathcal{C}$ by applying a 3×3 diagonal matrix of the following form:

$$\mathcal{D}^{u,v}\underline{p}^{u,e} \;=\; \underline{p}^{v,c} \tag{5.6a}$$

The diagonal elements are simply the coordinates of $\underline{p}^{v,c}$ divided by those of $\underline{p}^{u,e}$:

$$\mathcal{D}^{u,v}_{kk} \;=\; \frac{p^{v,c}_k}{p^{u,e}_k}, \;\; (k = 1,2,3) \tag{5.6b}$$

The entire set of mappings taking $p^{u,e}$ into $\mathcal{C}$ is denoted as $\mathcal{C}/\underline{p}^{u,e}$. Since the diagonal matrices have only 3 non-zero entries, the members of $\mathcal{C}/\underline{p}^{u,e}$ can be represented as 3-vectors instead of matrices. Using this representation it follows that:

$$\mathcal{C}/\underline{p}^{u,e} \;=\; \{\mathcal{D}\underline{p} \;:\; \underline{p} \in \mathcal{C}\}, \quad \mathcal{D}_{kk} \;=\; \frac{1}{p^{u,e}_k}, \;\; (k = 1,2,3) \tag{5.7}$$

The points in $\mathcal{C}/\underline{p}^{u,e}$ are exactly the points in $\mathcal{C}$ mapped by the diagonal matrix $\mathcal{D}$ It follows that $\underline{p}^{u,e}$ can be mapped to any point in $\Gamma(\mathcal{C})$ by an appropriate convex combination of $\mathcal{C}/\underline{p}^{u,e}$; since $\underline{p}^{u,e}$ mapped by the elements of $\mathcal{C}/\underline{p}^{u,e}$ is simply $\mathcal{C}$. The set of convex combinations of $\mathcal{C}/\underline{p}^{u,e}$ is denoted $\Gamma(\mathcal{C}/\underline{p}^{u,e})$.

The set of diagonal matrices, $\Gamma(\mathcal{C}/\mathcal{I})$, that simultaneously map all points of $\mathcal{I}$ into $\Gamma(\mathcal{C})$ is the intersection of all the sets mapping the individual points of $\mathcal{I}$ into $\Gamma(\mathcal{C})$:

$$\Gamma(\mathcal{C}/\mathcal{I}) \;=\; \bigcap_{j=1}^{m} \Gamma(\mathcal{C}/\underline{p}^{j,e}) \tag{5.8}$$

Figure 5.1 shows a 2-dimensional example of $\Gamma(\mathcal{C}/\mathcal{I})$ as the intersection of 3 triangles. Each triangle represents the convex hull of the set of 2-dimensional diagonal matrices mapping a vertex of $\mathcal{I}$ into $\Gamma(\mathcal{C})$.

In 3 dimensions the $\Gamma(\mathcal{C}/\underline{p}^{j,e})$ are convex polytopes and (5.8) is the common volume. The cost of intersecting two convex polytopes is determined by the number of hull points. Given 2 polytopes each with n hull points their intersection can be calculated in $O(n \log n)$. The number of hull points in the intersection is finite but can be much greater than the number for either of the original polytopes (it can be shown that the common intersection of m polytopes each with n hull points can have as many as $3mn - 6$ hull points), so calculating (5.7) can be computationally expensive. Consequently Forsyth's [For90] CRULE solves for an approximation of $\Gamma(\mathcal{C}/\mathcal{I})$.

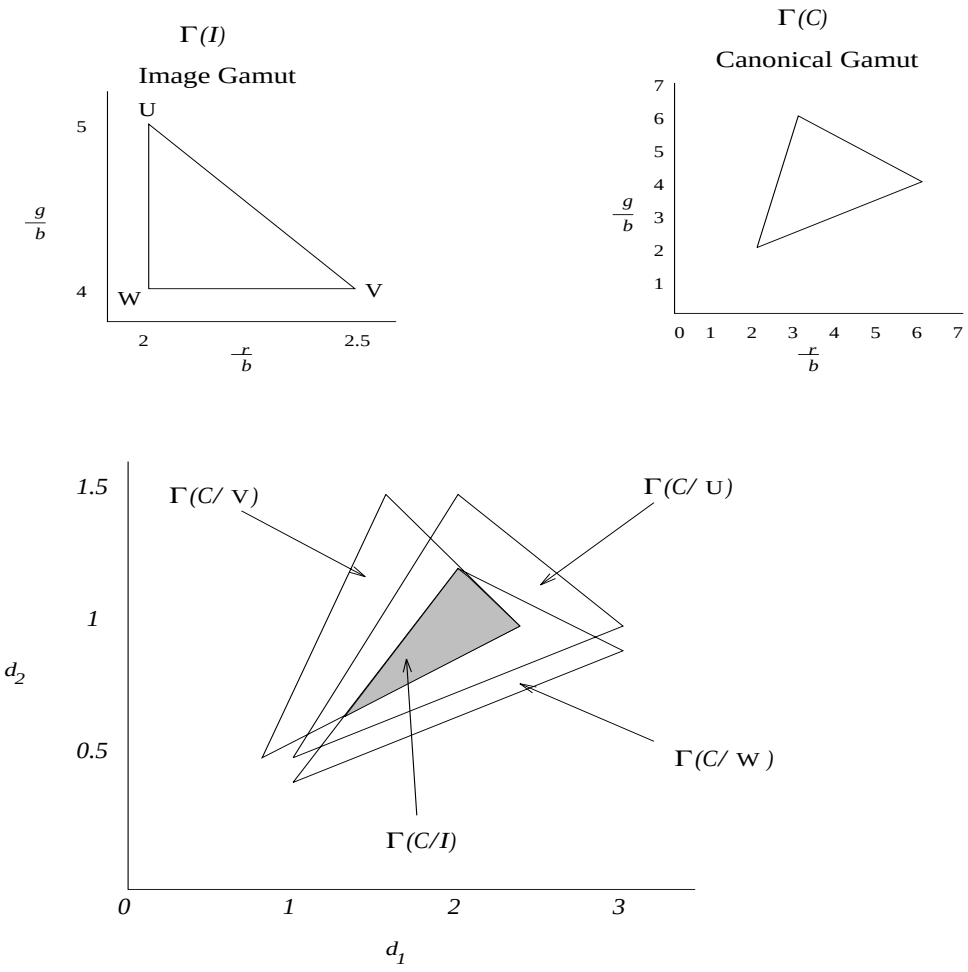

Figure 5.1: The shaded area represents the set of diagonal mappings taking the image gamut into the canonical gamut. It is the intersection of the 3 sets of mappings taking each of the image gamut's hull points U, V and W, into the entire canonical gamut.

$\Gamma(\mathcal{C}/\mathcal{I})$ is exactly the set of diagonal matrices which map the image gamut $\Gamma(\mathcal{I})$ into $\Gamma(\mathcal{C})$. This follows from the fact that the convex combinations in (5.6) are preserved under a diagonal transform:

$$\mathcal{D}\underline{p}^{x,c} = \sum_{j=1}^{n} \omega_j \mathcal{D}\underline{p}^{j,c} \quad , \quad \forall(j)\, \omega_j > 0 \; , \; \sum_{j=1}^{n} \omega_j = 1 \tag{5.9}$$

Hence if $\mathcal{I}$ is mapped within $\Gamma(\mathcal{C})$ by a diagonal matrix $\mathcal{D}$, then the convex combinations of $\mathcal{D}\mathcal{I}$ are also in $\Gamma(\mathcal{C})$, and these convex combinations are precisely the interior points of $\Gamma(\mathcal{I})$ under the mapping $\mathcal{D}$.

The image gamut serves to constrain the set of possible solutions. Unless the image contains a diverse and representative set of colors, $\Gamma(\mathcal{C}/\mathcal{I})$ will be large. To chose a good candidate from $\Gamma(\mathcal{C}/\mathcal{I})$ and thereby provide a unique estimate of the unknown illuminant, Forsyth employs the heuristic that the volume of $\Gamma(\mathcal{I})$ mapped within $\Gamma(\mathcal{C})$ be maximum.

CRULE provides an adequate estimate of the illuminant when applied to images of many reflectances provided that Forsyth's assumptions (planar surfaces of matte reflectance under uniform illumination).

5.3 Color in Perspective

Unfortunately, very few scenes will meet the restrictions imposed by CRULE. Since few surfaces are flat, perhaps the planarity restriction is the most serious one. It cannot be circumvented simply by moving to smooth surfaces that are locally approximately planar because locality competes with the need for a comprehensive image gamut. Constant illumination is also unrealistic and might be handled by assuming only that it is locally constant. Most surfaces have a significant specular component and even if only a small part of the scene creates a specularity it will ruin the CRULE's results for the whole image.

For a patch of constant reflectance, varying the shape effects the intensity but not the hue of the reflected light. The length of the measured *rgb* will change but not its direction. As long as the illumination is spectrally constant—which will be assumed

to be the case— variations in illumination intensity will create a similar effect.

Specularities also create an intensity problem for CRULE. In essence a specularity acts as a very white surface in that it reflects the incident light without significantly changing its spectral composition. However, a specularity is much more reflective than any physically realizable white surface [Bre89]. Therefore, in mapping the image gamut to the canonical gamut, the specularity will be mapped to the most reflective white in the canonical gamut forcing the other surfaces to be interpreted as very dark.

Curved surfaces, specularities and spatially varying illumination intensity all mitigate against retrieving the intensity of surface colors. What if we abandon intensity and concentrate instead on recovering only the correct orientation of the *rgb* vector? In this case color constancy reduces to a problem with only 2 degrees of freedom. In rephrasing the color constancy problem in 2 dimensions, there are many possible transformations that could be used to map the 3-dimensional *rgb* response vectors to 2-dimensional intensity-independent counterparts, but only some of them permit us to continue using the gamut-mapping method. One of them is the perspective transform given in equation (5.3).

In the 3-dimensional case, the canonical gamut is a convex set and the illumination mapping is calculated as the intersection of convex sets representing 3×3 diagonal matrices. The non-linearity of the perspective transform means that the 3-dimensional gamut constraint need not translate necessarily into a similar constraint involving convex sets of 2×2 diagonal matrices and this should be the case if we are to use the gamut mapping method.

For the gamut constraint to carry over to 2 dimensions, it must be the case that for each diagonal matrix mapping between a pair of rgb's there exists a corresponding 2-dimensional diagonal matrix mapping between that pair's perspective coordinates. Suppose that the 3×3 diagonal matrix $\mathcal{D}$ maps $\underline{p}^{j,1}$ to $\underline{p}^{j,2}$ (i.e. $\underline{p}^{j,2} = \mathcal{D}\underline{p}^{j,1}$). Let the perspective transformation be defined by equation (5.3). It follows that the perspective coordinates of $\underline{p}^{j,1}$ and $\underline{p}^{j,2}$ are:

$$\underline{q}^{j,1} = (p_1^{j,1}/p_3^{j,1}, p_2^{j,1}/p_3^{j,1})^t = \underline{q}^{j,2} = (p_1^{j,2}/p_3^{j,2}, p_2^{j,2}/p_3^{j,2})^t \quad (5.10)$$

Let $\mathcal{D}'$ be the 2-dimensional diagonal matrix with diagonal elements $\mathcal{D}'_{11} = \mathcal{D}_{11}/\mathcal{D}_{33}$

and $\mathcal{D}'_{22} = \mathcal{D}_{22}/\mathcal{D}_{33}$. Clearly, $\underline{q}^{j,2} = \mathcal{D}'\underline{q}^{j,1}$: a diagonal matrix maps between perspective colors. Under the perspective transform in (5.3) illuminant maps remain diagonal matrices.

The tractability of Forsyth's gamut mapping procedure is contingent on the fact that the gamuts are convex and that they can be described by a finite (usually quite small) number of hull points. These properties must also be preserved in perspective if we are to use the gamut mapping procedure. The notions of convexity and convex combinations are stated algebraically in eqn. (5.4). An equivalent geometric interpretation of (5.4) is:

$$\text{if } \underline{p}^{1,c}, \underline{p}^{2,c} \in \Gamma(\mathcal{C}) \text{ then } \forall \alpha \in [0,1] \ \alpha\underline{p}^{1,c} + (1-\alpha)\underline{p}^{2,c} \in \Gamma(\mathcal{C}) \tag{5.11}$$

If $\underline{p}^{1,c}$ and $\underline{p}^{2,c}$ are vectors, or points, in the convex set $\Gamma(\mathcal{C})$ then any point, or vector, on the straight line between $\underline{p}^{1,c}$ and $\underline{p}^{2,c}$ is also in $\Gamma(\mathcal{C})$. Under a perspective transform, lines in 3 dimensions are mapped onto lines in 2 dimensions [FvD90]. Let $\underline{q}^{1,c}$ and $\underline{q}^{2,c}$ be the perspective projections of $\underline{p}^{1,c}$ and $\underline{p}^{2,c}$. It follows that any point on the straight line between $\underline{p}^{1,c}$ and $\underline{p}^{2,c}$ in 3 dimensions is mapped to a point on the straight line between $\underline{q}^{1,c}$ and $\underline{q}^{2,c}$ in 2 dimensions. Simply, convex combinations, and convexity in general, are preserved under a perspective transform. The canonical gamut $\Gamma(\mathcal{C})$ in perspective can be represented as convex combinations of the perspective projections of the hull points $\mathcal{C}$.

The canonical and image gamuts of perspective colors are 2-dimensional convex polygons. As a result all the gamut-mapping can be carried out in terms of 2-dimensional convex polygons and this simplifies computation. Not only does it take fewer operations to intersect two convex polygons than intersecting two convex polyhedra (assuming the number of hull points is the same in both cases) but the number of points on the intersected hull does not grow as quickly as it does in the 3-dimensional case. Intersecting two convex polygons, each with n hull points costs $O(n)$ and the number of hull points defining the common intersection of m, n-vertex polygons can be shown to be mn. Moreover the actual implementation of intersection in 2 dimensions is much simpler. As a result, in 2 dimensions we compute $\Gamma(\mathcal{C}/\mathcal{I})$ exactly.

5.4 Illuminant Gamut Constraint

In the above, an illuminant map, is represented as the diagonal matrix transforming sensor responses under one illuminant to those that would be obtained under the canonical illuminant. During gamut mapping all diagonal matrices are considered as candidates for the unknown illuminant map, but clearly not all diagonal matrices correspond to actual illuminant maps. To restrict the solutions to actual illuminant maps, we introduce the idea of a canonical illuminant gamut.

Consider the set of rgb response vectors obtained by viewing a single standard surface s reflectance under a large set of n representative illuminants:

$$E = \{\underline{p}^{s,1}, \underline{p}^{s,2}, ..., \underline{p}^{s,n}\} \tag{5.12}$$

Let $\mathcal{E}$ denote the convex hull of E. The camera response vectors of the standard surface viewed under all possible convex combinations of the illuminants is denoted $\Gamma(\mathcal{E})$. If $\underline{p}^{s,c}$ denotes the response obtained from the standard surface under the canonical illuminant then, using the nomenclature of section 5.2, $\Gamma(\mathcal{E}/\underline{p}^{s,c})$ is the set of all diagonal maps taking the canonical illuminant to all other illuminants.

Suppose that $\mathcal{D}$ is a diagonal matrix taking the image into the canonical gamut, $\mathcal{D} \in \Gamma(\mathcal{C}/\mathcal{I})$. Such a mapping is permissible only if $\mathcal{D}^{-1} \in \Gamma(\mathcal{E}/\underline{p}^{s,c})$; that is $\mathcal{D}^{-1}$ takes the canonical descriptor of the standard surface to an observation in $\mathcal{E}$. Let $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ denote the set of diagonal mappings taking observations from the back to the canonical illuminant. Clearly, if $\mathcal{D} \in \Gamma(\mathcal{E}/\underline{p}^{s,c})$ then $\mathcal{D}^{-1} \in \Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$. The set $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ effectively places a gamut constraint on the illuminant.

For surfaces belonging to the canonical gamut (of surfaces) and illuminants belonging to the illumination gamut, the set of diagonal matrices representing the unknown illuminant is:

$$\Gamma(\mathcal{C}/\mathcal{I}) \bigcap \Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c}) \tag{5.13}$$

In contrast to $\Gamma(\mathcal{E}/\underline{p}^{s,c})$, $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ is not convex (see Figure 5.2). As a result the intersection required in (5.13) is more costly to compute with time complexity $O(n^2)$. Fortunately though, the illuminant-gamut constraint requires intersection

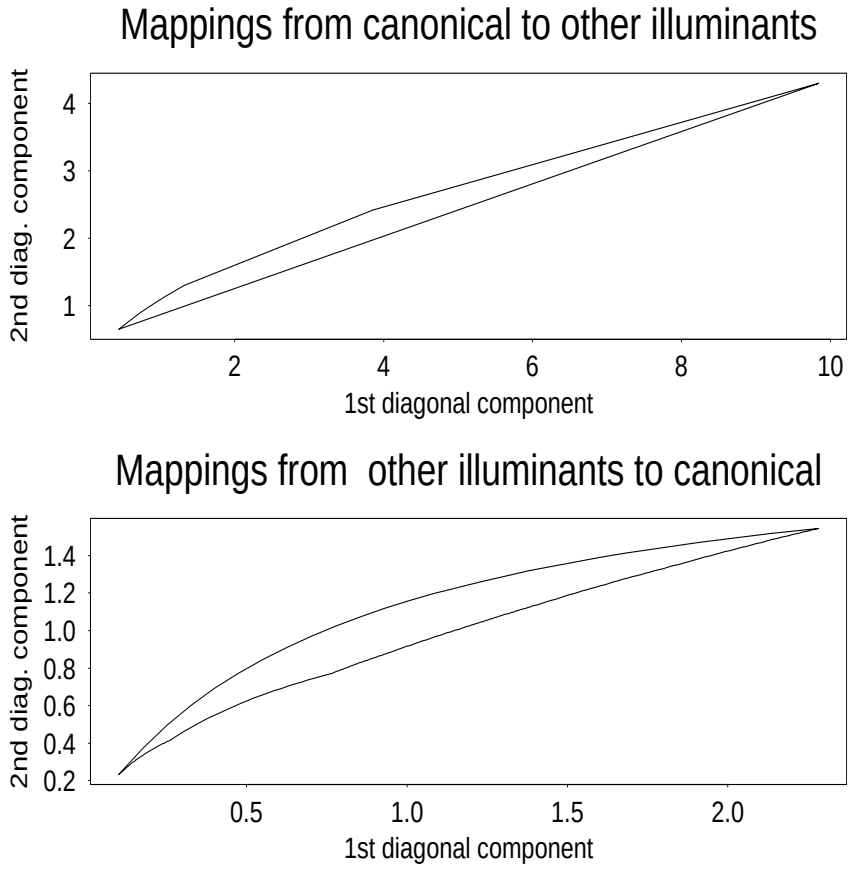

Figure 5.2: The set of mappings taking the canonical to other illuminants, top graph, is convex. The mappings taking other illuminants to the canonical, bottom graph, is non-convex.

only once. In contrast many, albeit simpler, intersections are calculated when enforcing the canonical surface gamut constraint at a net cost of $O(n^2)$. Therefore overall algorithm complexity remains at $O(n^2) = O(n^2) + O(n^2)$ when the illumination constraint is included. Using an illuminant gamut constraint in 3 dimensions appears to be a much more complex task.

5.5 Results

Before any color constancy experiments can be run the gamuts $\Gamma(\mathcal{C})$ and $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ must be created. To do this the chromaticities of real surfaces viewed under a real illuminant could be used as a basis for calculating $\Gamma(\mathcal{C})$. Similarly $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ could also be estimated given the camera chromaticities of some real surface under various test illuminants. In practice however, confounding factors such as camera noise will cause the measured camera responses to depart from the ideal. This led us to build the gamuts using a perfectly Lambertian mathematical model. Given a measured reflectance $S(\lambda)$, a measured illuminant $E(\lambda)$ and the 3-vector of camera sensitivities $\underline{R}(\lambda)$ the induced camera response is calculated as:

$$\underline{p} = \int_{\omega} \underline{R}(\lambda)E(\lambda)S(\lambda)d\lambda \tag{5.14}$$

The response functions of the SONY DXC-930 camera used in the experiments were derived via the methods of Sharma and Trussel [ST93]. For reflectances the 24 Macbeth color checker patches [BP88] were used, and for illuminants the 6 Judd daylight phases D48, D55, D65, D75, D100, D200 [JMW64], CIE standard illuminants [WS82] A, B, and C, a 2000K planckian black body radiator, and a uniform white. The uniform white was chosen to be the canonical illuminant; the 24 canonical chromaticities of the Macbeth color checker were calculated according to equations (5.14) and (5.3). The corresponding gamut is shown in Figure 5.3.

Although the Macbeth checker reflectances are often used to represent all reflectances, there will in practice be chromaticities which will lie outside this gamut. To account for this the gamut was increased by 5%. The new gamut, also shown in

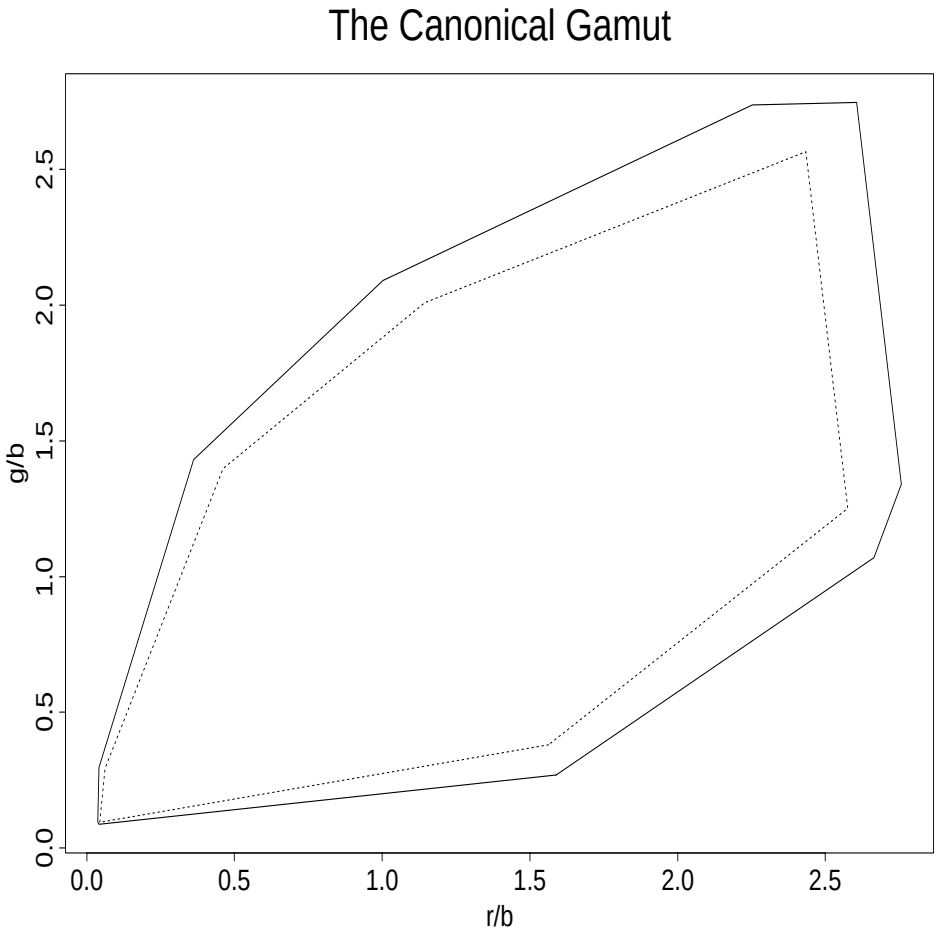

Figure 5.3: The dotted line is the gamut of chromaticities of the 24 Macbeth patches. The solid line delimits the same gamut expanded by 5%.

Figure 5.3, is sufficiently large that it contains the chromaticities of the 341 natural reflectances measured by Krinov [Kri47] and 452 of the 462 Munsell reflectances [Nic57].

In creating $\Gamma(\mathcal{E})$ a perfect white reflecting surface is used as the standard reflectance. Again using (5.14) and (5.3) the chromaticities of the standard white under each of the 11 illuminants were calculated; this gives a the set E from which $\Gamma(\mathcal{E}/\underline{p}^{s,c})$ and $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ are readily calculated (see Figure 5.2).

Three real images of the Macbeth color checker were taken relative to three different illuminating conditions: a tungsten light, cloudy sky outdoor illumination and blue sky outdoor illumination. The Macbeth color checker is shown by the window of our lab in Figure 5.4. Four checker patches were randomly selected to use in our constancy experiments; this set is denoted *small*. A further 4 checker patches were randomly selected from the Macbeth checker (duplicates were allowed) and added to *small* making a larger set which is denoted *medium*. Finally another 4 checker patches were added to *medium* making the set *large*. The a priori expectation is that color constancy should become easier (more constrained) the larger the set of reflectances.

Denote the gamut of chromaticities of these three sets under an unknown illuminant e as $\mathcal{I}^{\text{small},e}$, $\mathcal{I}^{\text{medium},e}$ and $\mathcal{I}^{\text{large},e}$. The input to the color constancy algorithm will be one of these three sets; the goal is to recover the chromaticities under the known canonical illuminant ($\mathcal{I}^{\text{small},c}$, $\mathcal{I}^{\text{medium},c}$ and $\mathcal{I}^{\text{large},c}$).

For the chromaticities measured under the tungsten illuminants $\Gamma(\mathcal{C}/\mathcal{I}^{\text{small,tungsten}})$, $\Gamma(\mathcal{C}/\mathcal{I}^{\text{medium,tungsten}})$ and $\Gamma(\mathcal{C}/\mathcal{I}^{\text{large,tungsten}})$ were calculated. These are shown at the top of Figure 5.5. The outermost convex polygon (lightly shaded) delimits all those possible maps which will take $\mathcal{I}^{\text{small,tungsten}}$ into $\Gamma(\mathcal{C})$. The next largest polygon (embedded in the first and shaded more darkly) takes $\mathcal{I}^{\text{medium,tungsten}}$ into $\Gamma(\mathcal{C})$. The third (most darkly shaded) and smallest polygon contains those maps that take $\mathcal{I}^{\text{large,tungsten}}$ into $\mathcal{C}$. The decreasing size of the sets of candidate mappings is expected—the more reflectances in each scene the more constrained the color constancy problem becomes. The illuminants which are in the illuminant gamut (those in $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$) must lie inside the region bounded by the dotted curves. It is clear that $\Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$ strongly constrains the possible mappings to the canonical, especially for the *small* and *medium* sets.

Figure 5.4: The Macbeth Color Checker illuminated by cloudy sky daylight.

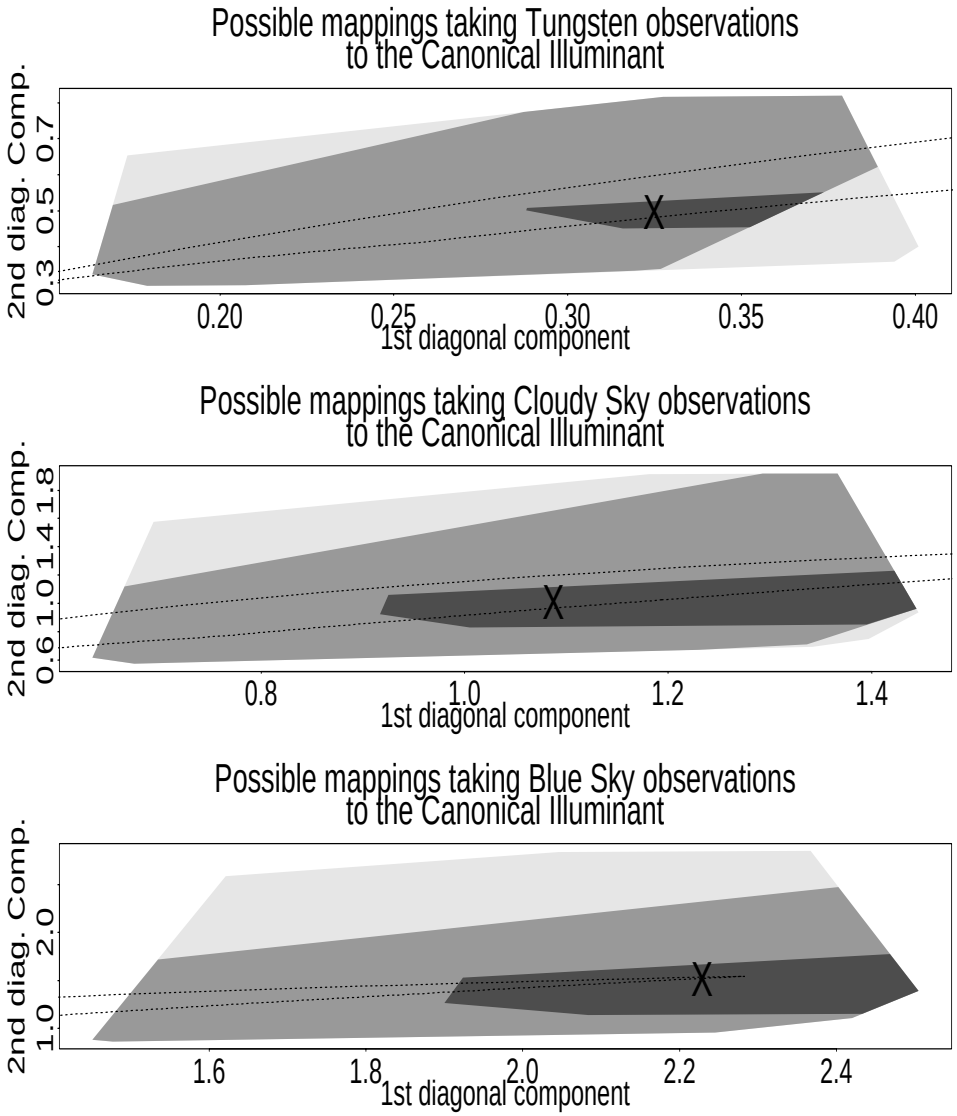

Figure 5.5: Plausible color constancy solutions for various sets of reflectances viewed under various illuminants—see text for explanation.

Finally the diagonal map which best takes all 24 measured chromaticities to their canonical counterparts in the least-squares sense is marked by a X. If Q^c and Q^e are the chromaticities of the 24 Macbeth checker patches under canonical and unknown illuminants, the best diagonal map $\mathcal{D}^{e,c}$ minimizes

$$||\mathcal{D}^{e,c}\mathcal{Q}^e - \mathcal{Q}^c||_F \tag{5.15}$$

where $||.||_F$ denotes the Frobenius norm; i.e., the square root of the sum of squared differences between $\mathcal{Q}^c$ and $\mathcal{D}^{e,c}\mathcal{Q}^e$.

The middle portion of Figure 5.5 shows the sets of illuminant maps calculated for each of the three reflectance sets for the cloudy white illuminant. Finally the sets of illuminant maps calculated for the blue sky illuminant are shown in the bottom graph in Figure 5.5. Note that the extent of the axes differ markedly for all three illuminants.

To evaluate the color constancy delivered by the algorithm it one must first estimate the cost of doing no color constancy processing. That is, quantify the extent to which colors change with a change in illumination. In computing this—because the chromaticity space is not metric relative to the original 3-dimensional color space—3-vectors calculated from the chromaticities are used. If (r', g') is the chromaticity then $(r', g', 1)$ is the 3-vector in the direction of the original 3-vector camera response, but with a different intensity (eqn. (3)).

Consider an *rgb* vector of some surface under some unknown illuminant. Under the canonical illuminant, its chromaticity must fall within the canonical gamut. However, without doing any color constancy processing one would be forced to guess where it falls in that gamut. Let $\underline{p}^{j,c}$ denote the correct response vector of surface j under the canonical illuminant and $\underline{p}^{\text{guess},c}$ a random guess. The worst case performance of doing nothing is defined as

$$nothing(j) \;=\; \max_{\text{guess}} \Theta(\underline{p}^{j,c}, \underline{p}^{\text{guess},c}) \;\;,\;\; \underline{q}^{\text{guess},c} \in \Gamma(\mathcal{C}) \tag{5.16}$$

where $\Theta()$ is a function returning the angle between two vectors and $\underline{q}^{\text{guess},c}$ is the chromaticity corresponding to the 3-vector $\underline{p}^{\text{guess},c}$. One might argue that the worst

case performance defined in equation (5.16) is too pessimistic; after all there is no constraint on the illuminant in (5.16). This consideration leads to a second measure of doing nothing. The maximum angle between the canonical camera response vector and the corresponding vectors of the same surface under any of our 3 test illuminants (tungsten, cloudy sky and blue sky) is calculated:

$$nothing2(j) = max(\{\Theta(\underline{p}^{j,c}, \underline{p}^{j,\text{tungsten}}), \Theta(\underline{p}^{j,c}, \underline{p}^{j,\text{cloud}}), \Theta(\underline{p}^{j,c}, \underline{p}^{j,\text{bluesky}})\}) \quad (5.17)$$

In evaluating the color constancy performance of the algorithm one must first take the set of plausible mappings and choose amongst them in order to have a single answer. To do this Forsyth's heuristic is used; the map which maximizes the area of the image gamut after it is mapped to the canonical illuminant is chosen. In terms of Figure 5.4 this roughly translates into choosing the points furthest from the origin for each set of mappings. The maximum area heuristic effectively forces the image to be as colorful as possible. For each illuminant and each reflectance set two mappings are calculated. The first is chosen using only the gaumt constraint on surfaces, i.e., the map in $\Gamma(\mathcal{C}/\mathcal{I})$ under which the image gamut has maximum area. The second, incorporating the illuminant constraint, is the maximum area mapping in $\Gamma(\mathcal{C}/\mathcal{I}) \bigcap \Gamma^{-1}(\mathcal{E}/\underline{p}^{s,c})$. Given a chromaticity under an unknown illuminant it is straightforward to calculate its chromaticity with respect to either of these maps and from this the corresponding 3-vectors relative to the canonical illuminant. The angle between these and the correct canonical response vector is a measure of the algorithm's color constancy performance.

In the top of Figure 5.6 the performance of the color constancy algorithm for the Macbeth color checker viewed under tungsten light is graphed.

The graph is in three sections, one for each data set (*small*, *medium* and *large*) and within each section there are five error bars. The leftmost bar (lightest shading), gives the range of worst-case errors as defined by equation (5.16) and the next bar (slightly darker) as defined by (5.17). The third bar (darker again) shows the recovery performance of the algorithm using only the surface gamut constraint. The fourth bar (still darker) shows the performance when the illuminant constraint is also

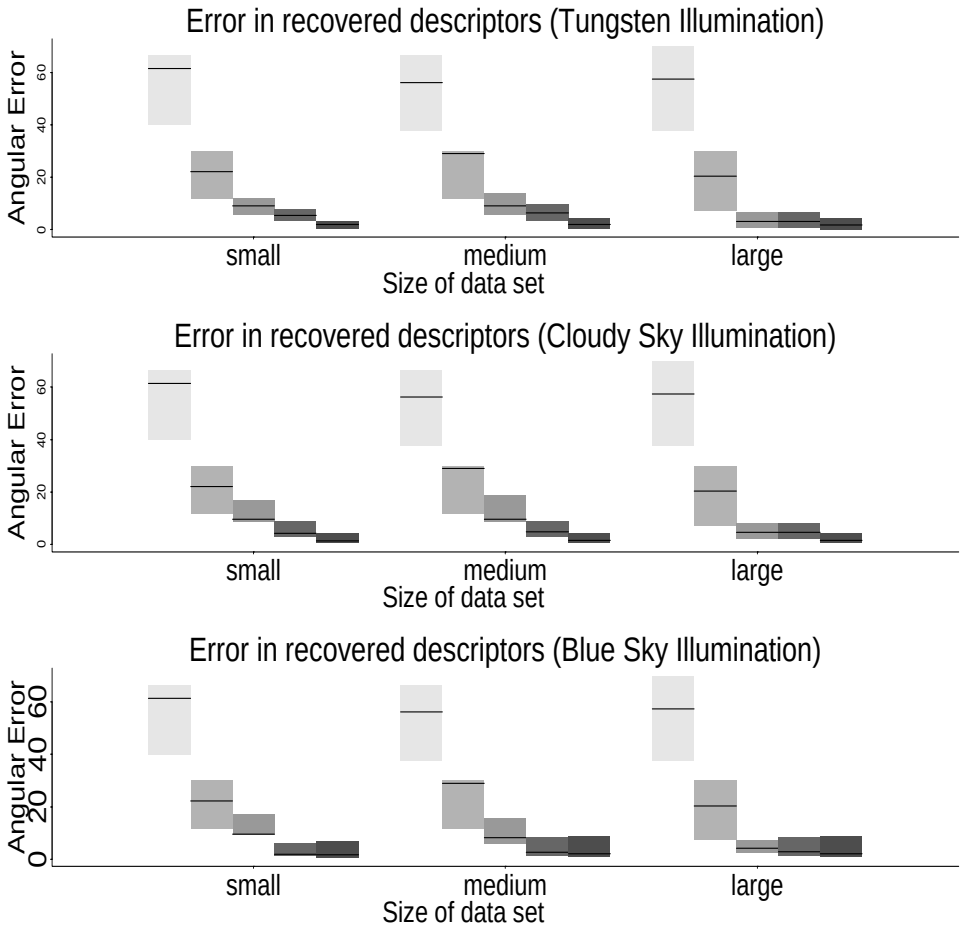

Figure 5.6: The color constancy error for various color constancy solution methods is displayed. The top graph records the performance returned when the input consists of three sets of reflectances are viewed under Tungsten illuminant. The middle and bottom graphs record performance when the same reflectances are viewed under Cloudy sky and Blue sky illuminants. A detail discussion of the experiment is given in the text.

used. Finally the fifth bar (darkest overall) shows the error performance when the best least-squares fit diagonal matrix is used (the X's in Figure 5.5). This last bar does not correspond to any algorithm but rather to the case when there is a known correspondence between canonical chromaticities and those of the unknown illuminant and thus serves as a control on how much color constancy can actually be attained. The straight lines on each of the bars shows the median angular error in the recovery.

Color constancy performance under a cloudy sky light is shown in the middle graph. Finally the bottom graph shows algorithm performance for the blue sky illumination.

It is clear that the algorithm returns results far better than doing nothing for all illuminants and all reflectance sets. Moreover it is apparent that a constraint on the illuminant is very powerful, especially for the *small* and *medium* reflectance sets—much better constancy is afforded than using only the surface gamut constraint. Moreover, in almost all cases, exploiting illuminant and surface constraints results in constancy of a similar order as that available in the control *best case*.

In real scenes the color of the illuminant changes from place to place. For example at the window of an office daylight is dominant. As one moves further to the interior of the room the component due to artificial illumination becomes stronger. It is imperative therefore, in the context of the current theory, that color constancy be delivered given only local *rgb* measurements. The more local these measurements, the fewer the chromaticities contributing to the image gamut. That such good color constancy is possible with a reflectance set of just 4 patches is extremely encouraging in this regard.

5.6 Conclusions

Color constancy has proven to be a very hard problem to solve. Existing methods are all limited in that they can only work for scenes satisfying very unrealistic assumptions. For example some approaches require a white reflectance in every scene while others assume that all surfaces are planar. Most assume there are no specularities. In this chapter an algorithm was developed which can deliver color constancy under

more realistic assumptions.

There are two observations. First, the range of camera responses induced by different reflectances varies with a change in illumination and second, that illumination can only vary within certain bounds. These observations are all that is assumed in the algorithm. Consequently color constancy becomes possible in images of real world scenes, including those with shape, specularities and illumination power variation.

The algorithm was tested on real images of Macbeth color checker patches. Extremely good constancy is possible even when there are few patches in a given scene. Indeed the degree of constancy is comparable to the best which is theoretically possible.

Chapter 6

Color Constancy and a Changing Illumination

G.D. Finlayson. Color Constancy and a Changing Illumination. *Proceedings: Human Vision, Visual Processing and Digital Display V.* SPIE, The International Society for Optical Engineering, Vol. 2179: 352-363, 1994.

Abstract

The color constancy problem has proven to be very hard to solve. This is even true in the simple Mondriaan world where a planar patchwork of matte surfaces is viewed under a single illuminant. In this chapter we consider the color constancy problem given two images of a Mondriaan viewed under different illuminants.

We show that if surface reflectances are well-modelled by 3 basis functions and illuminants by up to 5 basis functions then we can, theoretically, solve for color constancy given 3 surfaces viewed under 2 illuminants. The number of recoverable dimensions in the illuminant depends on the spectral characteristics of the sensors. Specifically if for a given sensor set a von Kries type, diagonal model of color constancy is sufficient then we show that at most 2 illuminant parameters can be retrieved.

Recent work has demonstrated that for the human visual system a diagonal matrix is a good model of color constancy given an appropriate choice of sensor basis. We might predict therefore, that we can recover at most 2 illuminant parameters. We present simulations which indicate that this is in fact the case.

6.1 Introduction

Under different illuminants the same surface reflects different spectra of light; however, despite this we see the same color. This is the phenomenon of *color constancy*. Despite extensive research there does not yet exist a computational theory sufficient to explain the color constancy performance of a human observer.

A common starting point for color constancy research is the following question:

"Given an image of several surfaces viewed under a single illuminant how can we derive illuminant independent surface descriptors?"

This basic problem is tremendously hard to solve. Edwin Land's famous retinex theory [LM71, Lan86, Lan77] (and Hurlbert's [Hur89] subsequent extension) are easily shown to be inadequate for the task [BW86]. Forsyth's recent theory [For90] though more powerful requires many chromatically distinct surfaces in each scene.

Consequently the basic color constancy question is relaxed and computational theories often incorporate other factors into the problem formulation. Statistical analyses of reflectance and illuminant spectra are at the heart of many approaches [Mal85, MW86, Buc80, GJT88, FD88]. Funt and Ho [FH88] demonstrated that the chromatic aberration inherent in every lens can provide useful color information, Shafer [Sha85] provides a method of determining the illuminant color given specularities and Funt et al [FDH91] have shown that mutual reflection occurring at a concave edge ameliorates the color constancy problem.

Tsukada et al [TO90] and D'Zmura [D'Z92] have considered the color constancy problem where the illumination changes. Specifically they asked: "Given an image of several surfaces viewed under **two** illuminants how can we derive illuminant independent surface descriptors?"

This question is particularly relevant since Craven and Foster [CF92] have recently demonstrated that an illumination change is easily discernible and moreover can be distinguished from reflectance (i.e. false illuminant) changes. D'Zmura [D'Z92] has shown that if illuminants and surface reflectances are well described by finite-dimensional models each of 3 dimensions (the *3–3* world) then the color constancy

problem can be solved given 3 surfaces seen under two illuminants. D'Zmura and Iverson [DI93] generalized this result and have presented an algorithm which can solve for color constancy given many different assumptions—for example different numbers of reflectances or different model dimensions. However the results of the present chapter are not part of D'Zmura and Iverson's general theory.

In this chapter we begin by providing an alternate analysis for the *3–3* problem. Unlike D'Zmura's method our approach generalizes to more than 3 patches. Further, in the case where reflectances and illuminants are not precisely 3-dimensional our approach provides a least-squares criterion in solving for color constancy (D'Zmura's analysis is for exact models only). We extend our basic method and demonstrate that it is theoretically possible to solve for 5 parameters in the illuminant and 3 for reflectances given 3 patches seen under two illuminants. Under this generalized model every bijective linear map corresponds to a valid illumination change.

We show that there are classes of sensors for which the 2 illuminant, *3–3* color constancy problem **cannot be solved**. Specifically if for a given sensor set a diagonal matrix is a good model of illuminant change then the color constancy problem cannot be solved.

The world is not precisely *3–3* nor even *5–3* (in illuminant and reflectance dimension); therefore we must ask where our computational model will succeed in solving the color constancy problem. Using the human eye sensitivities we test our algorithm's color constancy performance through simulation experiments. We come to the following surprising conclusion: illumination change is useful in solving for color constancy if we assume the world is *2–3*; that is we assume the **illumination is 2-dimensional**. Moreover the greater the change in illuminant color the greater the likelihood of correctly solving for color constancy.

Assuming a 2-dimensional illumination has widespread implications. Firstly leading from the work of Maloney-Wandell [MW86] it is straightforward to show that the color constancy problem can be solved given a single patch under 2 illuminants[1]. Secondly color constancy is still possible even where a diagonal matrix is a good model of illumination change. This is reassuring since Finlayson et al [FDF93b] have shown

[1]This case is examined in isolation elsewhere [FF94].

that under the *2–3* assumptions there exists a sensor basis for all trichromatic visual systems for which a diagonal matrix precisely models all illuminant change. The analysis presented in this chapter therefore, serves to strengthen diagonal matrix theories of color constancy.

In section 6.2 we formulate the color constancy problem under changing illumination. Our solution method for both the *3–3* and *5–3* method is detailed in section 6.3. In section 6.4 given the human eye cone sensitivities we consider when the color constancy problem can in practice be solved. Section 6.5 presents simulations which demonstrate that color constancy can be solved by assuming a *2-3* model and given large color shifts in the illuminant.

6.2 The Color Constancy Problem

In keeping with D'Zmura we develop our solution method for the Mondriaan world: a Mondriaan is a planar, matte surface with several different, uniformly colored patches. Light striking the Mondriaan is assumed to be of uniform intensity and is spectrally unchanging. Each Mondriaan is assumed to contain at least 3 distinct surfaces.

If $E(\lambda)$ is the illuminant incident to surface reflectance $S(\lambda)$, where λ indexes wavelength, then the reflected color signal is equal to:

$$C(\lambda) = E(\lambda)S(\lambda) \qquad \text{(6.1: matte reflectance)}$$

The value registered by the kth cone to the color signal $C(\lambda)$ is defined by the integral integral equation (where ω denotes the visible spectrum):

$$\underline{p} = \int_{\omega} C(\lambda)R_k(\lambda)d\lambda \qquad \text{(6.2: color response)}$$

The illuminant, reflectance or sensitivity functions are known only for a set of sampled wavelengths (in this chapter all spectra are in the range 400nm to 650nm measured at 10nm intervals). Therefore the integral, of equation (6.2), is approximated as a summation.

6.2.1 Modelling Reflectance and Illuminant Spectra

Both illuminant spectral power distribution functions and surface spectral reflectance functions are well described by finite-dimensional models of low dimension. A surface reflectance vector $S(\lambda)$ can be approximated as:

$$S(\lambda) \approx \sum_{i=1}^{d_S} S_i(\lambda)\sigma_i \tag{6.3}$$

where $S_i(\lambda)$ is a basis function and $\underline{\sigma}$ is a d_S-component column vector of weights. Similarly each illuminant can be written as:

$$E(\lambda) \approx \sum_{j=1}^{d_E} E_j(\lambda)\epsilon_j \tag{6.4}$$

where $E_j(\lambda)$ is a basis function and $\underline{\epsilon}$ is a d_E dimensional vector of weights.

Maloney [Mal86] presented a statistical analysis of reflectance spectra and concluded that between 3 and 6 basis vectors are required to model surface reflectance. We will assume a 3-dimensional reflectance model. A similar analysis for daylight illumination was carried out by Judd [JMW64]; daylight illuminants are well represented by 3 basis vectors.

Given finite-dimensional approximations to surface reflectance, the color response eqn. (6.2) can be rewritten as a matrix equation. A *Lighting Matrix* $\Lambda(\underline{\epsilon})$ maps reflectances, defined by the $\underline{\sigma}$ vector, onto a corresponding color response vector:

$$\underline{p} = \Lambda(\underline{\epsilon})\underline{\sigma} \tag{6.5}$$

where $\Lambda(\underline{\epsilon})_{ij} = \int_\omega R_i(\lambda)E(\lambda)S_j(\lambda)d\lambda$. The lighting matrix is dependent on the illuminant weighting vector $\underline{\epsilon}$, with $E(\lambda)$ given by eqn. (6.4). The lighting matrix corresponding to the ith illuminant basis function is denoted as Λ^i.

6.2.2 Color constancy under 2 illuminants

Let us denote the 2 illuminants by the weight vectors $\underline{\epsilon}^1$ and $\underline{\epsilon}^2$. Reflectances are denoted as $\underline{\sigma}_1, \underline{\sigma}_2, \cdots, \underline{\sigma}_k$ where $k \geq 3$. Arranging the k sigma vectors as the columns of the matrix $\Omega_{1,2,\cdots,k}$ we can write the color responses under the two illuminants as:

$$\mathcal{P}^1 = \Lambda(\underline{\epsilon}^1)\Omega_{1,2,\cdots,k} \quad \mathcal{P}^2 = \Lambda(\underline{\epsilon}^2)\Omega_{1,2,\cdots,k} \tag{6.6}$$

where the ith column of $\mathcal{P}^j$ is the response of the ith surface under the jth illuminant ($j = 1$ or $j = 2$). Given $\mathcal{P}^1$ and $\mathcal{P}^2$ we want to solve for $\underline{\epsilon}^1$, $\underline{\epsilon}^2$ and $\Omega_{1,2,\cdots,k}$.

6.3 The Color Constancy Solution

We propose solving for color constancy in 2 steps. First we calculate the linear transform mapping the color responses between illuminants. This transform is independent of $\Omega_{1,2,\cdots,k}$ and, as we shall show, provides an elegant means of determining $\underline{\epsilon}^1$ and $\underline{\epsilon}^2$. By calculating $[\Lambda(\underline{\epsilon}^1)]^{-1}$, we can easily recover $\Omega_{1,2,\cdots,k}$.

In the *3–3* world lighting matrices have 3 rows and 3 columns. An implication of this is that color responses under one illuminant can be mapped to corresponding responses under a second illuminant by the application of a 3×3 linear transform; we call this an *illuminant map*:

$$\mathcal{M}^{1,2}\Lambda(\underline{\epsilon}^1) = \Lambda(\underline{\epsilon}^2) \ , \ \mathcal{M}^{1,2} = \Lambda(\underline{\epsilon}^2)[\Lambda(\underline{\epsilon}^1)]^{-1} \qquad \text{(6.7: illuminant map)}$$

Theorem 6.1 The color constancy problem can only be solved if for each pair of illuminants $s_1\underline{\epsilon}^1$ and $s_2\underline{\epsilon}^2$ (s_1, s_2 are scalars) the corresponding illuminant map $\mathcal{M}^{1,2}$ is unique, up to a scaling, over all other illuminant pairs.

Proof of Theorem 6.1: Assume we have 4 illuminants $\underline{\epsilon}^a, \underline{\epsilon}^b, \underline{\epsilon}^c$ and $\underline{\epsilon}^d$ such that $\mathcal{M}^{a,b} = \mathcal{M}^{c,d}$ where $\underline{\epsilon}^b$ and $\underline{\epsilon}^d$ are linearly independent. If $\Lambda(\underline{\epsilon}^a) = s\Lambda(\underline{\epsilon}^c)$ then Theorem 6.1 follows since b and d are, by assumption, linearly independent and consequently $\mathcal{M}^{a,b} \neq \mathcal{M}^{c,d}$. Otherwise let $\Omega_{1,2,3}$ denote a matrix of 3 reflectances. Let us define $\Omega^*_{1,2,3}$ such that:

$$\mathcal{P}^b = \Lambda(\underline{\epsilon}^b)\Omega_{1,2,3} \equiv \mathcal{P}^d = \Lambda(\underline{\epsilon}^d)\Omega^*_{1,2,3} \tag{6.8}$$

Thus the reflectances $\Omega_{1,2,3}$ viewed under illuminant $\underline{\epsilon}^b$ cannot be distinguished from the reflectances $\Omega^*_{1,2,3}$ viewed under illuminant $\underline{\epsilon}^d$. That is, the color constancy problem

cannot be solved if $\mathcal{M}^{1,2}$ is not unique over all other illuminant pairs. If $\mathcal{M}^{1,2}$ is not unique we cannot hope to separately recover both the $\underline{\epsilon}$ vectors (or at least the $\mathcal{M}^{1,2}$ matrix) as well as the $\underline{\sigma}$ vectors.

■

Let us assume that all pairs of illuminants, in our 3-dimensional span, have a corresponding unique illuminant map. Given the mapping $\mathcal{M}^{1,2}$ then:

$$\mathcal{M}^{1,2}[\Lambda^1\epsilon_1^1 + \Lambda^2\epsilon_2^1 + \Lambda^3\epsilon_3^1] \;=\; [\Lambda^1\epsilon_1^2 + \Lambda^2\epsilon_2^2 + \Lambda^3\epsilon_3^2] \tag{6.9}$$

Let $\mathcal{A}$ denote a 9×3 matrix where the ith column contains the basic lighting matrix Λ^i stretched out columnwise. Similarly let $\mathcal{B}$ denote the 9×3 matrix where the ith column contains $\mathcal{M}^{1,2}\Lambda^i$. Rewriting equation (6.9):

$$\mathcal{B}\underline{\epsilon}^1 \;=\; \mathcal{A}\underline{\epsilon}^2 \tag{6.10}$$

The columns of the matrix $\mathcal{A}$ are a basis for a 3-dimensional subspace of 9-space. Similarly the columns of $\mathcal{B}$ are a basis for a 3-dimensional subspace of 9-space. The solution of equation (6.10) is the intersection of these two spaces. The intersection is easily found by the method of principal angles [GvL83]. This method finds the vectors $\underline{\epsilon}^1$ and $\underline{\epsilon}^2$ which maximizes:

$$\cos^2(\theta) \;=\; \left(\frac{(\mathcal{Q}\underline{\epsilon}^1).(\mathcal{P}\underline{\epsilon}^2)}{|\mathcal{Q}\underline{\epsilon}^1||\mathcal{P}\underline{\epsilon}^2|} \right)^2 \tag{6.11}$$

That is, $\underline{\epsilon}^1$ and $\underline{\epsilon}^2$ are chosen such that the angle θ between $\mathcal{Q}\underline{\epsilon}^1$ and $\mathcal{P}\underline{\epsilon}^2$ is minimized. Thus even when there does not exist an exact solution to equation (6.10), the method of principal angles provides a least-squares criterion for returning the *best* answer.

6.3.1 Robust Color Constancy

While reflectances may in general be well described by a 3-parameter reflectance model a particular set of three reflectances may be poorly modelled. Consequently $\mathcal{M}^{1,2}$ will be incorrectly estimated. In this case the color constancy algorithm may return

incorrect estimates for the reflectance and illuminant parameters. However if the parameters of the illuminant map are derived from the observations of many (greater than three) distinct reflectances then we would expect improved color constancy.

Let $\mathcal{P}^1$ denote a $3 \times n$ matrix of n reflectances observed under an arbitrary illuminant. Similarly $\mathcal{P}^2$ is the $3 \times n$ matrix of observations of the same reflectances viewed under a second illuminant. The best illuminant map, in the least-squares sense, taking $\mathcal{P}^1$ onto $\mathcal{P}^2$ is defined by the Moore-Penrose inverse:

$$\mathcal{M}^{1,2}\mathcal{P}^1 \approx \mathcal{P}^2 \quad , \quad \mathcal{M}^{1,2} = \mathcal{P}^2(\mathcal{P}^1)^t[\mathcal{P}^1(\mathcal{P}^1)^t]^{-1} \tag{6.12}$$

In section 6.5 we present simulations where the illuminant map is derived first from 3 and then from 6 reflectances. The greater the number of reflectances the better the color constancy performance.

6.3.2 Solving for more Illuminant Parameters

The columns of $\mathcal{A}$ and $\mathcal{B}$ of equation (6.9) are bases for 3-dimensional subspaces of 9-space. It is possible, therefore, that $\mathcal{A}$ and $\mathcal{B}$ have a null-intersection—the combined 9×6 matrix $[\mathcal{A} \quad \mathcal{B}]$ has full rank; all columns are linearly independent. A null intersection is indicative of the fact that the world is not *3-3*. We might ask therefore, if it is possible to extend our model assumptions such that any illuminant map falls within our model.

The columns of $\mathcal{A}$ correspond to the three basis lighting matrices; the columns of $\mathcal{B}$ correspond to this basis transformed by an illuminant mapping. Clearly if we increase the dimension of the illuminant model from 3 to 5 then $\mathcal{A}$ and $\mathcal{B}$ become 9×5 matrices and are bases for 5-dimensional subspaces. In the 5-dimensional case the combined matrix $[\mathcal{A} \quad \mathcal{B}]$ has 9 rows and 10 columns. If the first 9 columns of $[\mathcal{P} \quad \mathcal{Q}]$ are linearly independent then they form a basis for 9-space. Consequently the 10th column is guaranteed to be linearly dependent on the first 9. Thus in the *5-3* case intersection is assured; indeed all bijective linear maps correspond to a valid illuminant mapping.

It is interesting to note that the *5-3* world does not belong to the general color constancy formulation of D'Zmura and Iverson [DI93]. Our work, therefore, supplements this general theory.

6.4 When color constancy can be solved

So far we have assumed that the illuminant map is unique, and consequently from Theorem 6.1, there exists a solution to the color constancy problem. If the world is *3-3* and the illuminant mapping is not unique then we show here that there exists a sensor basis such that all illuminant mappings are diagonal matrices.

Theorem 6.2 If the illuminant mapping is non-unique then there exists a sensor transformation $\mathcal{T}$, such that for all illuminants 1 and 2, $\mathcal{T}\mathcal{M}^{1,2}\mathcal{T}^{-1}$ is a diagonal matrix.

Proof of Theorem 6.2: Assume we have 4 illuminants $\underline{\epsilon}^a$, $\underline{\epsilon}^b$, $\underline{\epsilon}^c$ and $\underline{\epsilon}^d$ such that $\mathcal{M}^{a,b} = \mathcal{M}^{c,d}$ ($\mathcal{M}^{a,b} \neq \mathcal{I}$) and $\Lambda(\underline{\epsilon}^a) \neq s_1\Lambda(\underline{\epsilon}^c)$ and $\Lambda(\underline{\epsilon}^b) \neq s_2\Lambda(\underline{\epsilon}^d)$ (where s_1 and s_2 are scalars). Because we are assuming a *3-3* world, there are only 3 linearly independent lighting matrices and therefore we can choose scalars α, β and γ such that:

$$\Lambda(\underline{\epsilon}^a) + \alpha\Lambda(\underline{\epsilon}^b) \;=\; \beta\Lambda(\underline{\epsilon}^c) + \gamma\Lambda(\underline{\epsilon}^d) \tag{6.13}$$

Denoting the identity matrix as $\mathcal{I}$, we can write the mapping of illuminant a to illuminant c as:

$$\mathcal{M}^{a,c} \;=\; [\beta\mathcal{I} + \gamma\mathcal{M}^{c,d}]^{-1}[\mathcal{I} + \alpha\mathcal{M}^{a,b}] \tag{6.14}$$

Rewriting both $\mathcal{M}^{a,b}$ and $\mathcal{M}^{c,d}$, by assumption they are equal, as $\mathcal{T}^{-1}\mathcal{D}\mathcal{T}$ and $\mathcal{I}$ as $\mathcal{T}^{-1}\mathcal{I}\mathcal{T}$ equation (6.14) becomes:

$$\mathcal{M}^{a,c} \;=\; \mathcal{T}^{-1}[\beta\mathcal{I} + \gamma\mathcal{D}]^{-1}\mathcal{T}\mathcal{T}^{-1}[\mathcal{I} + \alpha\mathcal{D}]\mathcal{T} \tag{6.15}$$

$$\mathcal{M}^{a,c} \;=\; \mathcal{T}^{-1}[\beta\mathcal{I} + \gamma\mathcal{D}]^{-1}[\mathcal{I} + \alpha\mathcal{D}]\mathcal{T} \tag{6.16}$$

The lighting matrices $\Lambda(\underline{\epsilon}^a)$, $\Lambda(\underline{\epsilon}^b)$ and $\Lambda(\underline{\epsilon}^c)$ are linearly independent and span the space of lighting matrices; any lighting matrix can be written as a linear combination of these three. Since $\mathcal{M}^{a,b}$ and $\mathcal{M}^{a,c}$ have the same eigenvectors all illuminant mappings from a must have the same eigenvectors:

$$\mathcal{M}^{a,s_1 a+s_2 b+s_3 c} = \mathcal{T}^{-1}[s_1\mathcal{I} + s_2\mathcal{D}^{a,b} + s_3\mathcal{D}^{a,c}]\mathcal{T} \tag{6.17}$$

where s_1, s_2 and s_3 are scalars defining an arbitrary illuminant and $\mathcal{D}^{a,b}$ and $\mathcal{D}^{a,c}$ are the diagonal matrices of eigenvalues for $\mathcal{M}^{a,b}$ and $\mathcal{M}^{a,c}$ respectively. It is a simple step to show that all illuminant mappings share the same eigenvectors. Consider the illuminant mapping $\mathcal{M}^{s_1 a+s_2 b+s_3 c,t_1 a+t_2 b+t_3 c}$, where s_i and t_j are arbitrary scalars. Employing equation (6.17) we can write this as:

$$\mathcal{M}^{s_1 a+s_2 b+s_3 c,t_1 a+t_2 b+t_3 c} = \mathcal{T}^{-1}[t_1\mathcal{I} + t_2\mathcal{D}^{a,b} + t_3\mathcal{D}^{a,c}]\mathcal{T}\mathcal{T}^{-1}[s_1\mathcal{I} + s_2\mathcal{D}^{a,b} + s_3\mathcal{D}^{a,c}]^{-1}\mathcal{T} \tag{6.18}$$

which is equal to

$$\mathcal{M}^{s_1 a+s_2 b+s_3 c,t_1 a+t_2 b+t_3 c} = \mathcal{T}^{-1}[t_1\mathcal{I}+t_2\mathcal{D}^{a,b}+t_3\mathcal{D}^{a,c}][s_1\mathcal{I}+s_2\mathcal{D}^{a,b}+s_3\mathcal{D}^{a,c}]^{-1}\mathcal{T} \tag{6.19}$$

This completes the proof of Theorem 6.2.

■

Corollary 6.1 If the illuminant mapping is non-unique, then from each lighting matrix $\Lambda(\underline{\epsilon})$ the set of all matrices of the form $\mathcal{T}^{-1}\mathcal{D}\mathcal{T}$ exactly characterize the set of valid illuminant maps. The spaces spanned by $\mathcal{P}$ and $\mathcal{Q}$ intersect in all 3 dimensions.

Theorem 6.2 provides a useful test (see Figure 6.1) for determining whether a change in illumination adds extra information to the color constancy process. If the eigenvectors of the mapping taking the first lighting matrix to the second *is different* from the eigenvectors of the mapping taking the first to the third then a change of lighting *adds* information to the color constancy process.

$$\mathcal{T}\mathcal{M}^{1,2}\mathcal{T}^{-1} = \mathcal{D} \ , \ \mathcal{T}\mathcal{M}^{1,3}\mathcal{T}^{-1} \neq \mathcal{D}'$$

Figure 6.1: Color Constancy Check for *3-3* world: If check fails then color constancy is as hard under 2 illuminants as under a single illuminant.

The question of when a change in illumination is useful in solving for color constancy raises a paradox. Entrenched in color constancy research is the notion that if illumination change is well modelled by a diagonal matrix then it is *easier* to solve for color constancy and this is in direct conflict with what we have shown.

In the next section we present simulations which go some way to resolving this paradox. We examine our algorithm's performance given reflectances viewed under pairs of illuminants. The best color constancy is attained assuming *2-3* conditions, as opposed to *3-3* or *3-5*. Under *2-3* conditions illumination change is always perfectly modelled by a diagonal matrix (with respect to an appropriate basis) [FDF93b], and more importantly by losing one degree of freedom in the illuminant model, every diagonal matrix corresponds uniquely to a pair of illuminants; that is, color constancy is soluble.

6.5 Simulations

For our illuminants we chose a set of 7 Planckian black body radiators with correlated color temperatures 2000K, 2856K (CIE standard illuminant A [WS82]), 4000K,

6000K, 8000K, 10000K and 20000K. Reflectances are drawn from the set of 462 Munsell [Nic57] spectra. Illuminant bases of dimension 2, 3 and 5 are derived from an ensemble set of 14 illuminants containing the Planckian radiators, 5 daylight phases and CIE B and CIE C. A 3-dimensional reflectance basis is derived from the complete Munsell set.

We proceed in the following manner. First, 3 reflectances are randomly drawn from the set of 462 Munsell Spectra. For each pair of black-body radiators, $B_i(\lambda)$, $B_j(\lambda)$, we simulate the color response of the eye (equation (6.2)) using the cone fundamentals measured by Vos and Walraven [VW71]. We run our color constancy algorithm three separate times; using the 2-, 3- and 5-dimensional illuminant models. We record the recovery error as the angle between $B_j(\lambda)$ and that returned by our algorithm ($\underline{\epsilon}^2$ of equation (6.10)):

$$err \;=\; angle(B_j(\lambda), \sum_{k=1}^{d_E} E_k(\lambda)\epsilon_k^2) \tag{6.20}$$

As the recovered spectra better approximates the actual spectra so the error decreases towards zero. This experiment was repeated 5-times and the average angular error calculated for each illuminant pair.

The 2000K black-body radiator is a red biased spectrum, the 2856K radiator is still red but has a greater blue component. This trend continues with each of the 4000K, 6000K, 8000K, 10000K and 20000K radiators becoming progressively bluer. If the first illuminant is the 2000K spectra and the second 2856K then these spectra differ by 1 red-blue position. If the second illuminant is 4000K then this distance is 2. Subject to this red to blue ordering in our illuminant set we further average the error values. We calculate the average angular error given a red-blue difference of 1, 2, 3, 4, 5 and 6. We graph the color constancy performance, given these red-blue distances, for 2-, 3- and 5-dimensional illumination models in Figure 6.2.

For our illuminant set the average angular error between illuminants for red-blue distances of 1 and 2 is 11.7 and 23.4 degrees. Therefore if the error in recovering illumination is on the order of 10 degrees then the recovered spectra might be too red or too blue by a red-blue distance of 1. Similarly when angular error is on the order

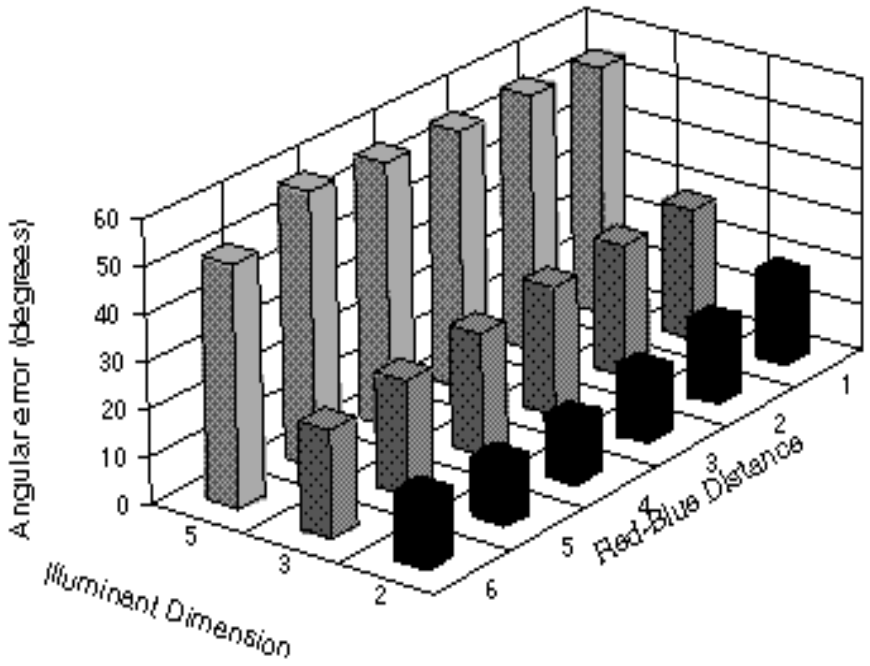

Figure 6.2: Randomly selected sets of 3 reflectances are imaged under 2 illuminants. The average angular error in the recovery of the second illuminant was calculated for illuminant models of 2-, 3- and 5-dimensions and red-blue distances of 1 through 6.

of 20 degrees the recovered spectra may be two units of red-blue distance from the correct answer. If recovery error is larger than this then the recovery procedure is not very effective since the same performance can be delivered by simple guess work. For example suppose we always guessed that the spectra is the 6000K black-body radiator then because its red-blue distance to all other spectra is less than or equal to three the recovery error of guessing will also be less than or equal to three.

In all cases the 2-dimensional assumption returns better color constancy. Moreover as the color constancy performance generally improves as the red-blue distance increases—a distance greater than 2 and the average angular error is less than or equal to 14 degrees. Under the 3-dimensional assumption the angular error is much

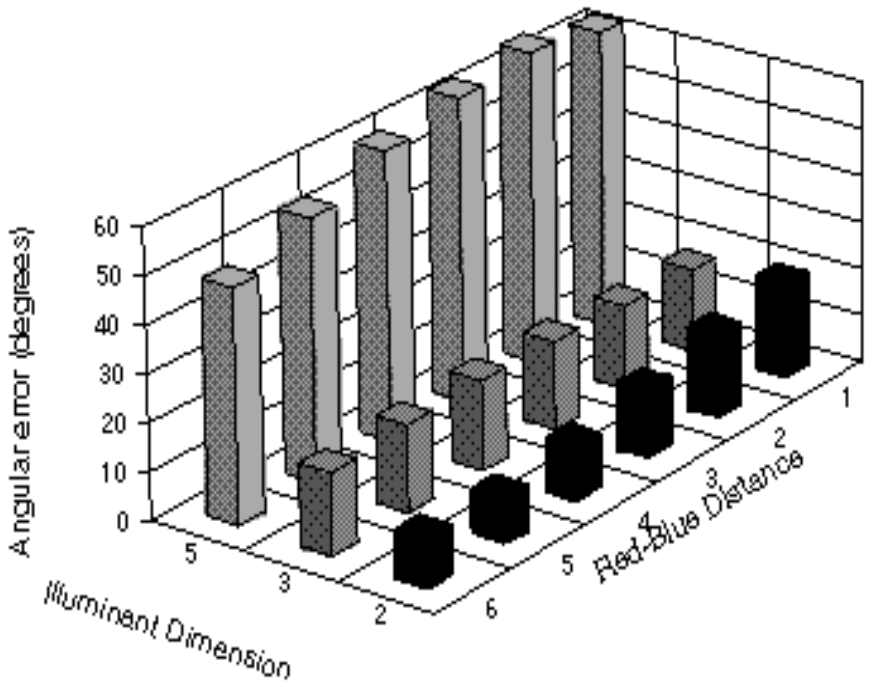

Figure 6.3: Randomly selected sets of 6 reflectances are imaged under 2 illuminants. The average angular error in the recovery of the second illuminant was calculated for illuminant models of 2-, 3- and 5-dimensions and red-blue distances of 1 through 6.

higher, on the order of 18 degrees throughout. However there is a discernible performance improvement given red-blue distances of greater than 4. The 5-dimensional assumption returns extremely poor color constancy with angular error always larger than 50 degrees.

We repeated this experiment for random selections of 6 patches. The results are graphed in Figure 6.3. Both the 2- and 3-dimensional assumptions show marked improvement; with the 2-dimensional assumption still supporting substantially better color constancy. As before the 5-dimensional assumption pays very poor dividends with the minimum angular error of 48 degrees. The error distribution is graphed in Figure 6.3.

That the 2-dimensional illuminant assumptions returns the best color constancy, at first glance, appears surprising. However previous simulations [MW92] have demonstrated that *2-3* assumptions provide a reasonable model for approximating cone responses. Moving to higher dimensional illumination models improves, slightly, on this approximation but at the cost of introducing very many more valid illuminant mappings. As the number of mappings increases so does the likelihood of a false match.

6.6 Conclusion

We have developed a computational framework for solving for color constancy under a change of illuminant. The framework is general in the sense that the computation remains the same under different illuminant model assumptions.

We derived a test to determine whether a change in illumination adds new information to the color constancy problem given a 3-dimensional illuminant. If a diagonal matrix, with respect to a sensor basis, is a good model of illuminant change then a change of illumination does not add new information. This is paradoxical in that it contradicts the established view that diagonal matrix color constancy is *easier* than non-diagonal methods.

Simulation experiments go some way to resolving this paradox. We show that a 2-dimensional illuminant assumption supports better color constancy than 3- or 5-dimensional assumptions. A 2-dimensional assumption is completely consistent with diagonal theories of color constancy.

Chapter 7

Contributions and future research directions

In this chapter I summarize the central contributions presented in my thesis. I also place this study in the wider context and suggest possible applications in other fields.

7.1 Contributions

1. I generalize the coefficient image model by allowing a sensor transformation prior to the application of scaling coefficients.

   Traditional coefficient theories of color constancy are based on the idea that an illumination change induces simple scalings of the color responses. Specifically if (r, g, b) is a color response for a surface viewed under one illuminant and (r', g', b') is the color response for the same surface viewed under a second illuminant then the two are related by simple coefficient scalings: $r' = \alpha r$, $g' = \beta g$ and $b' = \gamma b$ (where α, β, and γ are scalars). In the coefficient model the same scaling coefficients map the color responses for all surfaces between illuminants.

   Under the generalized coefficient scheme color responses are transformed to a new sensor basis prior to applying scaling coefficients. Each *coefficient channel* in the new basis is a linear combination of the original color response functions.

I demonstrated, in simulation experiments, that the coefficient model performs reasonably poorly for the human cones, the XYZ color matching curves, or the color-opponent type channels. In contrast when illumination change is modelled by a 3×3 matrix transform (i.e. the general image model) performance is excellent. A central goal of this thesis is to find coefficient channels which will enhance coefficient model performance; hopefully, to a level comparable to the general image model.

2. Several sets of coefficient channels are derived by different means. These are listed below where a short overview of their derivation is given.

   Sensor-based spectrally sharpened sensors. It is well known that the coefficient image model exactly characterizes color response for sensors which are sensitive to a single wavelength of light. Sensor-based spectral sharpening is a technique for determining the linear combination of sensors that is maximally sensitive (or narrow-band) for any interval over the visible spectrum. Coefficient image performance with sensor-based spectrally sharpened sensors is similar to general image model performance.

   Data-based Spectrally Sharpened sensors. Using the statistics of simulated color responses, data-based sharpening finds the optimal (with respect to a least-squares criterion) coefficient channels for any pair of illuminants. Given many illuminant pairs, the aggregate sensors are returned. Data-based sharpening markedly improves coefficient image model performance.

   Perfectly Sharpened sensors. Sensor-based sharpening and data-based sharpening operate for a given set of sensors. Perfect sharpening on the other hand is an algebraic argument defining the conditions that reflectance and illumination spectra must satisfy such that **every** set of trichromatic sensors has a basis where the coefficient model exactly describes color response. If illumination and reflectance are well described by 2- and 3-dimensional linear models then this is a sufficient condition for perfect sharpening. That daylight illumination and natural reflectances are reasonably described by such models indicates that the coefficient model will be approximately adequate regardless of the sensor set.

Generalized diagonal sensors. I show that when the perfect sharpening conditions are reversed—illumination and reflectance are described by 3- and 2-dimensional linear models— there still exists a sensor basis (though not the same one) where the coefficient model exactly characterizes color response.

Because these *3–2* conditions are at the foundation of many recovery algorithms it follows that they are all coefficient theories of color constancy. Operating in coefficient mode their exposition and implementation is simplified. I describe, in detail, how the Maloney-Wandell algorithm can be implemented from a coefficient perspective.

Maximum ignorance sensors. Sensor-based spectral sharpening is a method for finding linear combinations of the cones which appear narrow-band. Maximum ignorance sensors also aim at narrow-bandedness but from a different perspective. The aim here is to find the linear combination of the cones (or other sensors) which sample color signal spectra like a narrow-band set. If this goal can be achieved, coefficient color constancy must follow since the coefficient image model is perfect for narrow-band sensors.

The maximum ignorance sensors are the linear combination of the cones that operate most like a narrow-band set when all possible color signal spectra appear with equal likelihood. Maximum ignorance sensors markedly improve coefficient image model performance. The maximum ignorance assumption, while unrealistic (all color signals are not equally likely) has the advantage of simplifying computation since maximum ignorance sensors can be found in linear time.

Calibration sensors. Calibration sensors are derived in much the same way as maximum ignorance sensors. The only difference is that a calibration set of real color signal spectra is used. For a set of 426 Munsell reflectances [Nic57] illuminated by D48, D55, D65, D75, D100 daylights [JMW64] and CIE A [WS82] the derived calibration channels support good coefficient performance. Moreover, performance is significantly better than for the maximum ignorance sensors. However, increased performance is bought at the price of a computationally more expensive derivation. The computation is cubic for a trichromatic sensor

set.

3. Perspective colors are proposed as a suitable representation for investigating the color constancy problem. If (r, g, b) is a color response triple then $(r/b, g/b)$ is its corresponding perspective color (I use the term "perspective" since dividing by the third coordinate is analogous to the perspective projection used in computer graphics).

   In perspective, intensity information is factored out of the problem formulation. It follows that image processes which confound correct intensity recovery are of no consequence; these processes include specularities, shape and varying illuminant power. Thus, in theory, color constancy becomes possible in realistic world scenes.

4. I demonstrate that Forsyth's [For90] coefficient-based gamut-mapping recovery algorithm, CRULE, is an elegant vehicle for solving the perspective color constancy problem—if the coefficient model accurately models response vectors it will accurately model perspective colors as well. Furthermore, I show that problems inherent to CRULE are in large part mitigated. By discarding intensity recovery, color constancy becomes possible in real world scenes (Forsyth's CRULE runs in to difficulties in all but the simplest of worlds). Moreover in perspective, CRULE can be implemented with reduced computational complexity. Indeed while Forsyth's implementation of CRULE returns only an approximate answer, my 2-dimensional perspective calculation is exact and operates in a timely manner.

5. An illumination gamut restriction is added to the CRULE formulation. I demonstrate that this is an extremely powerful constraint in solving for color constancy. Moreover while the addition of the illumination gamut is straightforward in perspective, this is not true for the 3-dimensional CRULE. I call the recovery procedure which combines perspective CRULE with the illumination gamut constraint *color in perspective*.

6. **Color in perspective, when tested on real images, successfully and accurately recovered color descriptors. Moreover there was good recovery even when there was little color complexity in an image.** Most other recovery algorithms have not been tested on real images. Indeed many algorithms would appear not to have been tested at all. A key component of the recovery success is the new illumination gamut constraint.

7. I examine the multi-illuminant color constancy problem. Specifically I characterize the conditions under which viewing a scene under more than one illuminant can add information useful in descriptor recovery. I show that if there exists a sensor basis with respect to which the coefficient model accurately describes color response then no information is added. The only exception to this is when illumination is well described by a 2-dimensional linear model. In this case the generalized coefficient model holds for all sensor sets.

8. Given that there exist linear combinations of the cones with respect to which a coefficient image model performs well, I predict that at most 2 parameters of the illuminant can be successfully recovered. Experiments corroborate this prediction. With a 2-dimensional linear model assumption, some color constancy is possible. Assuming 3- (or higher) dimensional models results in poor recovery.

9. Theoretical and experimental comparisons of the various coefficient channels are given. Sensor-based sharpening is shown to be a theoretically sound method and is closely related to the derivation of maximum ignorance sensors.

   Experimentally, while all sets of coefficient channels enhance coefficient performance, they do so to different degrees. Performance is contingent on the accuracy of the underlying statistical models—the more accurate the statistics the better the performance. Data-based sharpening derived using simulated color response vectors performs best overall.

10. I have split the color constancy problem into two related parts: the image model and the recovery algorithm. The image model describes the interdependence of light, surface and sensor response. Taking illumination-dependent

sensor responses as input, the recovery algorithm attempts to extract illuminant-independent descriptors. The precise nature of recovery computation is defined by the image model.

An image model with few parameters may poorly account for sensor responses. A poor image model effectively limits the efficacy of any recovery procedure since descriptors can only be recovered up to the accuracy of the model. In contrast, while an image model with many parameters may accurately describe color response, it also necessitates a complex recovery procedure—there are many more parameters to recover.

11. To evaluate image model performance I developed the normalized fitted distance (or NFD) measure. This measure is applied in two stages. First, illuminant-dependent color observations are mapped (in a manner defined by the image model) to their corresponding illumination-independent descriptors with minimum least-squares error. At the second stage individual descriptors are compared with fitted observations. The NFD measure is defined as the Euclidean distance between fitted observation and descriptor normalized to the descriptor length.

7.2 Applications

1. *Color Graphics.* Color graphics is the inverse of color vision. The latter tries to derive scene properties from an image of the world whereas the former attempts to create an image from known scene properties. Of course both disciplines require a good image model. A poor image model bounds the accuracy of color recovery in color vision and limits the realism of color rendering in color graphics.

   In creating images it is, in principle, possible to use very accurate image models. Indeed full spectral data can be used. However, because image rendering is an exceptionally expensive computational [FvD90] task the simple coefficient image model is generally employed within the graphics community [Bor91]. However, the cone sensors are not used, rather, the coefficient channels are either the

XYZ color matching curves [WS82] or matching curves derived for monitor primaries [Bor91]. Neither of these color spaces are good choices for the coefficient image model. This has led Peercy [Pee93] to develop a many-parameter (many more parameters than the coefficient model) rendering scheme based on linear models of illumination and reflectance.

I propose that the coefficient channels derived in this thesis will support more realistic rendering for color graphics. Indeed I believe that, at least for simple scenes, they will be sufficiently accurate to circumvent Peercy's more complex method.

2. *Color balancing.* If the colors on a monitor or television "look" wrong then we actively change them. Color manipulation is either direct in that the gains of the beams for the three phosphors are independently adjusted, or indirect via changes in contrast and brightness. Both classes of manipulation operate across the image as a whole and in this sense simulate a changing illuminant color. If the colors in an image appear too red, as would be the case for an image taken under red light, then we would make them more blue thereby balancing the colors.

   Theoretical and experimental results presented in this thesis demonstrate that color balancing operating directly either on the cones, or on XYZ color matching functions, or on opponent channels, cannot correspond to physically plausible changes in illumination. If our goal is to correct for the illuminant, traditional balancing can have limited performance. In contrast color balancing relative to the coefficient channels corresponds precisely to physically plausible illumination changes. I propose the coefficient channels as the basis for color manipulation on display devices.

3. *Color reproduction: white point matching.* Color reproduction is the problem of taking colors displayed on one medium and correctly redisplaying them on another. For example a graphics artist may create a picture on a color monitor and then print a hard-copy on a color printer. Color reproduction is the art of

getting the printer to output colors which look the same.

The basic problem has proven very hard and is the source of much research (for example see [MB93]). As a stepping stone to accurate reproduction, the monitor-printer system (or other combination of devices) is often calibrated to reproduce white correctly. This can be achieved if the coordinates of white on each device are the same and each device outputs the same colors for the white coordinate. To fix the coordinate systems, device coordinates relative to white are used. One simply divides each color coordinate by the coordinate of white: e.g. if (r, g, b) is a device coordinate then relative to white the coordinate becomes $(\frac{r}{r_w}, \frac{g}{g_w}, \frac{b}{b_w})$ where (r_w, g_w, b_w) is the device coordinate of white. Crucially, different media have different white-points.

If device coordinates are defined relative to the cone basis or XYZ basis (the latter has become the standard) then correctly taking white across media will not guarantee the correct mapping of other colors. If, however, device coordinates are defined relative to coefficient channels, correctly accounting for white should also correctly account for most other colors.

4. *Color science: perceptually uniform space.* The eye adapts to the intensity of incoming light. Moreover, if a stimulus entering the eye is reddish the eye will become more red adapted. That is, adaptation depends on the color of the stimulus light.

   Therefore in quantifying the *appearance* of colored stimuli it is standard practice to include the color of the adapting light. The CIE Lab formulae attempt to predict how similar two test stimuli appear relative to some adaptation state. If a stimuli has an XYZ coordinate (x, y, z) and the adapting illuminant is (x_n, y_n, z_n) then the adapted tristimulus, the input to the CIE Lab formulae is $(\frac{x}{x_n}, \frac{y}{y_n}, \frac{z}{z_n})$. The input is effectively a white-point-relative coordinate system (see white-point matching above).

   By dividing by the coordinates of the adapting light coordinates it is hoped that the dependence of the test stimulus on the adapting light will be removed.

Of course as I have demonstrated, scaling the XYZ sensor basis can only approximately factor out illuminant color. If instead color coordinates are defined relative to the coefficient channels then, in principle, the effect of the illuminant can be removed with greater accuracy. I speculate that color appearance formulae defined in terms of adapted coefficient channel responses might better predict the psychophysical data.

5. *Color Machine Vision.* Because the color constancy problem is hard to solve, various authors have sought to extract illuminant independent color invariants from an image, instead of color descriptors. For example, under the coefficient image model, the ratio of color observations, effectively factors out the illumination dependent information producing a useful invariant. For example if $(\alpha r_1, \beta g_1, \gamma b_1)$ and $(\alpha r_2, \beta g_2, \gamma b_2)$ are two color observations where (α, β, γ) are the scaling coefficients to the corresponding descriptors then $(\frac{r_1}{r_2}, \frac{g_1}{g_2}, \frac{b_1}{b_2})$ is illumination independent.

   The more accurate the coefficient model the more accurate the ratio invariants. One would predict therefore that invariants calculated for coefficient channels would be more stable and hence more useful. Indeed it has been shown that *color constant color indexing* [FF95], a ratio-based object recognition technique, delivers only partial recognition with cone sensors but almost perfect recognition with coefficient channels.

7.3 Conclusion

By factoring a change of sensor basis into the coefficient image model I have shown that simple coefficient scalings suffice in solving for color constancy. I have developed a new coefficient color constancy algorithm called *color in perspective* which successfully solves the color constancy problem for real color images. In principal color constancy remains possible in the presence of specularities, varying illuminant power and shape. Finally I have shown that a changing illumination adds no new information to the color constancy problem. The only exception to this is when illuminants are quite

restricted; in this case the coefficient image model always suffices.

Appendix A

Image Performance in practice

In this thesis image model performance is quantified in two stages. First observations are mapped to descriptors with minimum error, where the form of the mapping is defined by the image model and error is defined as the sum of squared differences. At the second stage the Euclidean distance between each descriptor and fitted observation, normalized to the descriptor length, is calculated. The cumulative histogram of normalized fitted distances (or NFDs) is an effective vehicle for comparing image model performance.

While it is clear from NFD cumulative histograms that some models are better than others, it is not clear how this translates into visual performance. Is it significant that while the general linear model has NFD errors (Figure 1.6) always less than 4% the coefficient model (operating directly on the cones) can have greater than 10% error? This question is hard to answer for machine vision and would depend largely on the task at hand.

Suppose that some descriptor is the 3-tuple (200,200,200) (i.e. a white) and a fitted observation (220,220,220) then the NFD error is 10%. Assuming that a camera records discretely in the range 0 to 256, an error of 20 accounts for about 8% of the signal range. Such a high error rate would adversely affect any application demanding stable color descriptions, e.g. recognizing objects based on their color distributions [SB90].

A 4% error corresponds to a fitted observation of (208,208,208) but a pixel difference of 8 accounts for only about 3% of the signal range. Thus performance is

much better and this may be crucial. Indeed *color constant color indexing*, a recognition scheme based on the coefficient model of color constancy, delivers only partial recognition when the cones are used. Switching to the generalized coefficient model, thereby attaining general image model performance, results in almost perfect recognition [FF95].

In the context of human vision there are numerical formulae which can fairly accurately predict the perceptual difference between stimuli. In particular for a pair of XYZ color observations (the XYZ sensor basis is graphed in Figure 1.5) the CIE Lab formulae [WS82] provide a good estimate of how similar they would appear to a human observer. A Lab distance of one corresponds to a just noticeable perceptual difference. A distance of two is twice as noticeable. In practice a CIE Lab error of less than 2 is not noticeable [SFB92] for complex scenes.

To test the adequacy of an image model relative to the Lab perceptual measure I carried out the following experiment. I calculated the XYZ observations for the 462 Munsell spectra [Nic57] under 6 illuminants: 5 daylight phases (D48, D55, D65, D75 and D100) [JMW64] and CIE A [WS82]. Choosing D55 as the canonical illuminant, I then calculated the mappings taking observations to descriptors for three image models: the general linear model, the coefficient model operating on the cones, and the coefficient model operating on data-based sharpened sensors derived in chapter 2 (see Figure 2.1). Fitting the data for the coefficient models is carried out in two stages. First the XYZ observations are mapped to the coefficient basis (cones or data-based sharpened sensors) then the best scaling coefficients are determined and applied. The mapped observations are then transformed back to the XYZ basis. The reader is referred to chapter 2 for a more detailed discussion of this fitting procedure.

Rather than calculate NFD error between fitted observations and descriptors, I apply the CIE Lab formulae[1]. In this way I get a measure of the perceptual difference between fitted observations and descriptors. The cumulative histograms of CIE Lab error for each of the three image models are shown in Figure A.1.

[1]The CIE Lab formulae require the specification of illuminant. For both descriptors and fitted observations I use D55.

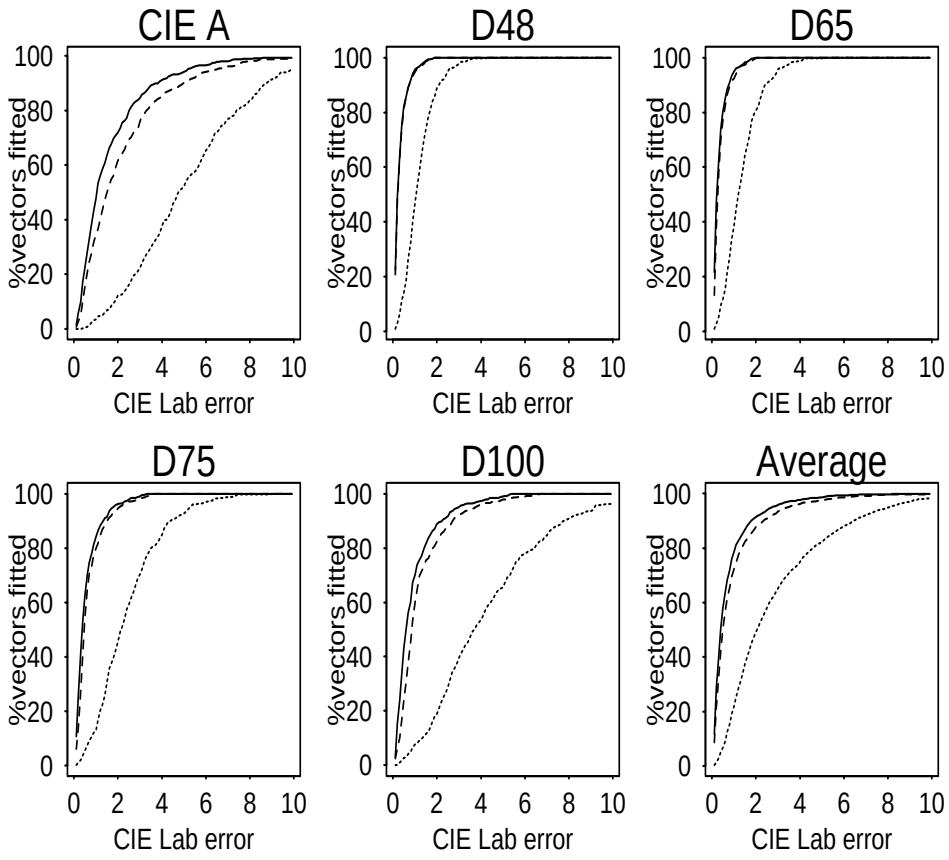

Figure A.1: Cumulative CIE Lab error histograms obtained for 5 test illuminants (CIE A, D48, D55, D65, D75 and D100) for general linear model (solid line), coefficient color constancy operating on the cones (dotted lines) and coefficient color constancy operating on data-based sharpened sensors (dashed lines).

On average both the general linear model and coefficient model operating on data-based sharpened cones map most observations with less than 2 CIE Lab error—an error small enough to go unnoticed by a human observer. In contrast coefficient color constancy operating directly on the cones incurs a much higher error. Most of the observations have larger than 2 CIE Lab error (thus they are visually noticeable) and some are as high as 10 (very noticeable).

In comparing Figure A.1 with cumulative histograms of NFD error (seen throughout the thesis) we see that they are qualitatively very similar. The NFD histogram curves ranging from 0 to 10% error are almost identical to the CIE Lab histogram curves ranging from 0 to 10 just noticeable differences. That there should be such a correspondence was quite unexpected.

People respond to color stimuli in a non-linear manner and CIE Lab equations were designed to account for this. For example a pair of green observations which are a little apart may look identical to a human observer. In contrast a pair of red observations similarly apart may appear quite different. The NFD error would not discriminate between the two circumstances. However, in looking at error over the entire set it appears that the linear NFD metric and non-linear CIE Lab metric are quantitatively quite similar. A NFD error of 1% on average correlates with a CIE Lab error of one just noticeable difference.

Bibliography

[Bar82] H.B. Barlow. What causes trichromacy? A theoretical analysis using comb-filtered spectra. *Vision Research*, 22:635–643, 1982.

[Bec72] J. Beck. *Surface Color Perception.* Cornell University Press, 1972.

[BF94] D.H. Brainard and W.T. Freeman. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *Human Vision, Visual Processing, and Digital Display V*, pages 364–376. SPIE, February 1994.

[BG84] G. Buchsbaum and A. Gottschalk. Chromaticity coordinates of frequency-limited functions. *J. Opt. Soc. Am. A*, 1(8):885–887, 1984.

[BH80] J.K. Bowmaker and H.J.A.Dartnell. Visual pigments of rods and cones in the human retina. *J. Physiol.*, 298:501–511, 1980.

[Bla85] A. Blake. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics, and Image Processing*, 32:314–327, 1985.

[Bor91] C. Borges. Trichromatic approximation for computer graphic illumination models. *Computer Graphics*, 25:101–104, 1991.

[BP88] R.S. Berns and K.H. Petersen. Empirical modelling of systematic spectrophotometric errors. *Color Res. Appl.*, (4):243, 1988.

[Bre89] G.J. Brelstaff. *Inferring Surface Shape from Specular Reflections.* PhD thesis, University of Edinburgh, Department of Computer Science, 1989.

[Bri78] M.H. Brill. A device performing illuminant–invariant assessment of chromatic relations. *J. Theor. Biol.*, 71:473–478, 1978.

[Bri80] M.H. Brill. Computer-simulated object-color recognizer. Technical Report 122, M.I.T. Research Laboratory of Electronics, 1980.

[Bri93] M.H. Brill. Can color-space transformations improve color computations other than Von Kries? In *Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 485–492. SPIE, 1993. Feb. 1–4.

[Buc80] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.

[Buc87] G. Buchsbaum. Color signal coding: Color vision and color television. *COLOR research and application*, 12(5):266–269, 1987.

[BW82] M.H. Brill and G. West. Constancy of munsell colors under varying daylight conditions. *DIE FABRE*, 30(1):65–68, 1982.

[BW86] D.A. Brainard and B.A. Wandell. Analysis of the reinex theory of color vision. *J. Opt. Soc. Am. A*, 36:1651–1661, 1986.

[BWC89] D.A. Brainard, B.A. Wandell, and W.B. Cowan. Black light: how sensors filter spectral variation of the illuminant. *IEEE Trans. Biomed. Eng.*, 36:140–149 and 572, 1989.

[CF92] B.J. Craven and D.H. Foster. An operational approach to colour constancy, July 1992.

[Cha93] F. Chatelin. *Eigenvalues of Matrices.* Wiley, 1993.

[Coh64] J. Cohen. Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci.*, 1:369–370, 1964.

[DI93] M. D'Zmura and G. Iverson. Color constancy. I. basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *J. Opt. Soc. Am. A*, 10:2148–2165, 1993.

[DI94] M. D'Zmura and G. Iverson. Probabalistic color constancy. In R.D. Luce, M. M. D'Zmura, D. Hoffman, G. Iverson, and K. Romney, editors, *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*. Laurence Erlbaum Associates, 1994.

[DL86] M. D'Zmura and P. Lennie. Mechanisms of color constancy. *J. Opt. Soc. Am. A*, 3:1662–1672, 1986.

[Dre93] M. S. Drew. Optimization approach to dichromatic images. *Journal of Mathematical Imaging and Vision*, 3:189–205, 1993.

[D'Z92] M. D'Zmura. Color constancy: surface color from changing illumination. *J. Opt. Soc. Am. A*, 9:490–493, 1992.

[Eas81] Eastman Kodak company. *Kodak filters : for scientific and technical uses*, 2nd edition, 1981.

[FB93] M. Fairchild and R. Berns. Image color-appearance specification through extension of CIELAB. *COLOR research and application*, 18(3):178–190, 1993.

[FD88] B.V. Funt and M.S. Drew. Color constancy computation in near–Mondrian scenes using a finite dimensional linear model. In *Computer Vision and Pattern Recognition Proceedings*, pages 544–549. IEEE Computer Society, June 1988.

[FDF93a] G.D. Finlayson, M.S. Drew, and B.V. Funt. Diagonal transforms suffice for color constancy. In *Proccedings of the fourth International Conference on Computer Vision*. IEEE Computer Society & European Vision Society, May 1993.

[FDF93b] G.D. Finlayson, M.S. Drew, and B.V. Funt. Enhancing von Kries adaptation via sensor transformations. In *Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 473–484. SPIE, 1993. Feb. 1–4.

[FDF94a] G.D. Finlayson, M.S. Drew, and B.V. Funt. Color constancy: Generalized diagonal transforms suffice. *J. Opt. Soc. Am. A*, 11:3011–3020, 1994.

[FDF94b] G.D. Finlayson, M.S. Drew, and B.V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *J. Opt. Soc. Am. A*, 11(5):1553–1563, May 1994.

[FDH91] B.V. Funt, M.S. Drew, and J. Ho. Color constancy from mutual reflection. *Int. J. Computer Vision*, 6:5–24, 1991.

[FF94] G.D. Finlayson and B.V. Funt. Color constancy with shadows. *Perception*, 23:89–90, 1994. Special issue on the 17th European Conference on Visual Perception, Eindhoven.

[FF95] B.V. Funt and G.D. Finlayson. Color constant color indexing. *IEEE transactions on Pattern analysis and Machine Intelligence*, 1995. to appear.

[FH88] B.V. Funt and J. Ho. Color from black and white. In *Proceedings of the Second International Conference on Computer Vision*, pages 2–8. IEEE Computer Society, December 1988.

[For90] D. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5:5–36, 1990.

[Fos81] D.H. Foster. Changes in field spectral sensitivities of red-,green- and blue-sensitive colour mechanisms obtained on small backround fields. *Vision Research*, 21:1433–1455, 1981.

[Fos84] D.H. Foster. Colour vision. *Contemporary Physics*, 25:477–497, 1984.

[FS83] D.H. Foster and R.S. Snelgar. Initial analysis of opponent-colour interactions revealed in sharpened field sensitivities. In J.D. Mollon and L.T.

Sharpe, editors, *Colour Vision : Physiology and Psychophysics*, pages 303–312. Academic Press, 1983.

[FvD90] J.D. Foley and A. van Dam. *Fundamentals of interactive computer graphics*. Addison-Wesley, 1990. 2nd edition.

[GJT88] R. Gershon, A.D. Jepson, and J.K. Tsotsos. From $[r, g, b]$ to surface reflectance: Computing color constant descriptors in images. *Perception*, pages 755–758, 1988.

[GvL83] G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins U. Press, 1983.

[HF83] D.C. Hood and M.A. Finkelstein. A case for the revision of textbook models of colour vision. In J.D. Mollon and L.T. Sharpe, editors, *Colour Vision : Physiology and Psychophysics*, pages 385–398. Academic Press, 1983.

[HKB83] W. Jaeger H. Krastel and S. Braun. An increment-threshold evaluation of mechanisms underlying colour constancy. In J.D. Mollon and L.T. Sharpe, editors, *Colour Vision : Physiology and Psychophysics*, pages 545–552. Academic Press, 1983.

[Hor74] B.K.P. Horn. Determining lightness from an image. *Computer Vision, Graphics, and Image Processing*, 3:277–299, 1974.

[Hur86] A. Hurlbert. Formal connections between lightness algorithms. *J. Opt. Soc. Am. A*, 3:1684–1692, 1986.

[Hur89] A.C. Hurlbert. *The Computation of Color*. PhD thesis, MIT Artificial Intelligence Laboratory, 1989.

[JMW64] D.B. Judd, D.L. MacAdam, and G. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, 54:1031–1040, August 1964.

[KH90] M. Kalloniatis and R.S. Harwerth. Spectral sensitivity and adaptation characteristics of cone mechanisms under white-light adaptation. *J. Opt. Soc. Am. A*, 7:1912–1928, 1990.

[KKOP66] D.L. Kreider, R.G. Kuller, D.R. Ostberg, and F.W. Perkins. *An introduction to linear analysis.* Addison Wesley, 1966.

[Kri47] E.L. Krinov. Spectral reflectance properties of natural formations. *Technical Translation TT-439, National Research Council of Canada*, 1947.

[Lan77] E.H. Land. The retinex theory of color vision. *Scientific American*, pages 108–129, 1977.

[Lan83] E.H. Land. Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proc. Natl. Acad. Sci*, 80:5163–5169, 1983.

[Lan86] E.H. Land. Recent advances in retinex theory. *Vision Res.*, 26:7–21, 1986.

[LBS90] H. Lee, E. J. Breneman, and C. P. Schulte. Modelling light reflection for computer color vision. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 4:402–409, 1990.

[Lee90] H. Lee. Illuminant color from shading. In *Perceiving, Measuring and Using Color*, volume 1250. SPIE, 1990.

[LM71] E.H. Land and J.J. McCann. Lightness and retinex theory. *J. Opt. Soc. Amer.*, 61:1–11, 1971.

[Luc93] Marcel Lucassen. *Quantative Studies of Color Constancy.* PhD thesis, Utrecht University, 1993.

[Mal85] L.T. Maloney. *Computational Approaches to Color Constancy.* PhD thesis, Stanford University, Applied Psychology Lab, 1985.

[Mal86] L.T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am. A*, 3:1673–1683, 1986.

[Mal90] L.T. Maloney. Photoreceptor spectral sensitivities and color correction. In M.H. Brill, editor, *Perceiving, Measuring, and Using Color*, volume 1250, pages 103–110. SPIE, 1990.

[MB93] R.J. Motta and H.A. Berberian, editors. *Device-independent color imaging and imaging systems integration*, volume 1909. SPIE, 1993.

[McC94] J.J. McCann. Adaptation or contrast: The controlling mechanism for colour constancy. In *Proceedings of the John Dalton Conference on Colour Vision*, Sep 1994.

[MMD76] C.S. McCamy, H. Marcus, and J.G. Davidson. A color-rendition chart. *J. App. Photog. Eng.*, pages 95–99, 1976.

[MW86] L.T. Maloney and B.A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A*, 3:29–33, 1986.

[MW92] D.H. Marimont and B.A. Wandell. Linear models of surface and illuminant spectra. *J. Opt. Soc. Am. A*, 9(11):1905–1913, 92.

[Nic57] D. Nickerson. Spectrophotometric data for a collection of munsell samples, 1957.

[NNJ43] Newhall, Nickerson, and D.B. Judd. Final report of the OSA subcommittee on the spacing of the munsell colors. *J. Opt. Soc. Am. A*, 33:385–418, 1943.

[NS90] C.L. Novak and S.A. Shafer. Supervised color constancy using a color chart. Technical Report CMU-CS-90-140, Carnegie Mellon University School of Computer Science, 1990.

[NSW90] C.L. Novak, S.A. Shafer, and R.G. Wilson. Obtaining accurate color images for machine vision research. In M.H. Brill, editor, *Perceiving, Measuring, and Using Color*, volume 1250, pages 54–68. SPIE, 1990.

[Pee93] Mark S. Peercy. Linear color representations for full spectral rendering. *Computer Graphics*, 27:191–198, August 1993.

[PHJ89a] J. P. S. Parkkinen, J. Hallikanen, and T. Jaaskelainen. Characteristic spectra of munsell colors. *Journal of the Optical Society of America A*, 6:318–322, 1989.

[PHJ89b] J.P.S. Parkkinen, J. Hallikainen, and T. Jaaskelainen. Characteristic spectra of munsell colors. *J. Opt. Soc. Am. A*, 6:318–322, 1989.

[PS85] F.P. Preparata and M.I. Shamos. *Computational Geometry: an introduction*. Springer Verlag, 1985.

[PW90] A.B. Poirson and B.A. Wandell. Task-dependent color discrimination. *J. Opt. Soc. Am. A*, 7:776–782, 1990.

[RS89] J. Rubner and K. Schulten. A regularized approach to color constancy. *Biological Cybernetics*, 61:29–36, 1989.

[SB90] M.J. Swain and D.H. Ballard. Indexing via color histograms. In *Proceedings: International Conference on Computer Vision, Osaka , Dec.4-7/90*, pages 390–393. IEEE, 1990.

[SFB92] M. Stokes, M.D. Fairchild, and R.S. Berns. Precision requirements for digital color reproduction. *ACM Transactions on Graphics*, 11(4):406–422, october 1992.

[SH71] H.G. Sperling and R.S. Harwerth. Red-green cone interactions in the increment-threshold spectral sensitivity of primates. *Science*, 172:180–184, 1971.

[Sha85] S.A. Shafer. Using color to separate reflection components. *Color Res. Appl.*, 10:210–218, 1985.

[Sne83] D.H. Foster R.S. Snelgar. Test and field spectral sensitivities of colour mechanisms obtained on small white backgrounds: action of unitary opponent-colour processes? *Vision Research*, 23:787–797, 1983.

[SSS92] B. Smith, C. Spiekermann, and R. Sember. Numerical methods for colorimetric calculations: Sampling density requirements. *COLOR research and application*, 17(6):394–401, 1992.

[ST93] G. Sharma and H.J. Trussel. Characterization of scanner sensitivity. In *IS&T and SID's Color Imaging Conference: Transforms & Transportability of Color*, pages 103–107. 1993.

[TO90] M. Tsukada and Y. Ohta. An approach to color constancy using multiple images. In *Proc. Third International Conference on Computer Vision*. IEEE Computer Society, December 1990.

[Tom94] S. Tominaga. Realization of color constancy using the dichromatic reflection model. In *The second IS&T and SID's Color Imaging Conference*, pages 37–40. 1994.

[TV91] H.J. Trussel and M.J. Vrhel. Estimation of illumination for color correction. In *Proceedings of the International Conference on Acoustics Speech and Signal Processing*, pages 2513–2516. IEEE, 1991.

[TW89] S. Tominaga and B.A. Wandell. Standard surface-reflectance model and illuminanat estimation. *J. Opt. Soc. Am. A*, 6:576–584, 1989.

[VGI94] M.J. Vrhel, R. Gershon, and L.S. Iwan. Measurement and analysis of object reflectance spectra. *COLOR research and application*, 19(1):4–9, 1994.

[Vrh93] M.J. Vrhel. *Mathematical methods of color correction*. PhD thesis, North Carolina State University, Department of Electrical and Computer Engineering, 1993.

[VT93a] P.L. Vora and H.J. Trussel. Measure of goodness of a set of color-scanning filters. *J. Opt. Soc. Amer. A*, 10:1499–1508, 1993.

[VT93b] M.J. Vrhel and H.J. Trussel. Physical device illumination correction. In *Device-Independent Color Imaging and Imaging Systems Integration*, volume 1909, pages 84–91. SPIE, 1–3 Feb. 1993.

[VW71] J.J Vos and P.L. Walraven. On the derivation of the foveal receptor primaries. *Vision Research*, 11:799–818, 1971.

[WB82] G. West and M.H. Brill. Necessary and sufficient conditions for von Kries chromatic adaption to give colour constancy. *J. Math. Biol.*, 15:249–258, 1982.

[WB86] J.A. Worthey and M.H. Brill. Heuristic analysis of von Kries color constancy. *Journal of The Optical Society of America A*, 3(10):1708–1712, 1986.

[Wil65] J.H. Wilkinson. *Algebraic Eigenvalue Problem*. Monographs on numerical analysis. Oxford University Press, 1965.

[WS82] G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulas*. Wiley, New York, 2nd edition, 1982.