# NOVEL COLOUR CONSTANCY ALGORITHMS FOR DIGITAL COLOUR IMAGERY

by

Lilong Shi
Bachelor of Applied Science, Simon Fraser University, 2003
Master of Computing Science, Simon Fraser University, 2005

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the
School of Computing Science


SIMON FRASER UNIVERSITY
Fall 2009

# APPROVAL

**Name:** **Lilong Shi**

**Degree:** **PhD of Computing Science**

**Title of Thesis:** **NOVEL COLOUR CONSTANCY ALGORITHMS FOR DIGITAL COLOUR IMAGARY**

**Examining Committee:**

**Chair:** **Greg Mori**
Assistant Professor

_____

**Brian Funt**
Senior Supervisor
Professor

_____

**Tim Lee**
Supervisor
Adjunct Professor

_____

**Mark S. Drew**
Internal Examiner
Professor

_____

**Alexander Logvinenko**
External Examiner
Professor
Glasgow Caledonian University

**Date Defended/Approved:** _____December 15, 2009_____

# Declaration of
# Partial Copyright Licence

# ABSTRACT

Colour constancy algorithms differ in their derivation, implementation, performance and assumptions. The focus of the research presented in this thesis is to discover colour constancy solutions to recover surface colours, or equivalently, to estimate the illumination, of single light source in a given scene.

Several colour constancy models will be proposed. These methods have different methodologies and constraints. For example, a method can be constrained on a particular model surface material, on blackbody radiation light source, on dichromatic model, and on spatial variation of the illumination and the reflectance. The methods to be discussed include, for instance, a method of identifying achromatic surfaces, which can then be used as known references for estimating the scene illumination. A second method examines the colour of human skin and its dependence on its hemoglobin content, melanin content, and the illuminating light. The corresponding basis of these three factors can be represented linearly in logarithm space, where the colour of the light can then be estimated. A third method, uses the fact that the colours reflected by an inhomogeneous dielectric material lie on a plane spanned by the colour of the specular component reflected from the air-surface interface and the colour reflected from the body of the material. Once these planes are detected by a Hough transform, their intersection line represents the scene illumination. A fourth method is based on the independence and difference in the rate of spatial

variation of the luminance and the surface reflectance in a given scene, from which image features can be separated via non-negative matrix factorization to reveal the true surface reflectance. A fifth method is based on learning the correspondence between an image's colour content and its illumination via thin-plate-spline interpolation so that the chromaticity of the light can be calculated. Finally, a quaternion-based curvature measure approach is developed that can be used as a complement to colour constancy methods that use information from spatial edges. In this thesis, these various methods are proposed to overcome drawbacks in existing approaches for better performance and improved robustness and efficiency.

**Keywords:** Colour Constancy, Illumination Estimation, Dichromatic Model, Curvature

# ACKNOWLEDGEMENTS

I offer my enduring gratitude to the faculty, staff and my fellow students at the SFU School of Engineering Science, who have combined to create a stimulating synergy for research in our field. I owe particular thanks to Dr. Brian Funt, for his supervision and support throughout my studies.

I thank Samsung Inc. for their support in the year 2006.

Special thanks are owed to my parents, whose support throughout my education years has been unbroken, and multi-faceted.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# 1: INTRODUCTION

Colour is an important, efficient and useful feature of images that has been widely utilized in the computer vision community for machine vision and image processing applications, such as segmentation, object recognition and object tracking. Although colour is often used together with other image features in these tasks, accurate reproduction of the surface colour is critical.

The response of a colour imaging acquisition device to lights depends on three factors: the underlying physical properties of the objects, the nature of the illumination incident on the objects, and the characteristics of the sensors of the imaging system itself. Problems arise when capturing the same scene under changing illuminant conditions. For example, images appear to be reddish if a scene is captured under tungsten illumination or a bluish if captured under fluorescent lighting. As a consequence, one must recover surface colour and reduce colour variation that appears in different views of the same scene in order to extract accurate information from images. Such a process is identified with the classical term, colour constancy.

The focus of the research presented in this thesis is to discover colour constancy solutions for recovering surface colours, or equivalently, to estimate the colour of the light. The colour constancy algorithms to be studied differ in derivation, implementation, performance and assumptions. In Chapter 1, background material is given on the basics of colour vision, colour formation and

characteristics of common light sources. In the first part of Chapter 2, some common existing colour constancy methods will be reviewed. These approaches involve separating the illumination from the reflectance and extracting the accurate colour of the objects. In the second half of Chapter 2, a list of the topics of my research will be briefly introduced, including five novel colour constancy models for illumination estimation under a single light source. Detailed discussion of these models, their implementation and performance will be covered in Chapters 3 to Chapter 6. In Chapter 7, a quaternion-based vessel detection method for colour images is provided as a potential aid for colour constancy models that use information from spatial edges.

## 1.1 Vision Basics

Colour perception is described as a sensation created in response to excitation of our visual system by the visible region of the electromagnetic spectrum. James Clerk Maxwell showed that light is essentially a form of electromagnetic radiation that contains radio waves, visible light, and X-rays. All of these types of radiation can be represented as a spectrum of radiation; the electromagnetic radiation that includes radio waves at one end and gamma rays at the other. The visible radiation wavelength range differs across species. For humans, the visible spectrum wavelength occupies the portion of the electromagnetic spectrum, ranging from approximately 400 nm to 700 nm.

As the important components of human visual system, eyes detect light by photoreceptors in the retina that convert it into electrical signals for the brain to process. The retina is an important component of the visual system, where two

kinds of photoreceptor cells--- cones and rods--- can be found. Having three types of cones, namely, L, M and S, each mostly sensitive to a certain portion of the spectrum, we are able to distinguish a great variety of colours. These cones are the most sensitive at wavelengths of 580nm (red), 540nm (green), and 480nm (blue) respectively, but are not activate in a dim environment. The rods are, on the other hand, responsible for dark-adapted vision. As mentioned before, a colour signal received by the eye depends on three factors: the light source that initiates the sensory process of vision; the object that absorbs and reflects the light source at the selected wavelengths determined by physical and chemical properties; and the human photoreceptors. This implies colour from the same objects or scenes may differ due only to illuminant changes. Since the intensity and spectral characteristics of natural or artificial lights can make significant changes, the colour from a surface with a fixed reflectance may vary substantially from one lighting condition to another. Interestingly, this is not true for our colour perception. For example, a red apple under different lighting conditions (sunlight, the light of a fire, or a harsh electric light) still appears red to us, in spite of the fact that the physical spectra are quite different.

The definition by Foster et al. [28] states that “Colour constancy is the constancy of the perceived colours of surfaces under changes in the intensity and spectral composition of the illumination.” Also, according to Wandell [58], the colour signal is the light arriving at the eye, which is the product of the spectral power distribution of the ambient light and the surface reflectance function. Theories have been proposed to understand how reflectance information is

extracted from colour signals. These theories are sometimes referred as “discounting the illuminant” because it separates the contribution of the reflectance from that of the illuminant in colour signals. A perfect colour constancy algorithm should accurately recover object reflectance under arbitrary lighting conditions by completely de-correlating illumination from surface reflection characteristics, and therefore be invariant to changes in the surface’s location, composition and lighting. However, no colour constancy algorithm is likely to be able to perfectly recover object reflectance due to the fact that the problem to solve is ill-posed. Nevertheless, efforts have been made in solving colour constancy problems through developing colour constancy models and corresponding algorithms that will be discussed in the following chapters.

## 1.2 Colour Signals

The colour signal defines all information received by the eye to understand reflectance at different surface locations[58]. The ultimate stage of the imaging system is to build a mathematical model embodying the predominant phenomena occurring in the formation of colour images. Therefore, all of the light source, the object, and the optical system should be quantified. In an imaging system, the light source is represented by its spectral power distribution $E(\lambda)$; a surface by its percent surface spectral reflectance function $S(\lambda)$; and the optical system by the spectral sensitivity function $\Gamma(\lambda)$ of each sensor. Usually, an optical system contains more than one sensor class. The human retina, for example, has three types of sensors, known as cones, which are sensitive to light of different wavelengths across the visual spectrum represented as different spectral

response curves. For this reason, a common imagining system contains three sensors, and thereby a colour is specified by a 3-component signal.

A colour signal, *c*, is generated as a consequence of the interaction between the light and a surface. More specifically, the colour signal is the product of spectral power distribution of the light source and the reflectance of the surface,

$$c(\lambda) = E(\lambda)^* S(\lambda). \tag{1.1}$$

In the second process, the colour signal reflected from object surface is captured by the sensors as an integration over all wavelengths across the visible spectrum. In a *k*-sensor imaging system, the colour signal is then represented as *k*-dimensional sensor response, as a result of the summation of the product of colour signal and sensors' response curves at each wavelength, as shown in Equation **1.2**.

$$p_k = \int \Gamma_k(\lambda) E(\lambda) S(\lambda) d\lambda \qquad k = R, G, B \tag{1.2}$$

Here $p_k$ stands for the camera sensor response. The vector **p** = <$p_G$, $p_G$, $p_B$> is called the *image colour*. When E(λ) = 1, **p** represents the *surface colour*. When S(λ) = 1, **p** represents the *illumination colour*. These terms will be widely used throughout this thesis. When narrowband sensors are used, (i.e., sensor sensitivity function, $\Gamma_k(\lambda)$, is approximately a Dirac delta function centred at

wavelength $\lambda_k$), Equation **1.2** can be simplified by eliminating the sensor sensitivity function and the integral, such that

$$P_k \approx S(\lambda_k)E(\lambda_k) \qquad k = \{R, G, B\} \tag{1.3}$$

In this model, the geometry of the object surface creates merely a scale factor in the function *E* since the surface is assumed to be Lambertian. Such a factor can be removed by normalization of this function. This model is correct unless severe specular reflection occurs. This model will be widely used throughout this thesis.

Previously, Finlayson and Schaefer [24] measured spectra of 172 light sources, including daylights and fluorescents. They reported that the illuminant chromaticities approximately fall on a long thin band that is clustered around the locus of chromaticities formed by in the chromaticity plane by a Planckian Blackbody radiator of varying temperature. The chromaticities of general purpose light sources (e.g., daylight and fluorescent light) have a small deviation from the Blackbody radiator of corresponding *correlated colour temperature*(CCT). A light source with a high CCT gives the material a bluish appearance (e.g., skylight) whereas a low CCT gives a reddish appearance (e.g., during sunset). The Planckian *spectrum power distribution* (SPD) is a smooth function that provides a good approximation for tungsten/halogen lamps and sunrise/sunset lighting (Hunt 1987). This fact suggests that commonly used light sources can be approximately modelled as blackbody radiators. This assumption is particularly

useful as illumination then becomes a function of a single scalar, i.e., temperature *T*. The Wien's approximation to the Planckian formula can be expressed as

$$E(\lambda,T) = Ic_1\lambda^{-5}e^{-\frac{c_2}{T\lambda}}, \qquad 1.4$$

where *I* is the power of radiation of the illumination, *T* is the blackbody radiator temperature, and $c_1$ and $c_2$ are two constant of value $3.74183*10^{-16}Wm^2$ and $1.4388*10^{-2}mK$, respectively. Based on Equation **1.2** and Equation 1.4, pixel intensity *P* from a sensor *i* of diffuse reflectance *S* as imaged by a linear digital colour camera (excluding any subsequent gamma correction) can be described [23] by,

$$P_k = \int S(\lambda)Ic_1\lambda^{-5}e^{-\frac{c_2}{T\lambda}}\Gamma_k(\lambda)\partial\lambda, \qquad k=\{R,G,B\} \qquad 1.5$$

For the case of narrowband sensors, we have

$$P_k \approx S(\lambda_k)Ic_1\lambda_k^{-5}e^{-\frac{c_2}{T\lambda_k}}, \qquad k=\{R,G,B\} \qquad 1.6$$

In the rest of this thesis, we will see that Equation 1.3 and Equation 1.6 will be extensively used since they serve as the foundation to many colour constancy models in very simple forms, as they were used by Finlayson [23][25].

It should be noted that the assumption that the sensors' spectral sensitivity functions are very narrowband, being represented by delta functions, means that if the assumption were to hold then there would be significant observer metamerism between such cameras and human observers. In other words, the "colour vision" of such cameras would be very special, deviating enormously from human colour vision. However, in Finlayson's methods [23][25] and in those introduced here, the narrowbandedness assumption is used only for the derivation of the methods--- the methods are then tested using standard, commercially available cameras whose spectral sensitivities deviate substantially from narrowbandedness.

The rest of this thesis is organized as follows. First of all, the basics of colour vision, colour formation and characteristics of common light sources are reviewed. Then, the most common existing colour constancy methods will be revisited. A list of the research topics that will be studied in the rest of thesis is made, including five novel colour constancy models for illumination estimation. These models, their implementation and performance will be covered in detail. Finally, a quaternion-based vessel detection method in colour images is provided as a potential aid for colour constancy models that use information from spatial edges.

# 2: COLOUR CONSTANCY MODELS

The process of solving colour constancy can be defined as the transformation from a source image taken under an unknown illuminant to a target image of the same scene as if it would have been obtained by the same camera under a known illuminant. The complete procedure involves two steps. The first step estimates the scene illumination of the source image. Based on this estimate the second step corrects the image colours pixel-by-pixel to the colours under the known illuminant. The focus of this thesis is on the first step of such colour constancy process.

Many assumptions have been incorporated into the various colour constancy algorithms because colour constancy is generally known to be an ill-posed problem due to the fact that the surface colour and illumination colour are not uniquely separable. According to their assumptions and techniques, colour constancy algorithms are mainly divided into two categories. The first is to recover object surfaces in canonical condition. In this category, the object image can be viewed as a result of objects under a certain canonical condition, normally the “white” illumination having equal energy at all wavelengths. Colour constancy solutions in this category recover the original object image from a given colour image under unknown illumination, removing illumination effects. The second category is to estimate the illumination colour based on the colour distribution in

the scene. Here illumination values are estimated, either as 2 chromaticity (e.g., *r*,*g*) or 3 colour values (e.g., R,G,B).

On the other hand, these algorithms may be divided into two groups based on whether prior statistical information is important: unsupervised and supervised. Unsupervised algorithms predict the illumination information solely based on the content in a single image with certain assumptions about the general nature of colour images; while supervised ones always require two stages: (1) learn training data by building a statistical model associating the input images to the corresponding illuminations, (2) predict the unknown illumination of any given image based on this statistical model.

In the first half of this chapter, a survey of a number of the most common colour constancy algorithms will be provided. The second half of this chapter briefly introduces the proposed algorithms in this thesis.

## 2.1 Unsupervised Illumination Estimation

In this section, five illumination estimation algorithms requiring no training will be reviewed, namely, Retinex, Max-RGB, Gray-World, Shades-of-Gray, and the Gray-Edge-Hypothesis. The last four approaches are similar in a way by making the assumption that the content of the image has a certain colour. Max-RGB and Gray-World are the simplest, yet perhaps the most widely used. Shades-of-Gray unifies and generalizes the former two, to improve the result, while the Gray-Edge-Hypothesis is considered as a more advanced approach by emphasizing pixels close to spatial edges.

### 2.1.1 Retinex

*Retinex*, as one of the best known colour constancy algorithms, originated from Land's research work on human vision [51]. This method received its name "Retinex" from "retinal" and "cortex", as Land believed that such mechanism is due to interaction between both components. Land discussed the fact that colour appearance depends on the relative absorption of light by the visual cones and its spatial pattern on the eye, rather than the absolute values of photo-pigment absorption. As a consequence, colour vision is illumination-independent at various locations, but dependent on the path followed by which the light reaches the eye.

The basic idea of Retinex is to separate the illumination from the reflectance in colour channel, $I_k$ ($k$ = R,G,B), independently. Using sharp sensors, according to Equation 1.3, the intensity value at location ($x,y$) can be represented as a product of illumination and reflectance:

$$I_k = E_k \cdot S_k \qquad k = R, G, B \tag{2.1}$$

Here, $E_k$ and $S_k$ are illumination and reflectance values in the $k$-th colour channel at location ($x,y$). For simplicity, the Retinex computation is implemented in the logarithm domain where multiplicative operations become additive. Retinex also assumes spatially smooth in the illumination field (i.e., the illumination changes smoothly across the scene), while the reflectance image corresponds to abrupt intensity changes in the image.

Many variants of Retinex have been proposed. Stockham [81] and Faugeeras [21] separated illumination from surface reflectance by applying a homomorphic filter on the input image for low- and high-pass results in the logarithmic domain. Horn [43] formalized Retinex in terms of differentiation, thresholding, and re-integration in the logarithm domain. Multi-resolution versions of Retinex were introduced for efficiency [30]. Kimmel [47] proposed a variational model for the Retinex problem. This model unifies previous Retinex solutions by formulating the illumination problem as a Quadratic Programming problem. Two versions of Retinex have been given standardized definitions in terms of Matlab code [33]. Parameters, such as path length, number of paths, and how a path is calculated, are very important in 'Retinex.' A discussion about their tuning can be found in [32][33][51].

### 2.1.2 Max-RGB

The Max-RGB algorithm, also referred as the *White-Patch* hypothesis, assumes that the maximal signal values in each colour channel, assumed to be the responses from a white surface, represent the scene illumination [22]:

$$E_k = \max_{x,y \in \Omega}(I_k(x,y)) \qquad 2.2$$

The Max-RGB solution can be viewed as a special limiting case of Retinex. Obviously, this method requires a scene that contains either a single surface that is maximally reflective throughout the range of the sensitivity of the imaging device (i.e., a white surface) or a number of surfaces that are maximally

reflective throughout the range of each of the three imaging sensors individually [40].

In spite of its simplicity, Max-RGB usually does not provide a satisfactory estimate for a real world scene due to violation of the assumptions and the limitation of the devices such as overflow duration signal acquisition [26].

### 2.1.3 Gray-World

The Gray-World algorithm assumes that the average colour of surfaces in the world is gray. Given an image with sufficient surface colour variance, or with a uniformly gray surface, the average colour of all surfaces tends to be gray. The offset from the gray is then due to the effect of lighting. Therefore, the average values in each band of the image are taken as the estimate of the illumination colour [9]. If we average Equation 1.3 at every pixel location, then

$$\bar{c}_k(\lambda) = E_k(\lambda)\bar{S}_k(\lambda) \qquad 2.3$$

Since the average reflectance is gray, the $\bar{S}_k(\lambda) = 1/\zeta$ , which is a constant number at all wavelengths λ. Hence, $\zeta \cdot \bar{c}_k(\lambda) = E_k(\lambda)$ , or

$$E_k = \sum\sum_{x,y} I_k(x,y)/N \qquad 2.4$$

As the surface colour is random and independent, it is reasonable to say that given a sufficiently large number of samples, the average surface colour should converge to gray. Therefore, if an image is taken with a camera under

yellow lighting, the output image can be expected to have a yellow cast over the entire scene. The effect of this yellow cast disturbs the Gray-World average on the original image and produces a shift from gray. By enforcing the assumption on the camera output image, the yellow cast may be removed to reveal better the actual colours of the surfaces.

### 2.1.4 Shades-of-Gray

The Max-RGB and Gray-World algorithms work well only if their assumptions are satisfied, that is, the existence of a white patch in the scene and the average surface colour in the scene is gray, respectively. Another approach named *Shades-of Gray* has been proposed by Finlayson et al. [26] as a more generalized approach for illumination estimation using the Minkowski norm. His assumption is that average surface colour in a scene tends to be gray after a nonlinear invertible transformation (*p*-norm function is selected here) of the pixels in each channel.

Without loss of generality, consider a single channel $I_k$ of an $N$-pixel multi-channel image. According to Finlayson's approach, the average colour of the scene raised to a power of $p$ tends to be gray so the illumination intensity of the $k^{th}$-channel can be estimated as the following:

$$E_k = \sqrt[1/p]{\frac{\sum\sum_{x,y}[I_k(x,y)]^p}{N}} \qquad 2.5$$

Here $E_k$ is the estimated illumination of the $k^{th}$ colour channel. This equation is in the form of Minkowski *p*-norm definition. We can use a similar method to find the illumination estimation for other colour channels as well.

Max-RGB and Gray-World algorithms are two special cases of Shades-of-Gray. By setting $p = \infty$ in the Minkowski norm, Equation **2.5** turns into Equation **2.2** of the Max-RGB approach. On the other hand, by setting $p = 1$ in the Minkowski norm, Equation **2.5** turns into Equation **2.4** of the Gray-World approach.

### 2.1.5 Gray-Edge-Hypothesis

The Gray-Edge Hypothesis proposed by J. Weijer and Th. Gevers [89] assumes that "the average reflectance difference in a scene is achromatic". Therefore, given that pixel value at location (*x*,*y*) of the *k*-th colour channel is $I_k(x,y)$, the illumination intensity at channel $I_k$ can be estimated by

$$E_k = \sum \sum_{x,y} \left\| \nabla I_k(x,y) \right\| / N \,, \qquad 2.6$$

with the magnitude of derivatives defined as

$$\left\| \nabla I_k(x,y) \right\| = \left\| \left( \tfrac{\partial}{\partial x} I_k(x,y), \tfrac{\partial}{\partial y} I_k(x,y) \right) \right\| = \sqrt{ \left( \tfrac{\partial}{\partial x} I_k \right)^2 + \left( \tfrac{\partial}{\partial y} I_k \right)^2 } \,. \qquad 2.7$$

Weijer justified his hypothesis in *Opponent Colour Space* (OCS), where the three axes correspond to the three principle axes extracted from all possible

RGB colours in the world via PCA. For a tri-band RGB colour image, the colour derivatives can be transformed into *Opponent Colour Space* by

$$\begin{cases} OC_1 = (\|\nabla R\| - \|\nabla G\|)/\sqrt{2} \\ OC_2 = (\|\nabla R\| + \|\nabla G\| - 2\|\nabla B\|)/\sqrt{6} \\ OC_3 = (\|\nabla R\| + \|\nabla G\| + \|\nabla B\|)/\sqrt{3} \end{cases} \quad 2.8$$

Under canonical illumination, the longest axis from the colour derivative value distribution should coincide with $OC_3$ . The solution of Gray-Edge hypothesis can be interpreted as skewing the colour derivatives distribution such that the average output corresponds to the white light direction in the *Opponent Colour Space*. So the average value of all the colour derivative values from the whole image gives the illumination colour. Additionally, this method can also be extended by incorporating it into the *p*-th Minkowski norm, just as Shades-of-Gray colour constancy does for the Gray-World algorithm, with improved results [88][89].

## 2.2 Supervised Illumination Estimation

In this category of colour constancy methods, the scene illumination is estimated by learning training data sets according to statistical models built on the input images paired with known illuminations. Three supervised illumination estimation methods will be reviewed in this section: Neural Network, Colour-by-Correlation, and Support Vector Regression.

### 2.2.1 Neural Network (NN)

In this method, a multi-layer neural network was established to learn the relationship between illumination chromaticity and colour distribution in the image, and then to predict the unknown illumination from an image [31][10]. The training input is a binary histogram of image chromaticities. The (*r*,*b*) space is divided into cells 0.02 units wide so that it includes 2500 bins as input layer nodes. A '1' or '0' in each bin represents the 'presence' or 'absence' of a given chromaticity. The neural net has two hidden layers: one has 400 nodes and the other has 30 nodes. Two output nodes with real value are the corresponding illumination chromaticities. It is trained with the back-propagation algorithm with a sigmoid activation function.

### 2.2.2 Colour-by-Correlation (C-by-C)

*Colour-by-Correlation*, proposed by Finlayson et al. [24], builds the correlation matrix to correlate the probability of image colours with each possible illuminant. This matrix is built on a large set of colour images and corresponding known illuminations. To cancel out the effects of intensity, geometry, and shading, these images' colours are converted into chromaticities and then mapped to histograms bins. The rows in the matrix are predefined chromaticities; the columns are known illuminants from the training data set. Each entry of the matrix indicates the frequency of occurrence of a given chromaticity under a given light. Assuming equal likelihood of occurrence of the lights, given the image data $C_{im}$, the probability that $E$ was the scene illuminant, $Pr(E|C_{im})$ is estimated as follows:

$$\Pr(E \mid C_{im}) = k \prod_{c \in C_{im}} \Pr(c' \mid E)\,, \qquad 2.9$$

where *k* is some constant and *c'* is a chromaticity. In the logarithm domain, the multiplication becomes addition.

In the training stage, three steps are needed in building a correlation matrix M: (a) Characterize which image colours (chromaticities) are possible under each reference illuminant. (b) Build a probability distribution for each light. (c) Encode these distributions in the columns of the matrix. The result of training is an $n^h$ x $n^E$ correlation matrix M that contains the frequency of each observed colour under each different reference illuminant, where $n^h$ is the size of the histogram, and $n^E$ is the number of reference illuminants. Solving for colour constancy also takes three steps: (a) Transfer the image colours into a binary vector in which '1' or '0' indicates the presence or absence of the corresponding chromaticity in the image. (b) Correlate this histogram *h* with each column of the correlation matrix, by simply applying a dot-product in log space. (c) This information is used to find an estimate of the unknown illuminant, for example, the illuminant which is most correlated with the image data.

The other contribution of Finlayson's work [24] is it proves that this framework is general and can be used to describe many other existing algorithms. Barnard et al. [4] improved the 'Colour by Correlation' method by extending it into the 3D colour space, where, in addition to chromaticity, the pixel brightness is used as extra information.

### 2.2.3 Support Vector Regression (SVR)

In this method, Xiong et al. [92] suggested using Support Vector Regression(SVR) as a tool for learning the relationship between illumination chromaticity and the colour distribution in the image. SVR estimates a continuous-valued function that encodes the fundamental interrelation between a given input and its corresponding output in the training data. This function then can be used to predict outputs for given inputs that were not included in the training set. This is similar to a neural network. However, a neural network's solution is based on empirical risk minimization. In contrast, SVR introduces structural risk minimization into the regression and thereby achieves a global optimization while a neural network achieves only a local minimum.

In the training stage, a given colour image is converted into a binary histogram *h* that uniformly covers the chromaticity space. Then the standard Support Vector Regression technique is applied to find a mapping function from image histograms to illuminant chromaticity. Here is the general form of the regression function *f*(*h*):

$$f_k(h) = \sum_{i=1}^{l} w_{k,i} K(h_i, h) \qquad 2.10$$

where *w*'s are the Langrange multipliers that need to be determined in the training stage, and $h_i$'s are the binary histograms in the training set containing *l* data points. The function $f_k(h)$ is the illuminant chromaticity estimation function for $k = r$ or $g$ (two chromaticity values), each associated with its own set of

parameters. The kernel function $K(h_i,h_j)$ measures the non-linear distance between two data points $h_i$ and $h_j$. The performance of SVR is dependent on the choice of kernel functions and the corresponding parameters. Rows 1-4 in Table 2.1 list a set of common choices of kernel functions for SVR:

| Name | Definition | Parameters |
|---|---|---|
| Linear | $K(h_i, h_j) = h_i^T \cdot h_j$ | None |
| Polynomial | $K(h_i, h_j) = (h_i^T h_j + 1)^\alpha$ | α |
| Radial Basis Function(RBF) | $K(h_i, h_j) = e^{-\alpha \|h_i - h_j\|}$ | α |
| Sigmoid | $K(h_i, h_j) = \tanh(h_i^T h_j + \alpha)$ | α |
| Thin-Plate-Spline (TPS) | $K(h_i, h_j) = \|h_i - h_j\| \cdot \log(\|h_i - h_j\|)$ | None |

**Table 2.1 Some common kernel functions for Support Vector Regression.**

## 2.3 Models Proposed in this Thesis

Depending on different methodologies, separation of the illumination from the reflectance of the objects can be achieved by supervised and non-supervised approaches. In this thesis, the main focus is on the category of non-supervised methods that require no training. I will explore non-statistical estimation solutions constrained by the characteristics of surface material as well as the illumination. The following characteristics will be studied: surface colour of certain materials, such as human skin and achromatic surfaces; illumination spectra due to blackbody radiation; the formation of colour imaginary according to the dichromatic model of reflection; and the independence of spatial variance of illumination and surfaces in a given scene. Additionally, I will discuss one supervised approach that uses prior information from a training set. Here is the list of topics to be represented:

### Chapter 3. Gray Surface Identification

In this chapter, the proposed extension first identifies colours that are likely to be from truly gray surfaces, and then averages only those colours. The trick is in the identification of gray surfaces in a colour coordinate system that encodes illumination and surface reflectance along different axes is used, namely, LIS colour coordinates.

### Chapter 4. Use of Skin Colour as a Constraint

The colour of skin in an image depends strongly on the colour of the incident light. To account for the induced shifts in skin colour, I propose a skin-illumination model in a linear space constrained by both the blackbody radiator locus and the skin locus. Based on this model, the range of chromaticities of arbitrary skin under all illuminant colour temperatures can be transformed into a new coordinate system defined with two independent axes: an illumination axis as a function of colour temperature, and a skin axis as a function of melanin content.

### Chapter 5. Solving Dichromatic Model

In this Chapter I propose a robust method for determining the illumination axis. The method detects dichromatic planes while placing few restrictions on the image content, such as the number of surfaces, the surface colours, or the identification of specular regions. The approach involves two Hough Transforms in sequence that result in a histogram representing the likelihood that a candidate

intersection line is the image illumination axis. The final illumination estimate is determined by intelligently choosing from amongst the most likely candidates.

**Chapter 6. Reflectance Recovery via Matrix Factorization**

In this chapter, I propose a new approach to illumination estimation for colour constancy and automatic white balancing, by separating the image into illumination and reflectance components based on the technique of nonnegative matrix factorization (NMF) with sparseness constraints (NMFsc). The image data is then organized as a matrix to be factored into two nonnegative components---illumination and reflectance. The approach provides a pixel-wise estimate of the illumination chromaticity throughout the entire image.

**Chapter 7. Illumination Estimation via Thin-Plate-Spline Interpolation**

In this Chapter, thin-plate spline interpolation is used to interpolate the chromaticity of the incident scene illumination from an image of the scene. Thin-plate splines interpolate over a non-uniformly sampled input space, which in this case is a set of training images and associated illumination chromaticities. Tests of the thin-plate spline method on a large set of real images demonstrate that the method estimates the colour of the incident illumination quite accurately.

**Chapter 8. Quaternion Colour Curvature**

Edges have been proven useful in colour constancy algorithms. For better detection of edge and curvature structures in colour images, I propose a novel approach to measuring curvature in colour or vector-valued images based on

quaternion singular value decomposition of a Hessian matrix. Test results show the effectiveness of quaternion colour curvature in generating a vesselness map.

These algorithms are applied on several datasets (described in *Appendix B*) to compare the performance with other existing methods in terms of estimation errors (details in *Appendix A*).

# 3: GRAY SURFACE IDENTIFICATION

The key to automatic white balancing of digital imagery is to estimate accurately the colour of the overall scene illumination. Many methods for estimating the illumination's colour have been proposed [9][10][22][24][26][34]. Although not the most accurate, one of the simplest and quite widely used methods is the gray world algorithm [9]. Borrowing on some of the strengths and simplicity of the gray-world algorithm, this chapter introduces a modification of it that significantly improves on its performance while adding little to its complexity.

## 3.1 Introduction

As reviewed in Section 2.1.3, the standard gray world algorithm is based on the assumption that the average surface colour in a scene is gray so that when an image's colours are averaged, any departure from gray reflects the colour of the scene illumination. In this chapter, the proposed extension first identifies colours that are likely to be from truly gray surfaces, and then averages only those colours. The trick is in the identification of gray surfaces. Note that one must make a distinction between the colour of the surface as it would appear under white light and the image colour of that same surface under the unknown scene illumination. Simply averaging of the image colours that are gray would not tell us anything other than that gray colours are gray. To find the surfaces that are gray, but do not necessarily appear gray in the image because of the

effect of the illumination, a colour coordinate system that encodes illumination and surface reflectance along different axes is used.

## 3.2 LIS Colour Coordinates

LIS coordinates, used by Finlayson et al [23][91], represent luminance, illumination colour and surface reflectance as separate dimensions, hence the designation “LIS” coordinates [23][91]. The goal of the coordinate system is to represent the three components of a colour in terms of the underlying physical components that generated the colour, in particular, luminance/intensity, incident illumination colour, and the surface reflectance colour. Of course, this goal cannot actually be met without additional information, but it can be approximated to a useful extent. Tests with the LIS representation of colours from images showed that LIS points having an S coordinate of zero were generally gray. They are not just gray in RGB image space, but they represent gray surface colours because they are in the reflectance space. To the extent that the S coordinate actually does represent reflectance and truly is independent of the illumination, this means that one can identify gray surfaces in an image independent of whether or not they have R=G=B.

The strategy for the proposed new gray-surface-identification (GSI) method of automatic white balance, therefore, is to use the LIS coordinates to identify gray surfaces in the image, and then use these grays to estimate the illuminant colour. This final step is done in the RGB colour space, converted back from the LIS space, and involves averaging the chromaticities of the surfaces that have been identified as gray. The details are described below.

LIS coordinates [23][91] are based on three assumptions (as discussed in Section 1.2): (1) that the illuminants are blackbody radiators; (2) that the camera's response functions are narrowband and can be modelled as Dirac delta functions; and (3) that no specular reflection occurs. The implication of the first assumption is that the illuminants can be modelled as a function of a single parameter, namely, the blackbody temperature. The implication of the second assumption is that each of the RGB channels is affected by only a single distinct wavelength of the incoming spectrum. Under these assumptions, Finlayson and Hordley [23] show that for a given camera, [*log*(R/G), *log*(B/G)] is an illumination-invariant colour chromaticity space in which the values of same surface under various illuminations tend to fall on a straight line and lines from different surfaces are parallel. Similarly, for a fixed surface reflectance, varying the intensity and colour temperature of the illumination incident on it causes the logarithm of the camera response [*log*(R), *log*(G), *log*(B)] to move within a plane. Different surface reflectances yield parallel planes. The S axis, which correlates with surface reflectance, of the LIS system is defined perpendicular to these planes. The L (luminance) axis and I (illumination 'colour') axis are then orthogonal to the S axis. An example of the LIS coordinate is shown in Figure 3.1, where the synthesized colours from three surfaces span three parallel planes: cyan, yellow and magenta planes in the 3D log space.

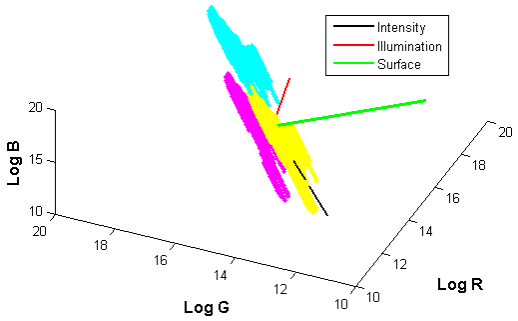



**Figure 3.1. An example of the LIS coordinate. Synthesized colours from three distinct surfaces span three parallel planes: cyan, yellow and magenta planes in the 3D log space.**

Although in theory the logarithm of the camera responses [*log*(R), *log*(G), *log*(B)] obtained from a given surface under all possible colours and intensities of illumination are predicted to lie in a plane, do they in practice? Clearly, the blackbody-radiator and Dirac-delta assumptions are strong ones, and are likely to be violated. However, for [*log*(R), *log*(G), *log*(B)] data synthesized based on the SONY DXC-930 sensitivity functions, the 102 illuminant spectra from the Simon Fraser University database [5] and the surface reflectance of the 24 Macbeth colour checker surface patches, PCA (principal component analysis) determines the plane and establishes that the first 2 dimensions explain 99.1% percent of the variance. These 102 illuminants are not specifically blackbody radiators, but common light sources found around a university campus. Similarly, although the camera sensitivity functions [5] are relatively sharp with little overlap between them, they are taken from a real camera, and certainly violate the Dirac-delta assumption. Despite violating the assumptions, the fit of a plane to the data is surprisingly good.

## 3.3 GSI Implementation

The first issue in terms of implementing the GSI colour constancy algorithm is that the LIS system is camera dependant and must be determined for the camera being used. There are two methods to do this depending on whether or not the camera's spectral sensitivity response functions are known. If they are known, then they can be used to calculate camera responses for spectra synthesized as the product of illuminant and reflectance spectra chosen from a database of spectra. If the camera's spectral sensitivity curves are not known, then real values can be obtained by using the camera to take images of a gray card under several different illuminants. PCA is then applied to the logarithm of RGBs from the gray card. The vector corresponding to the maximal eigenvalue forms the intensity axis, the next vector forms the illumination axes, and the vector corresponding to the least eigenvalue is the surface reflectance axis because there is no variation in surface reflectance.

To estimate the illumination for an image of $N$ pixels $[R_i, G_i, B_i]$, each pixel is first classified as to whether or not it belongs to the class of gray pixels. To classify a pixel, the logarithm of each channel is taken producing $[log(R_i), log(G_i), log(B_i)]$, which is then projected onto the S axis of the LIS coordinate system via vector inner product. If the resulting value is less than a specified threshold value then the pixel is classified as gray.

The GSI method estimates the colour $[R_e, G_e, B_e]$ of an image's illumination according to

$$\begin{cases} R_e = \dfrac{1}{N}\sum_i w_i R_i \\ G_e = \dfrac{1}{N}\sum_i w_i G_i \quad if\ \ isgray([R_i, G_i, B_i])\ then \quad w_i = 1\ else\ w_i = 0 \\ B_e = \dfrac{1}{N}\sum_i w_i B_i \end{cases}$$

where '*isgray*' is the test that classifies pixels as gray or not.

An example of the GSI method is shown in Figure 3.2. The *isgray* test identifies as gray those pixels from Figure 3.2(a) that are shown in white in Figure 3.2 (b). The true scene illumination measured from a gray card is [0.2476, 0.2910, 0.4614]. The standard gray-world method averages the RGBs of all pixels so that the estimated illumination is found to be [0.4748, 0.2348, 0.2903]. The GSI method, however, averages only the RGB of pixels that pass the *isgray* test with the result that the illumination is estimated to be [0.2810, 0.3290, 0.3899]. Clearly, this latter estimate is much closer to the true value. This example shows the potential of the GSI method; rigorous tests are presented in the next section.

Figure 3.2 shows an example of the gray-pixel detection results. The detected pixels are marked in white in Figure 3.2(b). The chromaticities of these pixels in Figure 3.2(a) can be averaged to obtain the colour of the illuminant for AWB.

(a) (b)

**Figure 3.2 Gray-pixel Detection Results. (a) Input image; (b) Pixels identified as gray are indicated in white.**

## 3.4 Tests of GSI

The GSI method was implemented in MATLAB. To evaluate GSI's illumination estimation and compare it to other methods, the algorithm was tested on two datasets of real images. (*See Appendix B*) The first one includes the 321 images of the SFU dataset [5], which are of scenes in a laboratory setting. The second set is the much larger and more varied image collection that Ciurea et al. [13] built using a digital video camera.

In evaluating performance, the error measure is based on Euclidean distance and angular difference between the estimated and true illumination chromaticity values. (*See Appendix A*)

The first test uses Barnard's 321 images [5] captured using a calibrated SONY DXC-930 camera. For the synthetic case, RGB values are synthesized for the measured percent spectral reflectance of 24 Macbeth colour checker patches and the spectral power distributions of 102 illuminants at 15 different intensities values. Applying PCA to this data, we find the LIS axes as row vectors:

[0.5994 0.5871 0.5441],
[0.6421 0.0482 -0.7651],
[0.4729 -0.8132 0.3358].

To compute the LIS coordinates from real data, the RGB values from the gray card under the 11 different illuminants are used. These RGBs are then scaled by 15 different factors to create a RGBs that vary in intensity. PCA is applied to the logarithms of the resulting 165 RGBs. The LIS axes obtained are:

[0.6040 0.5807 0.5459],
[0.6429 0.0499 -0.7643],
[0.4711 -0.8126 0.3432].

Clearly, the two methods produce very similar results. The advantage of the real data method is that it is generally easier to collect images of a gray card under a dozen or so different illuminants than it is to determine a camera's spectral sensitivity functions.

Having determined the LIS coordinates, the next step is to proceed to test the GSI method. Since 321 is a small number of images, it is more reasonable to use leave-one-out cross-validation [19] in evaluating its performance and that of competing methods. Each method is trained on 320 of the images and tested on the one remaining image. This procedure is repeated a total of 321 times so that each image can be tested. In the case of GSI, the training consists of choosing the optimal *isgray* threshold minimizing the median angular error over the training set. Table 3.1 compares GSI performance to that of Support Vector Regression

[34] both on RGB data (3D) and chromaticity data (2D), to Shades of Grey [26] with the optimal choice of norm, to Max-RGB [22] which takes the maximum in each of the 3 colour channels as the illumination colour, and to standard Grayworld [9].

Our second test is based on the Ciurea et al. [13] dataset. Since the camera was uncalibrated, we used the real data method to calculate the LIS coordinates for it based on RGBs from the gray ball, which can be found at the bottom right corner in every image.

The original image database includes 11,346 images. Here, we test on two subsets extracted from the whole dataset: subset A contains 3581 images, and subset B 4080 images (See *Appendix B*). First, subset A is used for training and subset B for testing, then vice versa. The errors from both tests are combined in the entries in Table 3.2.

| **Method** | **SVR Dimension/ Norm Power** | **Angular Distance** | | | **$L_2$ Distance x100** | | |
|---|---|---|---|---|---|---|---|
| | | **Median** | **RMS** | **Max** | **Median** | **RMS** | **Max** |
| GSI | | 3.91 | 10.11 | 33.79 | 2.71 | 7.15 | 22.65 |
| SVR | 2D | 4.65 | 10.06 | 22.99 | 3.41 | 7.5 | 16.41 |
| | 3D | 2.17 | 8.069 | 24.66 | 3.07 | 6.3 | 16.03 |
| SoG | 6 | 3.97 | 9.027 | 28.70 | 2.83 | 6.21 | 19.77 |
| Max RGB | | 6.44 | 12.28 | 36.24 | 4.46 | 8.25 | 25.01 |
| GW | | 7.04 | 13.58 | 37.31 | 5.68 | 11.12 | 35.38 |

**Table 3.1 Comparison of GSI performance to that of 2D and 3D Support Vector Regression, Shades of Grey, Max RGB, and Grayworld. The results involve real-data training and testing on the 321 SONY images. Errors are based on leave-one-out cross-validation evaluation and are reported in terms of both the RMS angular chromaticity and distance error measures.**

| Method | Angular Degrees | | | Distance($x10^2$) | | |
|---|---|---|---|---|---|---|
| | Median | RMS | Max | Median | RMS | Max |
| GSI | 5.46 | 7.95 | 38.71 | 4.15 | 6.23 | 31.93 |
| 3D SVR | 4.91 | 7.03 | 24.80 | 3.62 | 5.16 | 18.62 |
| SoG | 6.71 | 8.93 | 37.01 | 4.83 | 6.59 | 27.99 |
| MAX RGB | 9.65 | 12.13 | 27.42 | 6.86 | 8.80 | 21.72 |
| GW | 6.82 | 9.66 | 43.84 | 5.25 | 7.82 | 45.09 |

**Table 3.2 Comparison of GSI error to 3D SVR, SoG, Max RGB, and Grayworld. The results involve real-data training and testing on disjoint sets of 7,661 images taken from the Ciurea data set.**

## 3.5 Conclusion

A new colour constancy method, GSI, is proposed that is based on detecting pixels corresponding to gray surface reflectance—which is not necessarily the same as gray image colour—and using their average image colour as an indicator of the colour of the overall scene illumination. The gray surfaces are found by first transforming the image RGB values to a new LIS coordinate system with axes that roughly correspond to luminance, illumination 'colour' and reflectance. In LIS coordinates, values of S near zero tend to be gray. Tests on real images show the GSI method works better than Shades-of-Gray, Grayworld and Max RGB. This method does not require training, and is substantially simpler to implement.

# 4: CONSTRAINTS ON SKIN COLOUR

The colour of facial skin changes significantly with changes in the light incident upon it. Is it possible to use human skin, as it is such a special surface material, as a reference to detect the incident light, in a way similar to the Max-RGB methods, where a white patch reflects the scene illumination? On the other hand, is it possible to correct these skin tones to canonical conditions by removing the effect of the light?

By removing the colour of the light from the skin, the revealed true skin colour has to look natural. Such a constraint may set an upper bound for maximum error in illumination estimation. Therefore, in this chapter I propose a simple and inexpensive model for (1) illumination estimation using the skin surface from a human face as a reference, also for (2) normalizing the skin tones. The proposed model is based on two assumptions: (1) The surface reflectance of human skin is modelled as a function of the melanin content in the epidermis and the hemoglobin content in the dermis; (2) Light source spectra are approximated by a Planckian blackbody radiator as a function of temperature.

## 4.1 Introduction

The factors affecting facial skin colour have been previously studied in the context of image rendering in computer graphics; face detection and tracking in computer vision; diagnosis in dermatology; and makeup and skin care in

cosmetics. Skin colour is the most discriminative of skin attributes and depends on the skin's pigmentation, blood microcirculation, roughness, sebum, and perspiration, as discussed by Barel et al. [2]. Efforts in correcting skin colours under different lighting conditions have been made by Soriano et al. [79] and Marguier et al. [59]. In previous human skin models, such as Shimada et al. [78], it has been suggested that the colour of human skin is mostly determined by the concentration of melanin in the epidermal layer combined with the content of hemoglobin in the dermal layer. The change of melanin content in skin (e.g., caused by exposure to UV) happens more slowly than the change of blood content (e.g., after bathing). However, even for fixed melanin and hemoglobin content, skin colour can change significantly and quickly as the lighting conditions change. Results in Section 4.4.1 confirm that the changes in skin colour induced by changes in illumination colour are much larger than those due to biological factors.

To analyse the spectral reflectance of human skin from different ethnic groups, Shimada et al. [78] and Tsumura et al. [85] have used independent component analysis (ICA) to extract two independent colour components of skin--- the melanin component and the hemoglobin component--- such that all skin chromaticities can be represented as a linear combination of the two components in log-chromaticity space. As mentioned above, the colour of skin in an image depends strongly on the colour of the incident light. To account for the induced shifts in skin colour, I propose a skin-illumination model in a linear space constrained by both the skin locus and the blackbody radiator locus. Based on

this model, the range of chromaticities of arbitrary skin under all illuminant colour temperatures can be transformed into a new coordinate system defined with two independent axes: an illumination axis as a function of colour temperature, and a skin axis as a function of melanin content. This linear space will facilitate colour-based applications such as skin detection and face tracking.

In these coordinates, it becomes clear that the axis of change in skin colour caused by the hemoglobin content is almost the same as that of blackbody radiators of varying colour temperature. As a result, it is difficult to analyze whether the redness of skin in an image is the result of high hemoglobin content versus light of low colour temperature. On the other hand, the axis of skin colour variation caused by changing melanin content is at a very different orientation from the axis of illumination change. This suggests that chromaticities of skin along the melanin axis will be approximately invariant to illumination change. Therefore, the skin colour appearing in an image captured under a light of unknown colour temperature can be normalized to what it would be under a standardized light by shifting its chromaticity along the illumination direction to the projection point on the melanin axis. Since the projection point is determined by the melanin content alone, not the light's colour temperature, the shifted skin chromaticity is illumination insensitive.

Let's assume specularity is absent and shading can be normalized by pixel intensity. If the skin pixels within an image can be identified, using the Viola-Jones face detector by Viola and Jones [86], for instance, then the proposed colour correction can be applied to the identified skin pixels, as well as the entire

image. Furthermore, if an image area has been identified as containing skin, then both the chromaticity of the incident illumination and the melanin content of the skin can be estimated directly. This possibly provides an upper bound on the maximum error in illumination estimation by the model. Test results are reported showing that the proposed skin-illumination model leads to good estimates of the chromaticity of the incident illumination.

## 4.2 Background on Skin Modelling

Many skin models [1][37][85] model the reflectance of skin by two layers: a thin surface layer (the epidermis) and a thicker layer (the dermis). As in the dichromatic reflectance model [73], interface reflection takes place at the epidermis surface and is a constant 5% (wavelength independent). The body reflection component is due to light entering the skin being absorbed and scattered within the two skin layers. The scattering in the epidermis is considered negligible, it mainly absorbs light, hence it has the properties of an optical filter. The absorption depends on the melanin content in the epidermis. In the dermis layer the light is both scattered and absorbed, where the absorption depends on the blood content. The optical properties of the dermis are basically the same for all humans. The variations of skin colour are thus determined by transmittance properties of the epidermis and dermis. Hiraoka et al. [39] has used a spectrum-based skin model to describe reflection spectrum, $S$, of arbitrary skin:

$$S(\lambda) = \exp[-\rho_m \alpha_m(\lambda) l_m(\lambda) - \rho_h \alpha_h(\lambda) l_h(\lambda) - \xi(\lambda)], \quad 4.1$$

Here, for a compound skin, $\rho_m$, $\rho_h$ and $\alpha_m(\lambda)$, $\alpha_h(\lambda)$ are the pigment densities and spectral cross-sections of absorbance of melanin and hemoglobin respectively. The variables $l_m(\lambda)$ and $l_h(\lambda)$ are the mean path lengths of photons in the epidermis and dermis layers. Variable $\xi(\lambda)$ stands for the scattering loss and the absorbance of chromophores other than melanin and hemoglobin. It is reasonable to consider $\xi(\lambda)$ as a constant because it is essentially independent of $\rho_m$ and $\rho_h$. Using the two spectral absorptions of melanin and hemoglobin, it is possible to model the skin spectral reflectance for any melanin content.

Several researchers have measured the spectral reflectance of skin. It is natural to consider a principal components analysis or independent components analysis on these spectra since earlier studies have shown these techniques to be very useful for colour constancy (Maloney & Wandell [58]), colour correction (Lenz et al. [56]), and segmentation (Hauta-Kasari et al. [37]) . More recently, physics-based skin modeling based on basis functions has been proposed [78][82][85]. Angelopoulou et al. [1] showed that the spectral reflectance of skin can be modelled by a linear basis based on the measured reflectance spectra of the back of hands and palm. Shimada et al. [78] use regression analysis based on the modified Beer–Lambert law as a method of measuring melanin and blood content in human skin using basis functions. Tsumura et al. [85] propose a technique through which hemoglobin and melanin pigment content are extracted

from a single skin colour image by independent component analysis (ICA). In their technique, the scattering in the skin is modelled in a simple linear form in the optical density domain in which inverse optical scattering is performed by a simple inverse matrix operation.

Colour-based skin modelling and detection methods have been proposed for computer vision in unconstrained environments with changing illumination, e.g., for real-time tracking of faces [38][45]. In many methods, an area of skin is defined as a gamut that contains all possible skin chromaticities observed by a colour camera under a certain range of illuminations, where no specific distribution is assumed for the skin chromaticities. The melanin content for one subject is not constant but has a certain range. The lower and upper limits of the blood content are rather constant for all ethnic groups. The skin chromaticity distribution is approximated by the area bounded by chromaticities at the lower and upper limits of melanin and hemoglobin content modelled with the reflectance spectrum. It is possible to model the obtained skin gamut with nonlinear functions describing the boundaries, or with a look-up table [61].

Skin colour modelling also depends on the camera, particularly the camera sensor sensitivities functions. Moreover, an important issue in skin colour detection is the dependence on the illumination conditions. Usually, the colour gamuts for skin modeling are obtained by simulating skin colours under different illuminations, and using chromaticity space (e.g., *rg* chromaticity) to eliminate the influence of intensity. None of the colour spaces, however, normalizes for changing illumination colour. The problem of changing illumination colour has to

be solved by other means, for example, using an automatic white balance method. Nevertheless, all automatic white balancing methods are based on assumptions that may not be met in practice, and hence are subject to error.

## 4.3 The Proposed Skin-Illumination Model

As reviewed in Section 1.2, previous research by Finlayson and Schaefer [25] has shown that, given narrowband sensors, blackbody radiation models not only the spectra of direct sunlight and tungsten light bulbs, but also that of common daylight conditions. The camera's response to blackbody illumination by Wien's approximation assuming narrowband sensors can be expressed as in Equation 1.4. For simplification, this multiplicative model becomes additive in log colour space,

$$\log[E(\lambda)] = \log(I) + \log(c_1) + \log(\lambda^{-5}) - \frac{c_2}{T\lambda}, \quad 4.2$$

Similarly, the skin model specified in Equation 4.1 can also be treated linearly in log space:

$$\log[S(\lambda)] = -\rho_m \alpha_m(\lambda) l_m(\lambda) - \rho_h \alpha_h(\lambda) l_h(\lambda) - \xi(\lambda) \quad 4.3$$

According to the colour formation in Equation 1.3, with narrow-band sensors, combining Equation 4.2 and 4.3 provides us:

$$\log[P_k] = \log[S(\lambda_k)] + \log[E(\lambda_k)]$$
$$= -\rho_m \alpha_m(\lambda_k) l_m(\lambda_k) - \rho_h \alpha_h(\lambda_k) l_h(\lambda_k) - \xi(\lambda_k) + \log(I) + \log(c_1) + \log(\lambda_k^{-5}) - \frac{c_2}{T\lambda_k} \quad 4.4$$

Let **Π** represent the camera signals from the RGB channels in log space, based on Equation 4.4,

$$\mathbf{\Pi}(\rho_m, \rho_h, \tau, b) \approx \rho_m \boldsymbol{\sigma}_m + \rho_h \boldsymbol{\sigma}_h + \tau\boldsymbol{\omega} + b\mathbf{1} + \mathbf{c}, \quad 4.5$$

where

$$\mathbf{\Pi} = [\log(P_R), \log(P_G), \log(P_B)]^t,$$

$$\boldsymbol{\sigma}_m = [\alpha_m(\lambda_R) l_m(\lambda_R), \alpha_m(\lambda_G) l_m(\lambda_G), \alpha_m(\lambda_B) l_m(\lambda_B)]^t,$$

$$\boldsymbol{\sigma}_h = [\alpha_h(\lambda_R) l_h(\lambda_R), \alpha_h(\lambda_G) l_h(\lambda_G), \alpha_h(\lambda_B) l_h(\lambda_B)]^t,$$

$$\boldsymbol{\omega} = [\frac{c_2}{\lambda_R}, \frac{c_2}{\lambda_G}, \frac{c_2}{\lambda_B}]^t,$$

$$\mathbf{c} = [\log(c_1) + 5\log(\lambda_R) - \xi(\lambda_R), \log(c_1) + 5\log(\lambda_G) - \xi(\lambda_G), \log(c_1) + 5\log(\lambda_B) - \xi(\lambda_B)]^t,$$

and $b = \log(I)$, $\tau = 1/T$.

The observed signal $\mathbf{\Pi}$ therefore is represented by the weighted linear combination of the four vectors $\boldsymbol{\sigma}_m, \boldsymbol{\sigma}_h, \boldsymbol{\omega}$ and **1** plus a bias term **c**. These vectors correspond to the melanin, hemoglobin, the illumination chromaticity and the illumination brightness axis, respectively. The melanin and hemoglobin content, the blackbody temperature, and the intensity only vary along these four

directions. The negative signs in Equation 4.4 are ignored in Equation 4.5, since signs only indicate directions, and the variable can go either direction along an axis. The variable $\tau$ is the inverse of the temperature *T* and is measured in mired ($10^6 K^{-1}$). The vectors **ω** and **c** are sensor dependent and therefore stay constant for a given camera. This coordinate is illustrated in Figure 4.1, where a set of skin colours in log space are spanned by melanin and hemoglobin axes.

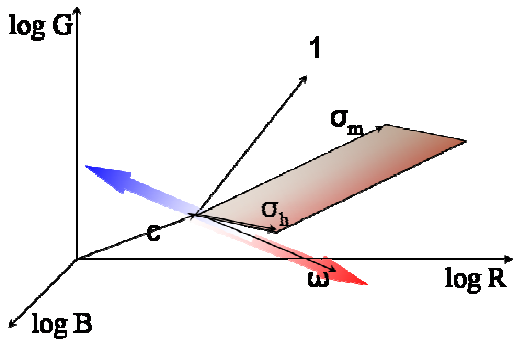



**Figure 4.1 An illustration of the skin-illumination coordinate in 3D log RGB space.**

My proposed skin model is based on simplifying Equation 4.5 for varying illumination colour temperature and melanin content. The result is

$$\mathbf{\Pi}(\rho_m, \tau) \approx \rho_m \boldsymbol{\sigma}_m + \tau \boldsymbol{\omega} \qquad 4.6$$

Here, the hemoglobin term in Equation 4.5 is dropped because the hemoglobin axis almost coincides with the illumination axis **ω**, both varying in the red-white-blue direction. Also, the brightness term is eliminated by intensity normalization. Equation 4.6 suggests that the melanin basis, $\boldsymbol{\sigma}_m$, and the blackbody illumination basis, $\boldsymbol{\omega}$, span the chromaticity space of arbitrary skin

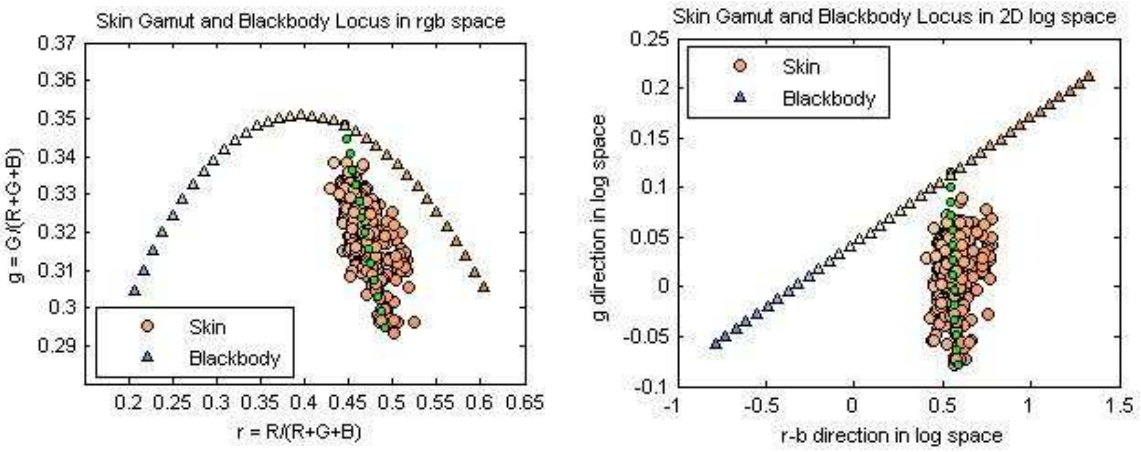
under different illuminations. It describes the skin-illumination model proposed in this chapter. As the blackbody radiator illumination basis, $\boldsymbol{\omega}$, can be easily calculated mathematically, the melanin basis, $\boldsymbol{\sigma}_m$, is obtained by using Independent Component Analysis(ICA) to separate it from hemoglobin basis [85], given a set of skin reflectance data.

## 4.4 Tests of Skin-Illumination Model

### 4.4.1 Skin Spectrum Analysis for Illumination

The first test is to analyze the set of 384 skin reflectance spectra of people of different ethnicity. A standard Independent Component Analysis tool called JADE[11] is used on this normalized skin data in logarithm space with gamma equal to one to extract two independent basis spectra – the melanin and hemoglobin – as predicted by the skin-illumination model, shown in Equation 4.3.

Figure 4.2 shows how the gamut (shown as a cloud of dots) of log chromaticities of real skin under canonical condition (pure white light) forms an area spanned by the hemoglobin and melanin axes, where the spectra were converted into standard RGB (sRGB) colour space. In Figure 4.2(b) the blackbody locus in log space becomes more or less straight and the hemoglobin axis (not shown) is roughly parallel to it. However, varying the melanin content leads to skin chromaticities (green dots) lying along a line that makes a sharp angle with respect to the blackbody locus.

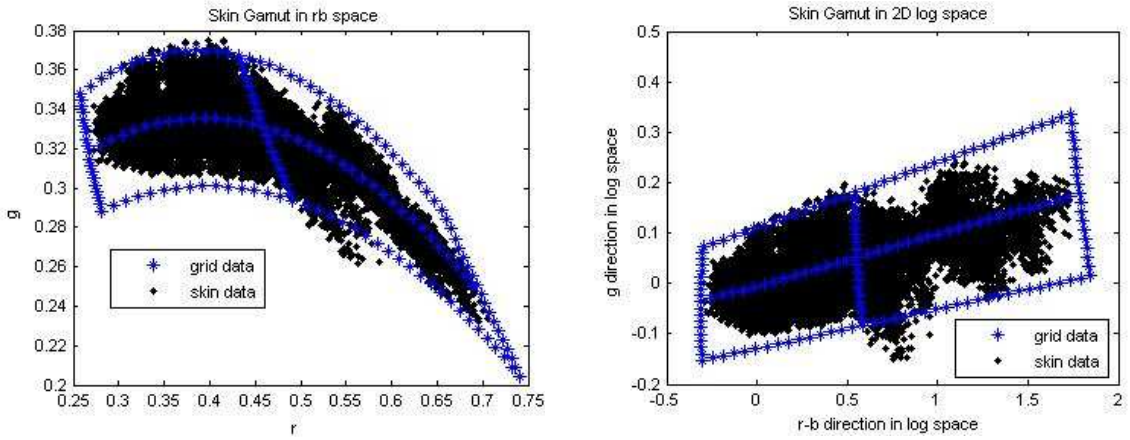



(a) (b)

**Figure 4.2 Real and synthesized skin reflectance chromaticities, along with blackbody radiator chromaticities plotted in the 2D plane. (a) Planckian blackbody locus and the gamut of 384 skin samples in rg-chromaticity space. Chromaticities of synthetic skin reflectance with only melanin variance are plotted with green dots. (b) Planckian blackbody locus and 384 skin samples in 2D log space.**

As discussed above, skin chromaticities can be modelled as a linear combination of two independent components in logarithm space. An important consequence of this is that it means that, given a skin chromaticity, one can decompose it into one component due to its melanin content, and a second component due to the incident illumination. From this decomposition, we thereby obtain a reasonable estimate of the colour of the incident illumination. To demonstrate that this works in practice, we first identify the boundaries and the axes of the skin-illuminant gamut to estimate the illumination for arbitrary skin types. Both spectra of real light sources and synthesized blackbody radiator illuminants are used. The blue stars in Figure 4.3(a) define a grid in *rg*-chromaticity space obtained with synthetic skin reflectances and Planckian illuminants. The black dots represent chromaticities from colour signal spectra synthesized as the product of 67 real light sources and 384 real skin

reflectances. By inspection, Figure 4.3 (b) shows how this grid is considerably straightened out into a more linear form in log space.




(a) (b)

**Figure 4.3 The grid of skin data denoted as blue stars, and test skin data denoted as black dots, plotted in (a) *rg*-chromaticity space, and (b) 2D log space.**

The non-linearity of the data in Figure 4.3 (a) makes decomposition into illumination and melanin components difficult. On the other hand, the strongly linear nature of Figure 4.3 (b), that is, this 2D space is spanned by two linear axes, makes linear projection for the colour temperature possible. Given this linear space, applications such as colour-based skin tracking can be simplified since all skin colours fall within a simple quadrilateral. In my test, this quadrilateral of skin colour boundaries (the blue dots in Figure 4.3 (b)) is transformed into a rectangle, so that the projections of the skin data (black dots) on the two axes of the rectangle tell the chromaticities of skin melanin and illumination.

This model assumes that the camera sensitivity functions are very narrowband. Given this assumption may be violated in practice, I tested the

above procedure to estimate illumination chromaticities with six different sets of sensor sensitivity functions: Sony DXC 930, Sony DXC 755p, Sony 828, Kodak 460, the Vos et al. cone fundamentals [87] and CIE XYZ. The estimation error (*See Appendix A*) between the estimated and real illuminations is shown in Table 4.1.

| Sensor Type | Angular Error | | | L2 Error | | |
|---|---|---|---|---|---|---|
| | Median | Mean | Max | Median | Mean | Max |
| Sony DXC 930 | 2.40 | 2.62 | 8.16 | 0.018 | 0.020 | 0.066 |
| Sony DXC 755p | 2.13 | 2.40 | 7.40 | 0.017 | 0.019 | 0.072 |
| Sony 828 | 1.20 | 1.44 | 7.03 | 0.0095 | 0.012 | 0.068 |
| Kodak 460 | 1.05 | 1.34 | 6.01 | 0.0086 | 0.011 | 0.058 |
| Cone LMS | 1.77 | 2.00 | 7.49 | 0.011 | 0.013 | 0.045 |
| CIE XYZ | 1.54 | 1.81 | 6.43 | 0.011 | 0.013 | 0.040 |

**Table 4.1 The illumination estimation error for 25,728 spectra calculated from 384 skin samples and 67 real illuminations. The camera responses are calculated based on four sets of camera sensors, cone LMS sensitivity functions, and CIE XYZ sensitivity functions.**

Table 4.1 shows that the model works well on real data. The errors that do arise originate from three sources: a) the simplified skin model, which assumes all variations in skin colour are due to variations in melanin and hemoglobin content, b) non-blackbody light sources, c) and broadband camera sensor sensitivities.

### 4.4.2 Skin Tone Correction

In addition to illumination estimation, the same model can also be used for skin tone correction. The idea is to shift the entire skin colour gamut of a given face image along the direction of the illumination basis until it is in the canonical condition (i.e., as it would be under pure white illumination). In this section,

testing was done using the University of Oulu Physics-Based(UOPB) Face Database provided by Marszalec et al. [60], which contains 357 measured spectra of human faces of 119 individuals of different races. In addition, this database contains images of 125 different individuals. An image series for one

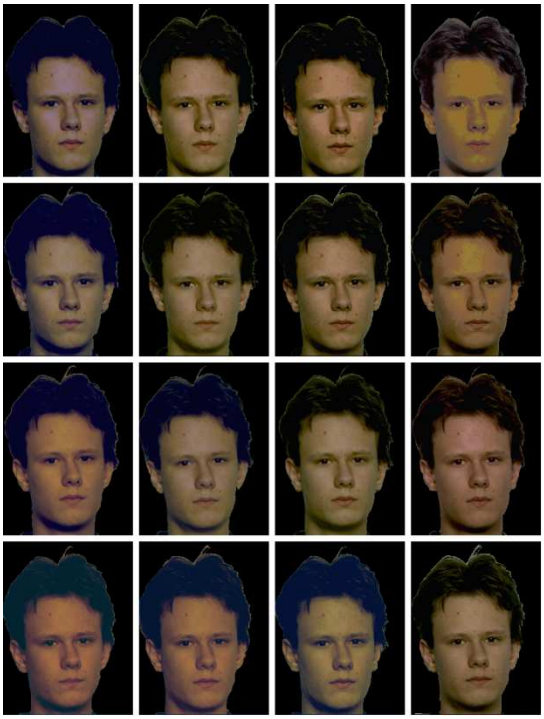


**Figure 4.4 Results based on an individual (no. 96) in the UOPB database. Skin tones are corrected based on the proposed model on a series of 16 face images under different camera calibration and illumination conditions (faces segmented from the background).**

person contains 16 frontal views, each of which is captured under a different combined illuminant and camera calibration condition. In my test, skin pixels in each linearized image are manually selected. Potentially, a face detector could be used for automated selection. The average of all skin pixels is translated to a point on the estimated melanin axis, along the blackbody axis. Then this point is shifted along the melanin axis until it is within the melanin value range of this

specific database if it is not (e.g., due to non-blackbody radiation illumination, camera overflow, etc). The same translation is then applied to all pixels of the entire image. As illustrated in Figure 4.4, the corrected images show that the skin tones of the face in the image are invariant to illumination change.

## 4.5 Conclusion

I have introduced a model of skin colour under varying illumination that allows easy decomposition of a skin chromaticity into two components: one due to the illumination, and the second due to the melanin content. As a computationally inexpensive model, the proposed method is useful for estimating the colour of the light illuminating the skin, and for normalizing skin in images for which the colour of the illumination is not known. This model assumes narrowband sensors, blackbody illuminants, and melanin and hemoglobin contents as being the dominant factors in skin colour; however, tests show that the model succeeds relatively well even when the assumptions are violated. In addition, this model provides a tool to normalize skin tones to be invariant to illumination change. The normalization is accomplished by shifting the colour of the entire image so that skin pixels lie on the pre-defined melanin axis.

# 5: SOLVING THE DICHROMATIC MODEL

In this chapter, attention is focused on the problem of estimating the colour of the light by exploiting the principles of colour image formation laid down by the dichromatic reflection model [73]. A new illumination-estimation method is proposed based on this model combined with Hough transform processing. Researchers have shown that using the dichromatic reflection model under the assumption of neutral interface reflection, the colour of the illuminating light can be estimated by intersecting the dichromatic planes created by two or more differently coloured regions. The proposed new method employs two Hough transforms in sequence in RGB space. The first Hough Transform creates a dichromatic plane histogram representing the number of pixels belonging to dichromatic planes created by differently coloured scene regions. The second Hough Transform creates an illumination axis histogram representing the total number of pixels satisfying the dichromatic model for each posited illumination axis.

Many of the existing methods rely on the assumption that there are sufficiently large, connected regions of a single, highly specular material in the scene. The proposed method should overcome the limitations of previous approaches that include requirements such as: that the number of distinct surfaces be known in advance, that the image be pre-segmented into regions of uniform colour, and that there exist distinct specularities. Comparing the

performance of the proposed approach with previous non-training methods on a set of real images, the proposed method yields better results, while requiring no prior knowledge of the image content.

## 5.1 Dichromatic Reflection Model

Dichromatic reflection model states that, in RGB space, the colours reflected by an inhomogeneous dielectric material lie on a plane that is spanned by two characteristic colours; namely, the colour of the specular component reflected from the air-surface interface, and the colour reflected from the body of the material. If neutral interface reflection is assumed [54], then the chromaticity of the specular reflection is the same as that of the illuminating light. As a result, the colour of the illuminant can be estimated by intersecting the planes that the set of RGBs from two or more different materials describe.

According to the dichromatic reflection model for inhomogeneous dielectric objects, the colour signal is composed of two additive components, one being associated with the interface reflection and the other describing the body reflection part [73], so it can be described as

$$C(\theta,\lambda) = m_I(\theta)C_I(\lambda) + m_B(\theta)C_B(\lambda) \qquad 5.1$$

where $C_I(\lambda)$ and $C_B(\lambda)$ are the spectral power distributions of the interface and the body reflection respectively, and $m_I$ and $m_B$ are the corresponding weight

factors depending on the geometry $\theta$, which includes the incident angle of the light, the viewing angle, and the phase angle.

Suppose $R$, $G$, and $B$ are the red, green, and blue pixel value outputs of a digital camera, then each colour vector $(R,G,B)^T$ is determined by a linear combination of a surface reflection component $(R_i, G_i, B_i)^T$ and a body reflection $(R_b,G_b,B_b)^T$ component. Equation 5.1 shows that the colour signal can be expressed as the weighted sum of these two reflectance components. Thus the colour signals for an object are restricted to a plane:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = w_i \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} + w_b \begin{pmatrix} R_b \\ G_b \\ B_b \end{pmatrix} \qquad 5.2$$

If we consider two objects within the same scene (and assume that the illumination is constant across the scene) then we end up with two RGB planes. Both planes, however, contain the same illuminant RGB. This implies that their intersection must be the illuminant itself. Although theoretically sound, dichromatic colour constancy algorithms do not always perform well on real images. For example, image noise may cause the intersection of two dichromatic lines planes to change quite drastically. In addition, textured and non-uniform surfaces may mean that the distribution of colours does not lie on distinguishable dichromatic planes.

Based on the dichromatic reflection model [73], Lee [53] introduced a method for computing the scene illuminant chromaticity by intersecting lines in chromaticity space. Although Lee's method performs sufficiently well on synthetic

images of spheres, its application to real-world scenes is sensitive to noise and inhomogeneities such as textured surfaces. Another approach using dichromatic regions of different coloured surfaces is called colour line search [52]. It involves automatic detection of specular regions, a Hough transform step, and consistency check step. However, this approach requires correct detection of regions of interest, and can fail when specular highlights are incorrectly identified or absent from the scene. The method proposed by Tan et al.[83] describes an inverse-intensity chromaticity space in which the correlation between illumination chromaticity and image chromaticity can be analyzed. Once again, this method relies on correctly identifying the highlight regions, and does not perform any better than competing methods. The proposal to solve for the intersection of the dichromatic planes directly as described by Toro et al.[84] assumes that in any patch of the given image, a fixed number of different materials coexist. The illumination colour can be calculated by solving a set of simultaneous linear equations using a Veronese projection of multilinear constraints. However, this approach assumes that the number of different surfaces in an image is already known. It also does not yield any better results than previous methods when applied to real images. The method proposed by Schaefer in [72] achieves competitive results, but the approach requires the illumination to be from a set of known light sources.

## 5.2 Hough Transform as a Solution

In the proposed solution, I use a Hough Transform [67] for dichromatic plane detection. In the 3D case, a plane is parameterized as:

$$\rho = (x\cos(\phi) + y\sin(\phi))\cos(\theta) + z\sin(\theta) \qquad 5.3$$

where $\rho$ is the distance between a plane and the origin, $\phi$ is angle relative to the *z*-axis, $\theta$ is angle relative to the *y*-axis. In the discrete case, the parameter space $(\rho, \phi, \theta)$ is quantized into bins, so the Hough Transform is represented as a three-dimensional histogram.

According to the dichromatic model, *all dichromatic planes should pass through the origin*. This implies that the "distance" $\rho$ in Equation 5.3 is zero, so the RGBs reflected from a dichromatic surface satisfy the following parametric plane equation

$$R\cos(\phi)\cos(\theta) + G\sin(\phi)\cos(\theta) + B\sin(\theta) = 0 \qquad 5.4$$

All pixels from the same surface belong to a single plane defined by the two angles $\phi$ and $\theta$. Hence a 2D Hough Transform can be used to create a *dichromatic plane histogram* $\mathbf{H_1}$. Each bin of the histogram represents the number of pixels belonging to a distinct dichromatic plane specified by the pair of angles $(\phi, \theta)$ satisfying Equation 5.4. A high value in the histogram implies the

existence of this dichromatic plane in the image, while a lower value implies its absence.

Since the illumination axis is the intersection of all dichromatic planes, it must be perpendicular to the normal of each dichromatic plane. Therefore, the axis perpendicular to the normals of the largest number of dichromatic planes is a good candidate for the illumination axis. To determine it, we use a second Hough Transform to create an *illumination histogram* $\mathbf{H_2}$ based on the data from $\mathbf{H_1}$. To use the data from $\mathbf{H_1}$, we first calculate the normals of the dichromatic plane in the dichromatic plane set. The normal of a dichromatic plane described by $(\phi,\theta)$ is $\mathbf{n} = (u,v,w)$, where

$$\begin{aligned} u &= \cos\theta\cos\phi \\ v &= \cos\theta\sin\phi \\ w &= \sin\theta \end{aligned} \qquad 5.5$$

When an illumination axis is represented in polar form by the two angles $\alpha$ and $\beta$, it is perpendicular to the normal $\mathbf{n}$ of a dichromatic plane if and only if it satisfies the following equation:

$$u\cdot\cos(\beta)\cos(\alpha)+v\cdot\sin(\beta)\cos(\alpha)+w\cdot\sin(\alpha)=0 \qquad 5.6$$

Based on Equation 5.6, a second 2D Hough Transform parameterized by $(\alpha, \beta)$ is used to create an *illumination histogram*. The count for a bin in the illumination histogram is calculated in the following manner. When the normal of

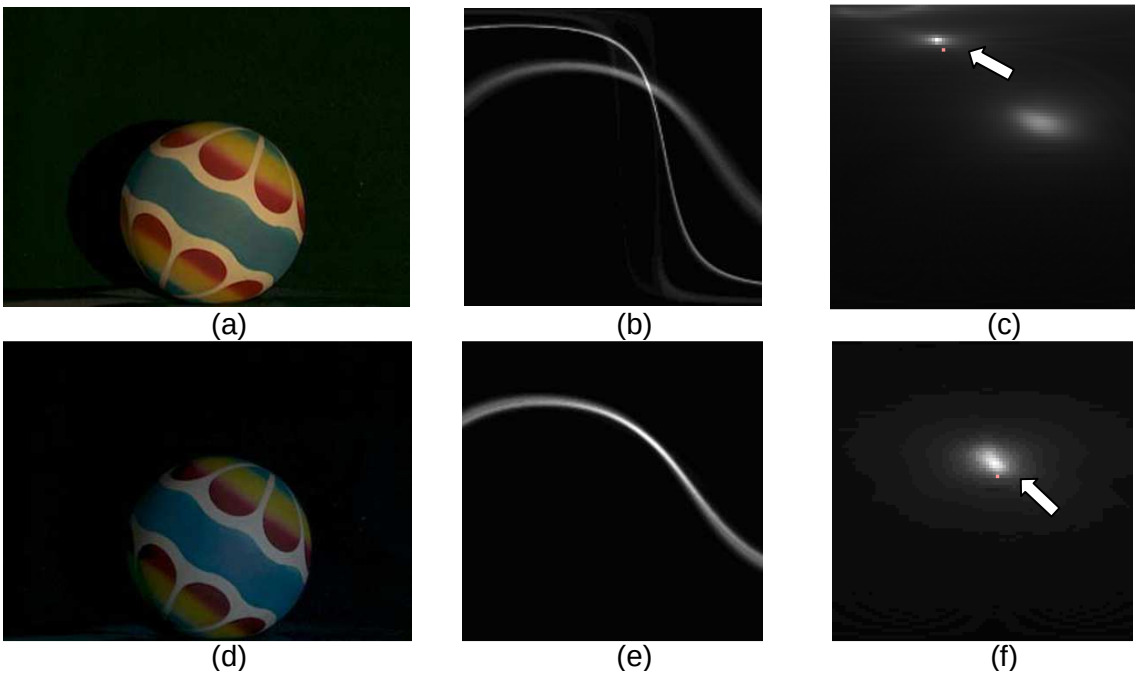



**Figure 5.1 Two images of the same object under different illuminants. (a) and (d) are the original images; (b) and (e) are the dichromatic plane histograms of the images in (a) and (d) respectively, after the first Hough Transform, with $\phi$ and $\theta$ ranging from 0 to 179; (c) and (f) are the illumination histograms of the images in (a) and (d) respectively, after the second Hough Transform, with $\alpha$ and $\beta$ range from 0 to 89. The arrows in the two figures indicate the locations of the true illuminants. The correspondence between the true illumination and the histogram peaks is evident.**

dichromatic plane ($\phi$, $\theta$) is perpendicular to illumination axis ($\alpha$, $\beta$), the count from the corresponding bin of $\mathbf{H_1}$ is added to that of the corresponding bin of $\mathbf{H_2}$. The bin count of a bin *b* in the resulting histogram indicates the number of image pixels that conform to the dichromatic model under the illumination that *b* represents in that contributing pixels all come from a collection of dichromatic planes that share a common intersection, and a common intersection represents a shared illumination. Therefore, a high bin count in $\mathbf{H_2}$ implies a high probability that the bin corresponds to the true scene illumination. Figure 5.1(c) and Figure 5.1(f) provide two examples of *illumination histograms* for the same object under two different illuminations.

In principle, the correct illumination can be determined by searching for the global maximum in the *illumination histogram*. However, due to noise and the non-dichromatic properties of some surfaces that may be present in the image, the global maximum of the *illumination histogram* does not always correctly correspond to the true scene illumination. Although the global maximum may not always indicate the correct illumination, generally one of the local maxima will. Hence the problem becomes how to select between the local maxima. Our strategy is to select the local maximum inside a bounding disc centered at the illumination as estimated by another illumination-estimation method. In particular, in the tests reported here we use the Shades of Gray (SoG) method [26]. The disc radius is based on the average and standard deviation of the error of this method.

The complete estimation consists of the following steps:

1. Normalize the image **I** (scale intensities, remove dark pixels, etc)
2. Transform 3D pixels in **I** into dichromatic Hough space $\mathbf{H_1}$ by Equation 5.4
3. Transform $\mathbf{H_1}$ into illumination Hough Space $\mathbf{H_2}$ using Equation 5.5 and 5.6
4. Estimate image illumination *L* by SoG
5. Find the nearest local maximum in $\mathbf{H_2}$ inside a bounding disc centered at *L*
6. Convert polar coordinate representation of this local maximum into chromaticity coordinates

In summary, the two Hough Transforms can be thought of as two voting procedures. First, each pixel votes for the candidate dichromatic planes that pass

through it. Second, each dichromatic plane in turn casts a weighted vote (weighted by the number of pixels on that plane) for each candidate illumination axis that passes through it. Finally, the illumination axes that receive the highest votes are considered likely candidates for the true illumination.

## 5.3 Test Results

The proposed method was tested on the Simon Fraser University colour image database [5], which contains 321 images of 32 scenes under 11 different illuminants (*See Appendix B*). In this test, an image is resized to 200x200 and normalized such that the range of intensity in any image is [0, 255], and then the first Hough transform is applied to all pixels (excluding pixels over 250 or under 10). The space of planes is defined by angles $(\phi,\theta)$, whose values are integers in [0 to 179]. The result of the first Hough transform, $\mathbf{H_1}$, is therefore a 180x180 2D histogram as shown in Figure 5.1(b) and Figure 5.1(e). The illumination axis space is defined by angles $\alpha$ and $\beta$ with integer values in [0, 89]. Hence, the illumination histogram $\mathbf{H_2}$ calculated by a Hough transform of $\mathbf{H_1}$ is a 90x90 2D histogram (Figure 5.1(c) and Figure 5.1(f)). The maximum in $\mathbf{H_1}$ is then found in a 13x13 bounding box centred at the illumination axis estimated by Shades of Grey [42].

The performance is evaluated in terms of the angular difference in degrees between the RGB of the estimated and actual illumination(*See Appendix A*). In Table 5.1, the proposed approach shows good performance when compared to competing illumination-estimation methods [3][42][84] [89][93][94].

| Methods | Training Required | Median Angular Error |
|---|---|---|
| Grey World | no | 7.0 |
| Max-RGB | no | 6.5 |
| Multilinear Constraint | no | 5.8 |
| Shades of Grey (n=6) | no | 3.7 |
| Grey Edge | no | 3.2 |
| 2nd order Grey Edge | no | 2.7 |
| GSI | no | 3.5 |
| Colour by Correlation | yes | 3.2 |
| Neural Networks | yes | 7.8 |
| TPS | yes | 0.6 |
| 2D SVR | yes | 4.7 |
| 3D SVR | yes | 2.2 |
| 3D Hough Transform | no | 1.7 |

**Table 5.1 Comparison of performance of the proposed method with that of other non-training methods ( Grey World, Max RGB, Multilinear Constraint, SoG, GSI, Grey Edge, 2$^{nd}$ order Grey Edge) and training methods ( Colour by Correlation, TPS, SVR 2D and SVR 3D, Neural Networks) measured in terms of median angular errors based on the SFU image dataset of 321 images. The entries for GW, Max RGB, SoG, Gray-Edge, 2nd Gray-Edge, Colour-by-Correlation, and NN are reproduced from Table II, page 2211 of [89].**

In general, the total complexity of our method is O($NM$+$MK$), where $N$ is the number of pixels in an image; $M$ is the cardinality of the candidate dichromatic plane set; $K$ is the cardinality of the candidate illumination axis set. In the test, the dichromatic planes and illumination axes were searched exhaustively with $M$ = 180x180 and $K$ = 90x90.

## 5.4 Conclusion

The presented illumination-estimation method uses the constraints provided by the dichromatic model in a new and quite robust way. The method is based on two Hough transform voting procedures. First, each image pixel votes for every dichromatic plane it could fall on. This results in a 2D histogram representing the likelihood of each plane. Second, each dichromatic plane votes for each candidate illumination axis that could pass through that plane. Finally,

an illumination axis is chosen from among those receiving the highest number of votes based on the resulting illumination being close to that of the SoG illumination estimate. This robust method creates a 2D *illumination axis histogram* that represents the likelihood of the possible illuminations. My approach makes no assumption about the number of surfaces or the surface colours, yet performs well in comparison to the other methods tested.

# 6: REFLECTANCE RECOVERY VIA MATRIX FACTORIZATION

In this chapter, a new approach to illumination estimation for colour constancy and automatic white balancing is developed based on the technique of Nonnegative Matrix Factorization (NMF) with sparseness constraints (NMFsc). In terms of accuracy, the method proposed in this contribution is no better than existing methods; however, it is interesting in that it approaches the problem from an entirely different mathematical perspective, a perspective in which the input colour image is viewed as a matrix to be factored subject to the constraints of non-negativity and sparseness. The resulting factors represent the scene's illumination (possibly spatially varying) and its reflectance. The nonnegative constraint on the factorization is important because illumination and reflectance are both nonnegative physical quantities. The sparseness constraints--- the reflectance component is sparsely encoded, while the illumination image is non-sparse--- force the factorization to obtain an illumination component that varies only slightly across the scene, while allowing the reflectance component to be more varied. The assumption that the illumination chromaticity remains either relatively constant or completely constant across the scene is a common to virtually all colour constancy methods to date.

One feature of the NMFsc-based method is that, like Retinex, it provides an estimate of the illumination and reflectance at every pixel. Although there are

still a few parameters to tune for optimal performance, the NMFsc illumination estimation method, like a few methods [9][22][26][33], avoids the major training step required by the many other methods that rely on image statistics [10] [24][34][70] or finite-dimensional models of spectra [58]. Finally, unlike methods where full-sized images are required, NMFsc can be applied to only a subset of the image pixels. This be helpful when only thumbnail images are available, or to reduce computation time. In the tests section, results are reported for a large set of images with the NMFsc method applied to full-sized images, to reduced 8x8 image thumbnails, and to randomly sampled image pixels.

## 6.1 Background on Non-Negative Matrix Factorization

Non-negative matrix factorization creates a non-negative approximation to a given set of non-negative input data that represents the data in terms of a linear combination of non-negative basis features. Sparseness constraints can be imposed to change the results of the factorization. Hoyer [44] provides an excellent introduction to non-negative matrix factorization, and the use of sparseness constraints. In the context of this chapter, factorization is used to re-represent the log image data in terms of a linear combination of log illumination and log reflectance.

Let us divide an image into $T$ sub-windows of $N$ pixels. Following Hoyer [44], assume that the data consists of $T$ measurements of $N$ non-negative scalar variables represented as $N$-dimensional measurement vectors $\mathbf{v}^t$ ($t$ = 1, . . . ,$T$). An $M+1$-dimensional linear approximation for each data vector is given by

$$\mathbf{v}^t \approx \sum_{i=0}^{M} \mathbf{w}_i h_i = \mathbf{W}\mathbf{h}^t, \quad 6.1$$

where **W** is an $N \times (M+1)$ matrix containing the basis vectors $\mathbf{w}_i$ as its columns, and $\mathbf{h}^t$ is the vector of scalar coefficients $h_i$. Arranging the measurement vectors $\mathbf{v}^t$ as columns of an $N \times T$ matrix **V**, we can write

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}, \quad 6.2$$

where each column of **H** contains the coefficient vector $\mathbf{h}^t$ corresponding to the measurement vector $\mathbf{v}^t$. Writing it in this form makes it apparent that this linear data representation is simply a factorization of the data matrix. Similarly, principal component analysis(PCA) can also be viewed as a matrix factorization method, but with a different choice of objective function than will be used in non-negative matrix factorization. NMF forces all entries of **W** and **H** to be non-negative. The advantage of this is that it means that the measurement data is modeled in terms of additive components only. Since imagery is based on light and there are no subtractive lights, non-negativity is a desirable feature of any model of it.

The idea of sparse coding originates in the neural network literature [27]. A sparse code is one that is based only a few neurons being active, with the majority being inactive. In the case of the representation defined in Equation 6.1, sparseness means that only a few of the elements of each $w_i$ are significant, with the majority being near zero. The expectation is that a sparse encoding will be

based on meaningful, somewhat localized features. In terms of our goal of separating the reflectance component of an image from its illumination component, we expect reflectance to be well represented by localized features, hence by a sparse coding, and illumination to be represented by global features, hence a non-sparse coding.

Hoyer [44] extended NMF to find non-negative factorizations with a specifiable degree of sparseness, and provides the following definition of NMF with sparseness constraints (NMFsc). Given a non-negative data matrix **V** of size $N \times T$, find non-negative matrices **W** of size $N \times (M+1)$ and **H** of size $(M+1) \times T$ minimizing

$$E(\mathbf{W},\mathbf{H}) = \| \mathbf{V} - \mathbf{WH} \|^2 \qquad 6.3$$

under optional constraints

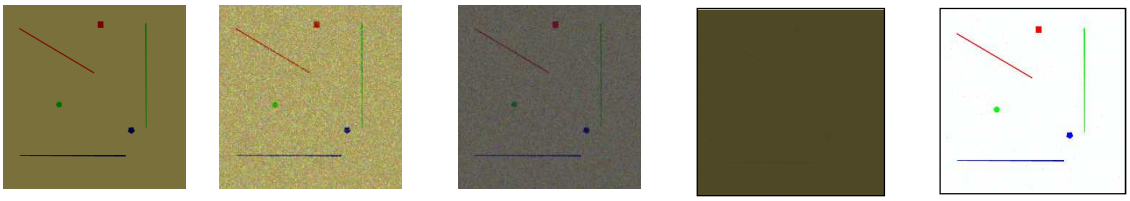
$$sparseness(\mathbf{w}_i) = S_w, \quad \forall i$$
$$sparseness(\mathbf{h}_i) = S_h, \quad \forall i,$$

where $\mathbf{w}_i$ is the $i^{th}$ column of **W** and $\mathbf{h}_i$ is the $i^{th}$ row of **H**. The three variables M, $S_w$ and $S_h$ are user controlled parameters, with M+1 denoting the number of components, and $S_w$ and $S_h$ denoting the desired sparseness of **W** and **H**.

Hoyer [44] uses a sparseness measure based on the relationship between the $L_1$ norm and the $L_2$ norm as

$$sparseness(\mathbf{x}) = \frac{\sqrt{n} - \frac{\Sigma|x_i|}{\sqrt{\Sigma x_i^2}}}{\sqrt{n}-1}, \qquad 6.4$$

where *N* is the dimensionality of **x**. $\Sigma|x_i|$ is the $L_1$ norm, and $\sqrt{\Sigma x_i^2}$ is the $L_2$ norm. This function evaluates to unity if and only if **x** contains a single non-zero component, and takes a value of zero if and only if all components are equal. According to the above definition of sparseness, a vector with all components almost equal should have a very small sparseness value. This allows us to use Equation 6.4 to measure smoothness of a vector and hence to enforce smoothness during factorization.



(a) (b) (c) (d) (e)

**Figure 6.1. Comparison of basis feature vectors obtained by NMF without sparseness versus NMFsc with sparseness for M+1=2. (a) Input image; (b-c) the two NMF basis features(without sparseness constraints); (d) NMFsc first basis feature constrained to have low sparseness; (e) NMFsc second basis feature constrained to have high sparseness.**

NMF with and without sparseness constraints has had empirical success in learning meaningful features from a diverse collection of real-life data sets [55]. For example, in Figure 1 of Hoyer [44], which shows how, when applied to a data set of face images, the representation by NMFsc consists of basis vectors encoding the intuitive features of faces such as the mouth, nose, and eyes. Figure 6.1 (of this thesis) shows an example of the difference in the basis vectors

obtained with and without sparseness constraints for the example of a 2-dimensional model (i.e., *M=1*).

## 6.2 The Model of Illumination and Reflectance

For a particular pixel in a colour image, the linear RGB sensor response is defined by the model in Equation 1.3. Taking the logarithm on both sides of the Equation 1.3 provides us a new equation in terms of addition

$$\log(p_k) = \log[E(\lambda_k)] + \log[S(\lambda_k)], \quad k = R, G, B \qquad 6.5$$

This has the advantage that the non-linear multiplicative combination of the illumination and reflectance becomes linear. Note also that if, as is common in digital cameras, the camera applies a non-linear 'gamma' function to the initial linear sensor responses such that the final output response becomes $(p_k)^{1/\gamma}$ with typically $\gamma \approx 2.2$, the consequence is simply that Equation 6.5 becomes scaled by the simple multiplicative constant of $1/\gamma$. Since this has no effect on the subsequent derivation, it will be dropped.

Arranging all three channels of the RGB image data as a vector, Equation 6.5 can be written

$$\log \mathbf{I} = \log \mathbf{E} + \log \mathbf{S}, \tag{6.6}$$

where **I** is the entire image, and **E** and **S** are the illumination and surface reflectance components of the image.

The reflectance image in log space, log **S,** can be further represented as a weighted linear combination of reflectance "features".

$$\log \mathbf{S} = \sum_{i=1}^{M} \mathbf{F}_i h_i \tag{6.7}$$

where the $\mathbf{F}_i$ are approximately independent reflectance features, and the $h_i$ are weighting coefficients. The variable *M* denotes the number of features used in modelling the reflectance component of the given image. Combining Equation 6.6 and Equation 6.7, the log image can be represented in terms of the illumination and *M* surface features as

$$\log \mathbf{I} = \log \mathbf{E} + \sum_{i=1}^{M} \mathbf{F}_i h_i \tag{6.8}$$

It is apparent that Equation 6.8 and Equation 6.1 are similar in structure, and in what follows we will exploit this similarity in separating an image into its illumination and reflectance components.

## 6.3 Separating Illumination and Reflectance Using NMFsc

Consider dividing the input image into subwindows, which can be overlapping or non-overlapping. Assuming the illumination is constant throughout the scene, then the contents of each subwindow can be modelled, according to Equation 6.8, in terms of an illumination component and a linear combination of reflectance features. However, what are the actual features making up the reflectance feature basis? This is where NMFsc is useful because from the collection of subwindows it provides a means of identifying the reflectance basis, along with the single illumination component.

Combining Equation 6.8 and Equation 6.1 yields the imaging model written in terms of an NMF approximation:

$$\mathbf{v}^t \approx \sum_{i=0}^{M} \mathbf{w}_i h_i = \mathbf{w}_0 h_0 + \sum_{i=1}^{M} \mathbf{w}_i h_i = \log \mathbf{E} + \sum_{i=1}^{M} \mathbf{F}_i h_i \qquad 6.9$$

In this case, $\mathbf{v}^t$ corresponds to $\log \mathbf{I}$ from Equation 6.8 and represents the data from an image subwindow. Since $\mathbf{w}_0 h_0$ takes the role of $\log \mathbf{E}$, the basis vector $\mathbf{w}_0$ is the "illumination" basis with weighting factor $h_0$. The symbol *M* represents the chosen dimensionality of the model. Since $\sum_{i=1}^{M} \mathbf{w}_i h_i$ takes the role of $\sum_{i=1}^{M} \mathbf{F}_i h_i$, the basis vectors $\mathbf{w}_i$ are the feature reflectance basis vectors with weighting factors $h_i$.

Let there be $T$ ($T \geq M+1$) sub-windows extracted from the input image. For each subwindow we append its log$R$, log$G$, log$B$ data into a column vector, and combine the columns of all the subwindows into the data matrix **V** of as in Equation 6.3. By applying NMFsc to solve for the basis matrix **W**, we obtain the illumination basis vector $\mathbf{w}_0$ and the feature reflectance basis vectors $\mathbf{w}_i$ ( $1 \leq i \leq M$ ). In other words, NMFsc decomposes the **V** into the illumination and reflectance components required for colour constancy and automatic white balancing.

Equation 6.8 is a purely additive model, and hence is appropriate for NMFsc since all basis vectors, including the feature reflectance images along with the illumination image, are required by the physics of light to be non-negative. Of course, the input data matrix **V** must be non-negative too. However, the logarithm of the original image data in [0, 255] could include both positive and negative values. The NMF input can be made non-negative by first scaling the original image data to [1/255, 1] to ensure that all pixel values in log space will be finitely negative or zero, and then replacing log with –log.

NMFsc allows the sparseness for each basis vector to be controlled individually. In my model, the illumination basis vector is forced to be non-sparse, making its components relatively similar, while the reflectance basis vectors are forced to be sparse. A minor modification of NMFsc allows the sparseness of each portion of a single basis vector to be controlled and evaluated separately. During each iteration of the original NMFsc method, sparseness is enforced on the entire basis vector. This modification is to enforce the same sparseness on

each third of the illumination basis separately. In other words, I evenly divided the illumination basis into three parts, apply the sparseness constraint to each part independently, then append them back into one vector. It is necessary to do this for the illumination vector because the R, G and B components of the illumination are all packed into a single vector in order to obtain a global solution across the three channels. However, when low sparseness is enforced uniformly across this vector, it leads to a solutions in which the three components are approximately equal, hence a prediction of a pure white illumination. To avoid this problem, the sparseness of the illumination basis vector needs is enforced separately for each of the R, G, and B segments of the vector. In addition, the sparseness for reflectance features is enforced on the rest basis vectors.

Hence, NMFsc is an approach for solving the illumination-reflectance model globally, in that the factorization aims to minimize the objective functions based on the data matrix that includes all three channels. This is an advantage over those methods that estimate the illumination and reflectance for each colour component independently.

The illumination-estimation algorithm based on using the NMFsc approach is summarized as follows:

1. Scale the input image values to be in the range [1/255,1]
2. Take *T* sample subwindows from the image
3. For each subwindow, concatenate the colour channels into a vector.
4. Take the negative of the logarithm of the elements of the vector from (3).

5. Choose *M* representing the total number of different reflectance features expected to appear in the *T* subwindows.
6. The true value of *M* is unknown in advance, and could differ from image to image. According to our test, we found that fixing *M* at 4 worked well.
7. Apply NMFsc to find *M*+1 basis vectors with the sparseness constraint of the $1^{st}$ basis vector set close to 0 so that it will represent the illumination(such constraint is enforced on three portions of this basis vector individually), and the sparseness constraints of the remaining basis vectors 2,..,(*M*+1) set significantly larger so they will represent the surface features.
8. Antilog the negative of the illumination basis
9. Separately average the R, G, B components of the illumination result to yield the estimate of the colour of the scene illumination.

Figure 6.2 and Figure 6.3 show one example of the result the above algorithm. More detailed test results will be described later.

In the above development, an image was assumed to contain *M* reflectance features. Data was collected from multiple subwindows to form the data matrix for NMFsc. However, instead of *M* reflectance features, suppose that we describe the scene as a single, more complex reflectance feature under a single illumination, and apply NMFsc to it. In this case, there is only one subwindow—the entire image—and there will be one illumination basis vector, and only a single reflectance basis vector.

For Equation 6.1 with $M = 1$ combined with Equation 6.8, the multiple-reflectance model can be simplified to be a single-reflectance model as

$$\mathbf{v}^t \approx \sum_{i=0}^{1} \mathbf{w}_i h_i = \mathbf{w}_0 h_0 + \mathbf{w}_1 h_1 = \log\mathbf{E} + \log\mathbf{S} \qquad 6.10$$

Here again, $\mathbf{w}_0 h_0 = \log\mathbf{E}$ so the basis vector $\mathbf{w}_0$ is the "illumination" basis with weighting factor $h_0$. Similarly, $\mathbf{w}_1 h_1 = \log\mathbf{S}$ so the basis vector $\mathbf{w}_1$ is the feature reflectance basis vector with weighting factor $h_1$. It is no longer necessary to estimate the parameter *M* because in this case it always equals 1.

Equation 6.10 expresses the idea that an input colour image can be separated into an illumination image and a reflectance image in log space. How NMFsc does the separation depends on the choice of sparseness constraints for the two components. Unfortunately, the optimal sparseness values are unknown for a given image. However, by setting a relatively low sparseness for one component, and a relatively high sparseness for the other, it is expected that the result of factorization to be a close representation of the true illumination (non-sparseness) and reflectance (sparse) components.

It is interesting to note that for the single-reflectance model the spatial location of the pixels becomes irrelevant. This is in contrast to the situation with the multiple-reflectance model where location does matter because pixel values are compared over multiple subwindows. Since the single-reflectance model does not rely on local spatial features, it can be applied to a set of un-ordered pixels, or even to a random subset of the input image pixels.

## 6.4 Tests of NMFsc

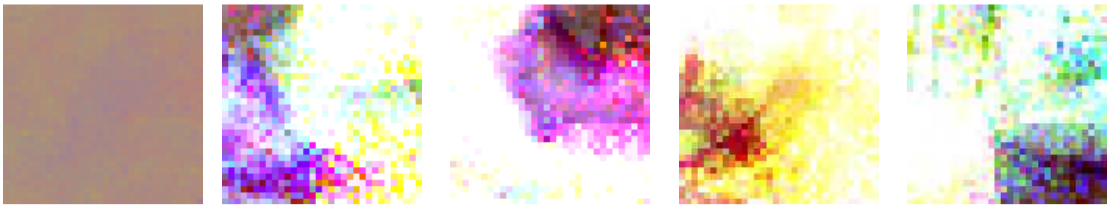
The first test is with the multiple-reflectance model. Figure 6.2 shows the result. This test is more illustrative of the method than a real test of its performance. Statistical results will be provided below. Figure 6.2 (a) is the original image of size 128x128. Figure 6.2 (b) shows the result of colour correcting the input image according to the actual illumination determined by measuring the colour of the gray reference card placed in the scene. Figure 6.2 (c) shows the colour correction based on NMFsc's illumination estimate using the multiple-reflectance model. The image in Figure 6.2 (a) was divided into 16 subwindows of size 32x32. NMFsc is applied with $M = 4$. Figure 6.3 shows the illumination and reflectance basis vectors from NMFsc. Figure 6.3 (a) is the illumination component (i.e., the antilog of the log$E$ term from Equation 6.9)), and is roughly similar across all subwindows, as it is expected to be. Figure 6.3 (b)-(e) are the feature reflectance basis vectors (i.e., the anti-log of the $w$'s in Equation 6.9).

The corresponding results for the single-reflectance model are shown in the top row of Figure 6.5. Once again the input image is Figure 6.2 (a). Figure 6.5(a) shows the colour-corrected image based on von-Kries scaling using the average colour of the illumination image (Figure 6.5 (b)) as the estimated scene illumination colour. Figure 6.5 (a) should be compared to the ground truth result shown in Figure 6.2 (b). Figure 6.5 (b) and Figure 6.5 (c) are the extracted illumination and reflectance images, respectively. The multiplication of Figure 6.5(b) and Figure 6.5 (c) pixel-by-pixel and channel-by-channel approximates

Figure 6.2 (a). For comparison, the corresponding results using McCann99 Retinex [33] are shown in the bottom row of Figure 6.5, where the R, G and B channels are processed independently for 5 iterations each. The purple and yellow pixels in Figure 6.5 (e) appear to relate to difficulties Retinex has in processing the almost-zero values in the dark shadows.

(a) (b) (c)

**Figure 6.2 The colour corrected image based on the multiple-feature reflectance model. (a) 128x128 input image; (b) Colour correction result based on the measured illumination; (c) Colour correction result based on illumination estimated using the NMFsc multiple-reflectance model.**



(a) (b) (c) (d) (e)

**Figure 6.3 The illumination and reflectance basis vectors (contrast enhanced for visualization) for the input image of Figure 6.4 (a) obtained via the multiple-feature reflectance model. (a) Illumination basis vector; (b)-(e) the reflectance feature vectors $w_i$.**

The second set of tests provides statistical results about the accuracy of NMFsc-based illumination estimation in comparison to some representative existing methods. Results are reported on three different image databases, and using full images versus thumbnails images as input (See *Appendix B*). The

single-reflectance-feature approach is also tested on a set of randomly drawn pixels. The first dataset is the Ciurea et al. [13] collection of 11,346 natural images representing a variety of indoor and outdoor scenes under different lighting conditions. The second test dataset is Cardei's [10] collection of 900 images of indoor and outdoor scenes. The third data set is Barnard's [5] 321 indoor images under 11 different illuminants. The images in the Barnard data set are linear, whereas, those in the Ciurea and Cardei sets have camera gamma applied. To make the tests consistent, we applied a gamma of 2.2 to the Barnard images.

For all three data sets, the algorithms tested include Grayworld, Max RGB, Shades of Gray (SoG) [26] with Minkowski norm 6, and Edge-Based [88], as well as both our multiple-reflectance and single-reflectance NMFsc methods. The Edge-Based method is included as representative of the state of the art since as the authors claim about their method, "…the newly proposed simple colour constancy algorithms obtain similar results as more complex state-of-the-art colour constancy methods." A comparison of the errors for all methods is given in Table 6.1, Table 6.2 and Table 6.3.

In the case of the multiple-reflectance-based estimation, each image is resized to be 64x64 pixels, and divided into sixteen 16x16 subwindows. The number of reflectance features $M$ is set to 4. The sparseness of the illumination and the reflectance bases are set to 0.01 and 0.25 respectively. The average computation time for processing one image is 0.8 seconds using Matlab on a 1.6 gigahertz computer. In the case of the single-reflectance-based estimation, each

image is also resized to 64x64. The sparseness of the illumination and the reflectance bases are 0.01 and 0.25 respectively. The average computational time for processing one 64x64 image is 2.4 seconds.

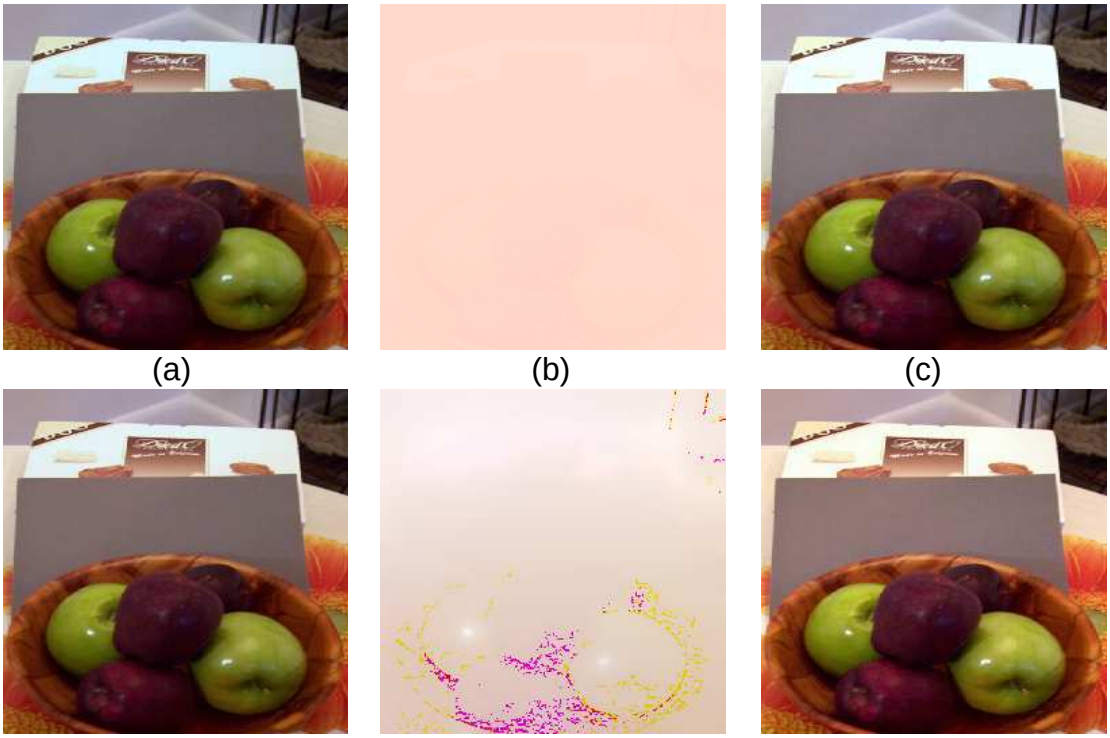



**Figure 6.5 Results using single-reflectance model on the input image from Figure 2(a). Top row NMFsc results: (a) colour corrected, (b) illumination component capturing the reddish colour cast in 2(a), (c) reflectance component. Bottom row Retinex results: (d) colour corrected, (e) illumination, (f) reflectance.**

The NMFsc illumination can also be applied to image thumbnails with comparable results. In particular, 8x8 thumbnails generated by down-sampling the full-sized input images are sufficient. As can be seen from Table 6.1 and Table 6.2, the illumination-estimation accuracy for the full-sized versus thumbnail images using the single-reflectance model is quite similar. Although not shown in the tables, similar accuracy is achieved on thumbnail images using the multiple-reflectance model too. One advantage of the 8x8 thumbnails is that they require

substantially less computation and storage. For the thumbnails, it takes on average 0.20 seconds (instead of 2.4) to process an image using the single-reflectance method.

| Method | Angular Degrees | | | Distance(x$10^2$) | | |
|---|---|---|---|---|---|---|
| | Median | RMS | Max | Median | RMS | Max |
| GW | 6.07 | 8.14 | 39.98 | 4.44 | 6.39 | 36.89 |
| SoG(p=6) | 4.06 | 6.16 | 30.37 | 3.05 | 4.63 | 28.57 |
| MAX RGB | 4.61 | 8.05 | 27.41 | 3.33 | 5.96 | 21.62 |
| Edge-Based | 4.27 | 7.89 | 30.67 | 3.22 | 5.81 | 24.10 |
| Multi-NMFsc | 5.41 | 7.80 | 31.80 | 3.76 | 5.68 | 24.34 |
| Single-NMFsc | 4.50 | 6.68 | 32.13 | 3.25 | 4.93 | 27.40 |
| Single-NMFsc (8x8 thumbnails) | 4.62 | 6.73 | 32.62 | 3.30 | 4.95 | 27.10 |
| Single-NMFsc (1% r.n. samples) | 5.01 | 7.41 | 37.44 | 3.66 | 5.51 | 36.05 |

**Table 6.1 Comparison of the NMFsc methods to the SoG, Max RGB, Grayworld and Edge-based methods based on the Ciurea [13] dataset of 11,346 images. Errors are reported in terms of both the RMS angular and distance error measures.**

| Method | Angular Degrees | | | Distance(x$10^2$) | | |
|---|---|---|---|---|---|---|
| | Median | RMS | Max | Median | RMS | Max |
| GW | 4.50 | 6.79 | 30.06 | 3.29 | 5.38 | 28.87 |
| SoG(p=6) | 2.96 | 4.85 | 17.95 | 2.20 | 3.71 | 15.33 |
| MAX RGB | 2.83 | 5.51 | 25.37 | 2.08 | 4.07 | 17.53 |
| Edge-Based | 3.16 | 5.41 | 29.90 | 2.37 | 4.11 | 26.50 |
| Multi-NMFsc | 3.95 | 6.27 | 24.40 | 2.85 | 4.63 | 19.73 |
| Single-NMFsc | 3.43 | 5.44 | 22.86 | 2.59 | 4.10 | 19.45 |
| Single-NMFsc (8x8 thumbnails) | 3.54 | 5.48 | 24.03 | 2.55 | 4.13 | 19.26 |
| Single-NMFsc (1% r.n. samples) | 3.48 | 5.46 | 22.57 | 2.51 | 4.08 | 19.53 |

**Table 6.2. Comparison of the NMFsc methods to the SoG, Max RGB, Grayworld and Edge-based methods based on the Cardei [10] dataset of 900 images. Errors are reported in terms of both the RMS angular and distance error measures.**

The sparseness measure is insensitive to the spatial location of pixels in the input, so in the case of the single-reflectance model, NMFsc can be used on a random subset of the pixels. I have tested the single-reflectance NMFsc method using as input a random sampling of only 1% of the pixels from the

original input image without a significant change in performance. The errors are reported in the last rows of Table 6.1 and Table 6.2.

| Method | Angular Degrees | | | Distance(x10$^2$) | | |
|---|---|---|---|---|---|---|
| | Median | RMS | Max | Median | RMS | Max |
| GW | 5.07 | 7.89 | 22.60 | 4.08 | 7.06 | 21.18 |
| SoG(p=6) | 2.57 | 4.42 | 13.03 | 1.99 | 2.96 | 8.69 |
| MAX RGB | 2.25 | 4.61 | 13.11 | 1.78 | 3.68 | 10.16 |
| Edge-Based | 4.33 | 5.36 | 12.87 | 3.37 | 4.38 | 10.93 |
| Multi-NMFsc | 5.46 | 7.20 | 17.57 | 3.96 | 5.24 | 12.84 |
| Single-NMFsc | 3.57 | 6.10 | 14.77 | 2.68 | 5.40 | 15.10 |
| Single-NMFsc (8x8 thumbnails) | 3.90 | 5.66 | 13.01 | 2.90 | 4.78 | 12.48 |
| Single-NMFsc (1% r.n. samples) | 4.37 | 5.86 | 15.03 | 3.20 | 4.47 | 11.47 |

**Table 6.3 Comparison of the NMFsc methods to the SoG, Max RGB, Grayworld and Edge-based methods based on the Barnard [5] dataset of 321 images. Errors are reported in terms of both the RMS angular and distance error measures.**

The parameter *M*, the number of reflectance features used in the model, is usually chosen to be a number smaller than 10. A larger number increases the computational cost, while a smaller number might not represent the correct number of features. I also explored how the sparseness settings affect the results. The assumption is that the illumination basis is quite uniform relative to the reflectance basis. What is important is not the absolute, but rather the relative values of the two sparseness settings. Therefore, the sparseness of the illumination is fixed at 0.01, varying only the sparseness of reflectance to find a good pairing. The relationship of the sparseness of reflectance and the average median angular errors are plotted in Figure 6.6 for both the 11,346 and 900 image datasets. As is evident from this plot the average median angular error increases with higher reflectance sparseness, but remains relatively stable below 0.5. As a result, we set the sparseness of reflectance to be 0.25 for all tests.

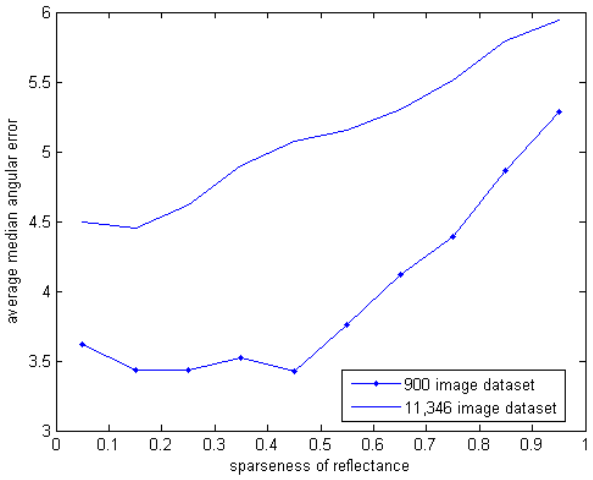



**Figure 6.6 The average median angular errors of both 11,346 and 900 image datasets plotted as a function of the sparseness of the reflectance component for the single-reflectance NMFsc on full size images. The sparseness of the illumination component is held fixed at 0.01.**

## 6.5 Conclusion

The problem of separating an image into its illumination and reflectance components was expressed in terms non-negative matrix factorization. The advantage of a non-negative factorization over other possible factorizations lies in the fact that reflectance and illumination are physically constrained to be non-negative. Since NMF factors a matrix into additive components, the matrix to factor is created from the negative of the logarithm of the input image data in (0, 1] is used. Sparseness constraints are imposed on the factorization so that it finds a component with little variation in its values, and a second component with significant variation. The sparseness constraints encode the assumptions that the scene illumination is roughly constant throughout the image, and that the surface reflectance is not constant.

Tests of the NMFsc approach to illumination estimation show that its performance is comparable to that of other methods. In particular, it is roughly comparable to the recent edge-based method [88], which is one of the most accurate illumination-estimation methods to date. Although it does not improve upon the state-of-the art in terms of accuracy, the NMFsc method is interesting in that it approaches the problem from an entirely different mathematical perspective and exploits a slightly different set of assumptions. In particular, the low sparseness constraint on the illumination does not directly imply spatially smooth illumination since the ordering of the data is irrelevant to the factorization. Similarly, there is no explicit assumption that either the average scene reflectance or its derivative be gray, nor that the maximal values of R, G and B represent the illumination colour in any way. Like Retinex, but unlike most other methods, NMFsc also provides a pixel-by-pixel estimate of the illumination colour.

# 7: ILLUMINATION ESTIMATION VIA THIN-PLATE SPLINE INTERPOLATION

In this chapter, we introduce a new approach to illumination estimation for colour constancy and automatic white balancing developed based on the technique of thin-plate spline interpolation. Different from the previous colour constancy models, we treat this estimation as a problem of interpolation over a set of training images. The illumination is described in terms of its chromaticity components $r$ and $g$, which can be viewed as a functions of the image $\mathbf{I}$; namely, $r = f_r(\mathbf{I})$ and $g = f_g(\mathbf{I})$. The problem of illumination estimation becomes that of estimating these two functions. In this case, TPS is used to map image information to the r-chromaticity and g-chromaticity values of the illumination.

## 7.1 Thin Plate Spline Method

Interpolation is a common problem and there are many well-established interpolation methods [48]. The majority of these methods, such as bilinear or bi-cubic interpolation, are based on interpolation over training data sampled on a uniform grid. However, we cannot uniformly sample the space of images, so interpolation over a non-uniformly sampled space is required. Thin-plate spline interpolation is an effective interpolation method under these conditions, and has been widely used in the context of deforming one image into registration with another.

As is typical of interpolation methods in general, thin-plate spline (TPS) interpolation constructs a function that matches a given set of data values $y_i$, corresponding to a given set of data vectors $\overline{X_i} = [X_{i,1}, X_{i,2}, ...X_{i,D}]$, in the sense that $y_i = f(\overline{X_i})$. Compared with other methods, TPS has been found to be quite stable and accurate in terms of finding a unique solution without having to tune a lot of parameters. Here, we apply it to the problem of estimating the chromaticity of a scene's overall incident illumination from an image of that scene. Many previous methods [10][34] have used a colour histogram as the input data; instead, for TPS we use image thumbnails as input. The thumbnails are 8x8 images created by averaging the underlying pixels in the original input image. These thumbnails in chromaticity coordinates become input vectors of size 8 x 8 x 2 = 128.

TPS for illumination estimation requires a "training" set of $N$ images along with their corresponding illumination chromaticity values $\{(\mathbf{I}_{i,1}, \mathbf{I}_{i,2}, \ldots \mathbf{I}_{i,128}), (r_i, g_i)\}$. TPS determines parameters $w_i$ and $a_j$ controlling the two mapping functions $f_r$, $f_g$, such that $(r_i, g_i) = (f_r(\mathbf{I}_{i,1}, \mathbf{I}_{i,2}, \ldots \mathbf{I}_{i,128}), f_g(\mathbf{I}_{i,1}, \mathbf{I}_{i,2}, \ldots \mathbf{I}_{i,128}))$.

According to TPS, the mapping function $f_r$, is defined as

$$f_r(I_1', I_2', ...I_{128}') = \sum_{i=1}^{N} w_i U(\| (I_1', I_2', ...I_{128}') - (I_{i,1}, I_{i,}, ...I_{i,128}) \|) + a_0 + \sum_{j=1}^{128} a_j I'_j \qquad 7.1$$

$$where \quad U(x) = x^2 \log x$$

The function $f_g$ is defined similarly. The weights $w_i$ control a non-linear term, and the $a_j$, control an additional linear term. Each training set pair (an image

plus its illumination chromaticity) provides 2 equations. For the $i^{th}$ training image we have

$$\begin{cases} r_i = f_r(I_{i,1}, I_{i,2}, ...I_{i,128}) \\ g_i = f_g(I_{i,1}, I_{i,2}, ...I_{i,128}) \end{cases} \quad 7.2$$

In addition, a smoothness constraint is imposed by minimizing the bending energy. In the original TPS formulation [48], the bending energy function was defined in 2D. Here we generalize it to higher dimensions defined as $\alpha_i$ ($i$= 1..128), the energy will be minimized when

$$\begin{aligned} &\sum w_i = 0 \\ &\sum I_{i,1} w_i = 0 \\ &\sum I_{i,2} w_i = 0 \\ &\quad .... \\ &\sum I_{i,128} w_i = 0 \end{aligned} \quad 7.3$$

For each of $f_r$ and $f_g$, we have ($N$+129) equations in ($N$+129) linear unknowns. Hence, the TPS parameters can be uniquely determined using matrix operations. If we define **L**, ***w***, ***k***, **Q** and $\overline{U}$ as follows:

$\mathbf{W} = [w_1, w_2, \ldots, w_N, a_0, a_1, a_2, \ldots, a_{128}]^T$

$\mathbf{K} = [r_1, r_2, \ldots, r_N, 0, 0, 0, \ldots, 0]^T$ or $[g_1, g_2, \ldots, g_N, 0, 0, 0, \ldots, 0]^T$

$$L = \begin{bmatrix} \overline{U} & Q \\ Q^T & 0 \end{bmatrix}, \quad where \quad Q = \begin{bmatrix} 1\ I_{1,1}\ I_{1,2} \ldots I_{1,128} \\ 1\ I_{2,1}\ I_{2,2} \ldots I_{2,128} \\ \ldots \\ 1\ I_{N,128}\ I_{N,128}\ \ldots I_{N,128} \end{bmatrix} \quad and \quad \overline{U} = \begin{bmatrix} 0 & U_{1,2} \ldots\ldots\ldots U_{1,N} \\ U_{2,1} & 0 \ldots\ldots\ldots\ldots U_{2,N} \\ .. & \\ U_{N,1} \ldots\ldots\ldots\ldots\ldots\ldots 0 \end{bmatrix}$$

where $\mathbf{U}_{ij} = \mathbf{U}(||(\mathbf{I}_{i,1} \ldots \mathbf{I}_{i,128}) - (\mathbf{I}_{j,1} \ldots \mathbf{I}_{j,128})||)$ and **0** is the 129x129 matrix of zeroes. These ($N$+129) equations system can then be written as **K**=**LW**, which can be solved as $\mathbf{W} = \mathbf{L}^{-1}\mathbf{K}$, to get the control parameters for functions $f_r$ and $f_g$ individually.

## 7.2 Tests of TPS

We have implemented the TPS illumination-estimation method in Matlab and conducted tests to compare its performance to that of other illumination-estimations methods.

Several different error measures are used to evaluate performance (See *Appendix A* for detail). For each image, the distance between the measured actual illumination chromaticity and that estimated is calculated as both the angular distance and the $L_2$-Euclidean distance. As with the distance measure, we also compute the RMS, mean, and median angular error over the test set of images.

The first test is from Barnard's calibrated 321 SONY images [5]. We evaluate the illumination error using the leave-one-out cross-validation procedure

[3]. In the leave-one-out procedure, one image is selected for testing, and the remaining 320 images are used for training in order to determine the required parameters. This is repeated 321 times, with a different image left out each time. The results with corresponding results for the Shades of Gray [26], Support Vector Regression(SVR)[26], Max RGB[22], and Grayworld [9] methods are listed in the Table 7.1.

We next consider Cardei's [10] set of 900 uncalibrated images(See *Appendix B*). Leave-one-out tests are used once again. The results are shown in Table 7.2.

The final test is based on the 7661 real images extracted from over 2 hours of digital video acquired with a SONY VX-2000 as used by Ciurea et al.[13]. We used Subset A for training and B for testing and vice versa. The results are listed in Table 7.3. The combined errors from both tests are shown in Table 7.4.

| **Method** | **Dimension/ Norm Power** | **Angle Distance** | | | **$L_2$-Distance(x$10^2$)** | | |
|---|---|---|---|---|---|---|---|
| | | **Median** | **Max** | **RMS** | **Median** | **RMS** | **Max** |
| TPS | | 0.64 | 14.43 | 2.10 | 0.53 | 1.55 | 10.42 |
| SVR | 2D | 4.65 | 22.99 | 10.06 | 3.41 | 7.5 | 16.41 |
| | 3D | 2.17 | 24.66 | 8.069 | 3.07 | 6.3 | 16.03 |
| SoG | 6 | 3.97 | 28.70 | 9.027 | 2.83 | 6.21 | 19.77 |
| Max RGB | | 6.44 | 36.24 | 12.28 | 4.46 | 8.25 | 25.01 |
| GW | | 7.04 | 37.31 | 13.58 | 5.68 | 11.12 | 35.38 |

**Table 7.1 Comparison of TPS to 2D and 3D SVR performance, SoG, Max RGB, Grayworld performance. The results involve real-data training and testing on the 321 SONY images. Errors are based on leave-one-out cross-validation, and are reported in terms of both the RMS angular chromaticity and distance error measures.**

| Method | Dimension/ Norm | Angle Distance | | | L2-Distance(x$10^2$) | | | |
|---|---|---|---|---|---|---|---|---|
| | | Median | RMS | Max | Median | Mean | RMS | Max |
| TPS(rg) | 2 | 2.26 | 3.86 | 22.23 | 1.72 | 2.22 | 2.92 | 18.29 |
| SVR (no resampling) | 2D | 2.40 | 4.47 | 20.43 | 1.74 | 2.40 | 3.27 | 18.40 |
| | 3D | 2.02 | 3.94 | 17.46 | 1.40 | 2.09 | 2.94 | 15.42 |
| SVR(with resampling) | 3D | 2.07 | 3.91 | 10.57 | 1.55 | 2.03 | 2.72 | 6.42 |
| C-by-C | 2D | - | - | - | - | 2.92 | 3.89 | - |
| NN | 2D | - | - | - | - | 2.26 | 2.76 | - |
| SoG | 6 | 3.02 | 4.99 | 19.71 | 2.19 | 2.96 | 3.80 | 15.96 |
| Max RGB | | 2.96 | 6.39 | 27.16 | 2.17 | 3.36 | 4.75 | 22.79 |
| GW | | 4.34 | 6.65 | 31.44 | 3.17 | 4.12 | 5.26 | 29.99 |

**Table 7.2 Comparison of Composition Solution and TPS to that of SVR, Colour by Correlation, the Neural Network, SoG, Max RGB, Grayworld. The tests are based on leave-one-out cross validation on a database of 900 uncalibrated images. The entries for C-by-C and NN are from [10] (Table 7 page 2385).**

| Method | Training and Test Sets | Angular Degrees | | | Distance(x$10^2$) | | |
|---|---|---|---|---|---|---|---|
| | | Median | RMS | Max | Median | RMS | Max |
| TPS | Train: Subset A<br>Test: Subset B | 4.52 | 7.02 | 34.81 | 3.37 | 5.19 | 25.78 |
| 3D SVR | | 4.53 | 6.76 | 24.55 | 4.11 | 5.03 | 18.62 |
| SoG (norm = 6) | | 6.71 | 8.93 | 37.01 | 4.83 | 6.59 | 27.99 |
| MAX RGB | | 10.33 | 12.81 | 27.42 | 6.99 | 9.14 | 21.72 |
| GW | | 6.83 | 9.66 | 43.84 | 5.25 | 7.82 | 45.09 |
| TPS | Train: Subset B<br>Test: Subset A | 4.58 | 6.83 | 27.62 | 3.31 | 4.99 | 29.37 |
| 3D SVR | | 5.33 | 7.32 | 24.80 | 3.91 | 5.29 | 16.68 |
| SoG (norm = 6) | | 6.71 | 8.92 | 37.01 | 4.83 | 6.59 | 27.99 |
| MAX RGB | | 9.23 | 11.32 | 26.76 | 6.76 | 8.39 | 21.55 |
| GW | | 7.83 | 10.66 | 43.84 | 6.25 | 8.81 | 45.09 |

**Table 7.3 Comparison of TPS error to 3D SVR, SoG, Max RGB, and Grayworld. Training is based on all the images in the given subset.**

| Method | Angular Degrees | | | Distance(x$10^2$) | | |
|---|---|---|---|---|---|---|
| | Median | RMS | Max | Median | RMS | Max |
| TPS | 4.56 | 6.93 | 34.18 | 3.35 | 5.09 | 25.78 |
| 3D SVR | 4.91 | 7.03 | 24.80 | 3.62 | 5.16 | 18.62 |
| SoG | 6.71 | 8.93 | 37.01 | 4.83 | 6.59 | 27.99 |
| MAX RGB | 9.65 | 12.13 | 27.42 | 6.86 | 8.80 | 21.72 |
| GW | 6.82 | 9.66 | 43.84 | 5.25 | 7.82 | 45.09 |

**Table 7.4 Comparison of TPS error to 3D SVR, SoG, Max RGB, and Grayworld. The results involve real-data training and testing on disjoint sets of 7,661 images from the Ciurea data set.**

## 7.3 Conclusion

The problem of estimating the chromaticity of the overall scene illumination is formulated in terms of interpolation over a non-uniformly sampled data set. The chromaticity is viewed as a function of the image and the set of training images is non-uniformly spaced. Thin-plate-spline interpolation is an excellent interpolation technique for these conditions and has been shown to work well for illumination estimation in particular. TPS calculates its result based on a weighted combination of the entire set of training data. Hence, for efficiency it is important to keep that set as small as possible. Overall, the tests on real images show the accuracy of TPS illumination estimation to be very good.

# 8: QUATERNION COLOUR CURVATURE

Edges have been used in colour constancy algorithms such as the Edge-based Hypothesis approach reviewed in Section 2.1.5. For better detection of edge and curvature structures in colour images, in this Chapter, I propose a novel approach to measuring curvature for in colour or vector-valued images based on quaternion singular value decomposition of a Hessian matrix. This approach generalizes the existing scalar-image curvature approach which makes use of the eigenvalues of the Hessian matrix [29]. In the case of vector-valued images, the Hessian is no longer a 2D matrix but rather a rank 3 tensor. Here, quaternion curvature is used to derive vesselness measure for tubular structures in colour or vector-valued images by extending Frangi's [29] vesselness measure for scalar images. Test results show the effectiveness of quaternion colour curvature in generating a vesselness map.

## 8.1 Introduction

Hessian-based methods have been widely used from curvature measures to feature detection[6][12][14][16][18][29][63][64][66]. The Hessian matrix describes the second-order structure of gray-level variations around each pixel of the image. There are two main categories where a Hessian matrix is used. First, the Hessian and the related second-moment matrix have been applied in several operators (e.g., the Harris [36], Harris-affine [65], and Hessian-affine [64] detectors) to find "interest" points where the local image geometry changes in

more than one direction. Hessian-based blob detector in colour space is proposed in [66]. Second, since the eigenvalues of the Hessian matrix at a pixel measure the principal curvatures of the image intensity surface, it can be used to detect tubular (linear, vessel-like) structures, which is useful in many applications [12][14][15][16][18][29]. By smoothing with Gaussian kernels of various sizes, the normalized second-order derivatives indicate the scale and orientation of vessels. *Vesselness* is measured by a large curvature in the cross-sectional direction and a small curvature along the vessel. By eigen-analysis of the Hessian matrix, elongated objects (i.e., vessels) are detected wherever the first eigenvalue is positive (or negative) and prominent. This process generates a single response for both lines and edges, producing a clearer sketch of an image's structure than is usually provided by the magnitude of gradient.

Existing first-derivative point/blob detectors are applied to gray scale images. In the case of colour images, the basic approach has been to compute the derivatives of each colour channel separately, and then add them to produce the final result [66]. However, the first derivatives of a colour edge can be in opposing directions, so the summation can lead to cancellation of the derivatives. The same situation occurs in second-derivative-based Hessian detectors. Existing Hessian-based curvature methods are also based on gray scale images, whether the luminance image, or a single colour channel. For example, Hessian-based multi-scale segmentation or enhancement of vessels in retinal images has been extensively studied [12][15][18][29], where only the green channel is used.

To make use of the extra information in a colour image, I use the quaternion representation of colour to extend Hessian curvature measures to the colour domain. In particular, Frangi's [29] vesselness approach is extended by estimating principle curvatures in RGB colour space using quaternion operations. Sanqwine [8] introduced the quaternion representation of colour. Since quaternions, which are an extension of the complex numbers, consist of one real component and three imaginary components, a colour can be represented by a pure quaternion having a real component of zero, and imaginary components R, G and B. With colours encoded in quaternions, the entries of the Hessian matrix become quaternions that combine secondary derivatives from all colour channels in their imaginary components. Quaternion singular value decomposition (QSVD) [8][69] can then be applied to the Hessian matrix in order to find the principle curvatures as described by the two non-negative, real-valued singular values. These singular values can be used to measure vesselness or other features.

## 8.2 Curvature and Vesselness Measure

Viewing an image as an intensity surface, the local shape characteristics of the surface at a particular point can be described by the Hessian matrix. Lines (i.e., straight or nearly straight curvilinear features) and edges have high curvature in one direction and low curvature in the orthogonal direction, and this characteristic can be measured via the Hessian, **H**. For a 2D scalar image, **H** is a 2x2 matrix of the second derivatives of image **I**

$$H(\sigma) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \quad 8.1$$

The four entries of **H** are the second-order partial derivatives of the scalar image **I** evaluated at the pixel $\boldsymbol{p} = \langle x, y \rangle$, and $\sigma$ is the Gaussian scale of the partial derivatives.

The eigenvalues of **H** are called principal curvatures and are invariant under rotation. The eigenvectors of **H** can be used to define a coordinate system that is aligned with the dominant directions of curvature. Given the ordered eigenvalues of **H** such that $|\lambda_1| < |\lambda_2|$ with corresponding eigenvectors ($\boldsymbol{e}_1$, $\boldsymbol{e}_2$), the eigenvectors define an orthogonal coordinate system aligned with the direction of minimal $\boldsymbol{e}_1$ and maximal $\boldsymbol{e}_2$ curvature.

In the case of a vessel-like structure, $\boldsymbol{e}_1$ indicates the orientation of the vessel. Thus $\boldsymbol{e}_1$ represents the parallel curvature, and $\boldsymbol{e}_2$ the orthogonal curvature. As a vesselness measure for 2D images, Frangi [29] uses **H** to describe the curvature at each point in the image. The idea behind eigenvalue analysis of the Hessian is to extract the principal directions in which the local second-order structure of the image can be decomposed. Since this directly gives the direction of least curvature (along the vessel), application of several filters in multiple orientations is avoided.

Both eigenvalues play an important role in the vesselness measure. In particular, for a vessel we expect $|\lambda_1| < |\lambda_2|$, with $\lambda_2 < 0$ for bright vessels against a

dark background, and $\lambda_2$ >0 for the reverse. Finally the overall magnitude of the eigenvalues should be larger at vessels than in background regions. The Frangi filter combines these observations in the following two quantities

$$R_B = \frac{|\lambda_1|}{|\lambda_2|} \tag{8.2}$$

$$S = \| H_\sigma \| = \sqrt{\lambda_1^2 + \lambda_2^2} \tag{8.3}$$

Here, $R_B$ is the *blobness* measure in 2D. It is maximized for highly blob-like structures and decreases as the difference between the parallel and orthogonal curvature increases. $S$ is the norm of the Hessian matrix and measures the relative brightness/darkness of the structure. It should become large for vessels. In other words, it presents the “unlikelihood” that a pixel is from the background. These quantities are combined using exponentiation yielding a ‘‘vesselness’’ measure (for the bright-vessels-on-dark case) defined as follows:

$$V(\sigma) = \begin{cases} 0 & if\ \lambda_2 > 0, \\ e^{\frac{-R_B^2}{2\beta^2}} (1 - e^{\frac{-S^2}{2c^2}}) \end{cases} \tag{8.4}$$

The constants $\beta$ and $c$ are parameters which control the sensitivity of the filter to *blobness* and *backgroundness*.

## 8.3 Eigenvalues of the Colour Hessian Matrix

As mentioned above, when the gradient of a colour image is computed by adding up the first derivatives of the separate channels, the channel derivatives may point in opposing directions and cancel one another. DiZenzo [17] argues that a simple summation of the derivatives ignores the correlation between the channels. A similar problem arises in converting a colour image to a luminance image. As a solution in the first-derivative case, DiZenzo[17] and Kass[46] proposed the colour tensor by colour gradient, but it does not generalize to the colour Hessian matrix. The alternative of solving for the eigenvalues of the Hessian matrix separately in each colour channel generates three pairs of eigenvalues, but these then do not immediately fit into the schema of Frangi's vesselness measure. Ming [66] used a weighted combination of Hessian matrices over HSI colour channels to calculate a colour Hessian.; however, this approach does not eliminate the cancellation problem either. Our proposal, which uses the eigenvalues and eigenvectors of a colour Hessian matrix based on quaternion singular value decomposition [8][18], overcomes the cancellation problem.

In the quaternion representation of a 2D colour image, each pixel $\boldsymbol{p}$ = <*x*,*y*> is represented by a quaternion number $Q = I_1 \cdot i + I_2 \cdot j + I_3 \cdot k$ , where $I_n$ (with $n$ = 1,2,3) is the $n^{th}$ channel of the input image, and *i*, *j*, and *k* are three imaginary bases. The quaternion-valued Hessian matrix $\mathbf{H}_Q$ is constructed as

$$H_Q(\sigma) = \begin{bmatrix} \frac{\partial^2 I_1}{\partial x^2} & \frac{\partial^2 I_1}{\partial x \partial y} \\ \frac{\partial^2 I_1}{\partial y \partial x} & \frac{\partial^2 I_1}{\partial y^2} \end{bmatrix} \cdot i + \begin{bmatrix} \frac{\partial^2 I_2}{\partial x^2} & \frac{\partial^2 I_2}{\partial x \partial y} \\ \frac{\partial^2 I_2}{\partial y \partial x} & \frac{\partial^2 I_2}{\partial y^2} \end{bmatrix} \cdot j + \begin{bmatrix} \frac{\partial^2 I_3}{\partial x^2} & \frac{\partial^2 I_3}{\partial x \partial y} \\ \frac{\partial^2 I_3}{\partial y \partial x} & \frac{\partial^2 I_3}{\partial y^2} \end{bmatrix} \cdot k \qquad 8.5$$

QSVD [8][69] can be applied to $\mathbf{H}_Q$,

$$H_Q = V_Q^T \cdot \Lambda \cdot U_Q \text{,} \qquad 8.6$$

where $\mathbf{V}_Q$ and $\mathbf{U}_Q$ are two quaternion matrices of eigenvectors, and $\Lambda$ is a real-valued diagonal matrix containing two non-negative singular values $\xi_1$ and $\xi_2$. Given the assumption that quaternion eigenvector corresponding to the smaller singular value of the Hessian points along the direction of minimal curvature, and that the larger singular value points along the direction of the maximum curvature, we can continue using Equation 8.4, but now as a colour vesselness measure. It should be noted that the two singular values in Equation 8.6 are unsigned magnitudes. To apply the sign test in Equation 8.4, however, we must use the sign of eigenvalue $\lambda_2$ from the corresponding gray-scale image.

## 8.4 Tests and Results of Curvature Measure

The tests are based on a set of colour images consisting of photomicrographs [97], nature photos, and satellite imagery [95]. For each such image, a vessel map image is generated that can be used for detection and segmentation of tubular structures, and vessel segmentation and enhancement. The main purpose of the vessel map is to increase the separability of vessel

structures from the background. Segmentation can be obtained by thresholding the vessel map, and enhancement can be achieved by pixel-wise multiplication with the input image. Due to the variability in the scale of vessels, the vessel map is constructed using a multi-scale scheme. Five scales of Gaussian are employed for each image, with $\sigma$ = 1,2,3,4 and 5. Gamma-normalized derivatives are also used with $\gamma$ = 0.5 as in [29]. The *blobness* and *backgroundness* parameters $\beta$ and *c* are set to 0.5. The results are combined across the scales by the maximum rule [29], which is to use the maximum vesselness response across all scales.

Several results of tests on colour images are shown in Figure 8.1 - Figure 8.4. The vessel maps are generated based on both the colour-Hessian approach and the traditional grayscale-Hessian approach. For ease of comparison, the vessel maps are normalized by scaling vesselness intensity to [0,1] and then scaled for better visualization.

As can be seen by comparing Figure 8.1(b)-Figure 8.4(b) with Figure 8.1(a)-Figure 8.4(a), the colour Hessian achieves better results than the grayscale version in term of the vessel map. In the grayscale-derived vessel maps, there is low vesselness found for vessels that differ in colour from the background, but are nonetheless iso-luminant to it. However, even in the regions where luminance of the vessel and background differ, the results of the colour Hessian show higher vesselness contrast. Colour is an important discriminative property of objects, and the results demonstrate that it provides sufficient extra

information to distinguish between background and objects in cases where the traditional luminance-based method fails.

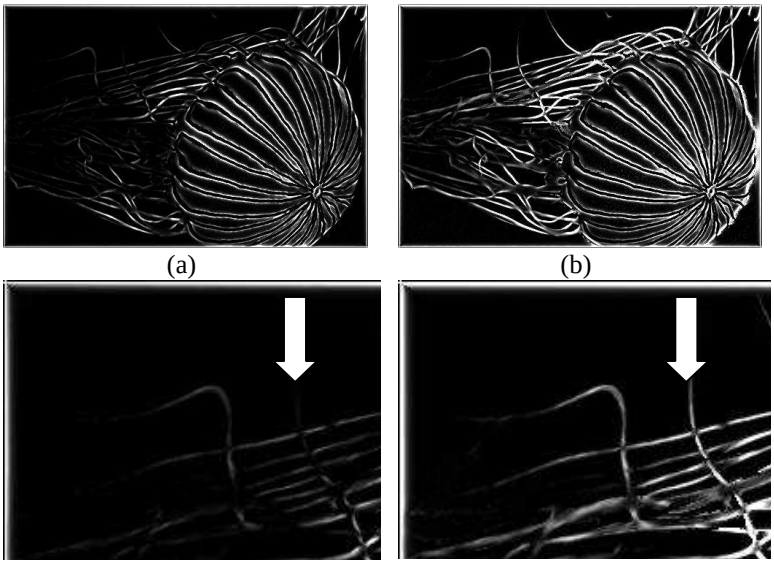



**Figure 8.1 Result based on a photo of a jellyfish [96] (a) Grayscale-based Hessian result in which the tentacles are not detected due to approximate iso-luminance. (b) Colour-based Hessian result in which the tentacles are more clearly delineated. (c)-(d) Scaled up version of the top-left corners of (a)-(b) respectively.**

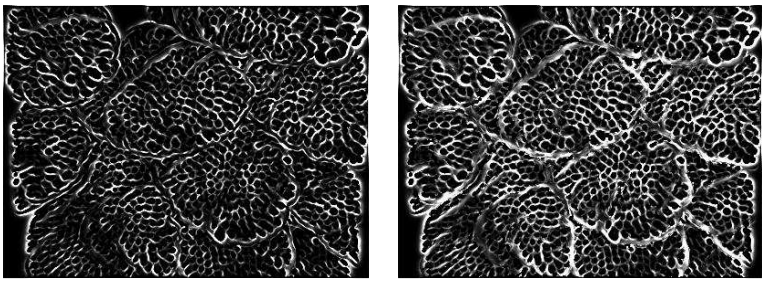
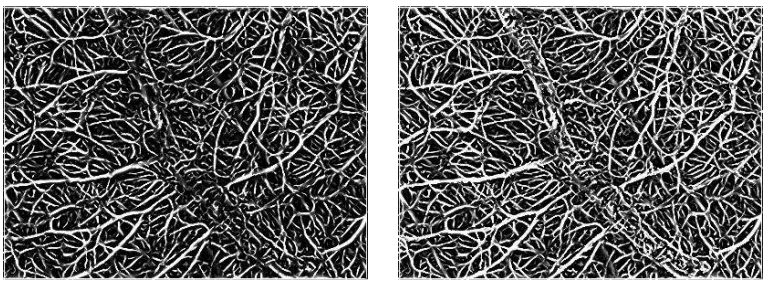
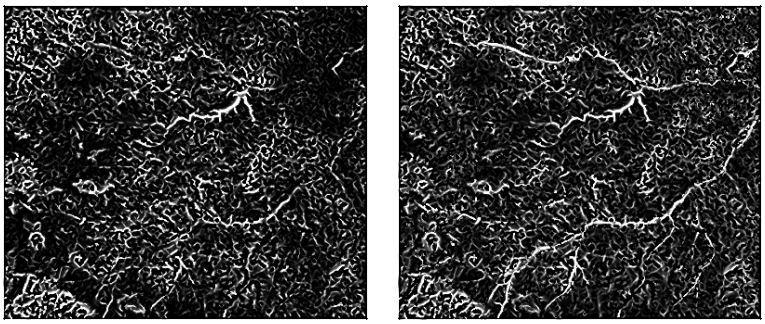



**Figure 8.2 Result based on a two-photon fluorescence microscopy image of villi of the mouse small intestine [97]; (a) Grayscale-based Hessian result in which the curvature measure around green contours is low because they are similar in intensity to the blue background; (b) Colour-based Hessian result in which the green tubular structures are clearly delineated.**



(a) (b)

**Figure 8.3 Result based on a fluorescence and confocal microscopy photo of rat retina astrocytes and blood vessels[97]; (a) Grayscale-based Hessian result where the method fails to detect the dominant vessel across the center line of the image. (b) Colour-based Hessian result.**



(a) (b)

**Figure 8.4 Result based on a satellite image[95]; (a) Grayscale-based Hessian result in which the green vessel-like structure is missed. (b) Colour-based Hessian result in which the green vessel-like structure is identified.**

## 8.5 Conclusion

The Hessian matrix can be used to estimate curvature and so provides a good foundation for identifying interesting image features such as tubular vessels and blobs. In generalizing the use of the Hessian from grayscale to colour images, however, the problem that arises is the possible information loss caused by cancellation of derivatives in opposing directions from the separate colour channels. To overcome this problem, I employ the quaternion representation of colour, which encodes an RGB colour in a single quaternion number. Information loss is avoided by extracting the eigenvalues from the quaternion-valued colour

Hessian matrix based via QSVD. The quaternion-based method demonstrates improved performance in term of the resulting vessel map, which is important for vessel segmentation and enhancement, as well as, for curvature/edge detection that can be potentially useful in improving the result of edge-based colour constancy algorithms.

# 9: OVERALL CONCLUSION AND CONTRUBITIONS

I have discussed several colour constancy methods in detail in the previous chapters. In this chapter, I will present a final review of these methods and enumerate my specific contributions.

I have explored non-statistical estimation solutions, constrained by the characteristics of surface material as well as the illumination. The following characteristics have been studied: surface colour of specific materials, such as human skin and achromatic surfaces; illumination spectra limited to blackbody radiation; the formation of colour signals according to the dichromatic model of reflection; and the independence of spatial variance of illumination and surfaces in a given scene. I also have discussed one supervised approach that uses prior information from a training set via thin-plate-spline interpolation. In summary:

In Chapter Three, a new colour constancy method, GSI, was proposed that is based on detecting pixels corresponding to gray surface reflectance—which is not necessarily the same as gray image colour—and using their average image colour as an indicator of the colour of the overall scene illumination. The gray surfaces are found by first transforming the image RGB values to the LIS coordinate system with axes which roughly correspond to luminance, illumination ‘colour’ and reflectance. The trick is in the identification of gray surfaces in a colour coordinate system that encodes illumination and surface reflectance along different axes, where values of S near zero tend to be gray. Tests on real images

show the GSI method works better than Shades-of-Gray, Grayworld and Max RGB. This method does not require training, and is substantially simpler to implement. This work has been published in CIC 2007 [93] as a result of the team effort of Weihua Xiong, Brian Funt and I.

Chapter Four discussed a model of skin colour under varying illumination that allows easy decomposition of a skin chromaticity into two components: one due to the illumination and the second due to the melanin content. As a computationally inexpensive model, the proposed method is useful for estimating the colour of the light illuminating the skin, and for normalizing skin in images for which the colour of the illumination is not known. Based on this model, the range of chromaticities of arbitrary skin under all illuminant colour temperatures can be transformed into a new coordinate system defined with two independent axes: an illumination axis as a function of colour temperature, and a skin axis as a function of melanin content. Tests show that the model succeeds relatively well even when the assumptions (i.e., narrowband sensors, blackbody illuminants, and melanin and hemoglobin contents as being the dominant factors in skin colour) are violated. The skin tone normalization is accomplished by shifting the colour of the entire image so that skin pixels lie on the pre-defined melanin axis. The corresponding result has been published in Shi and Funt, AIC 2008 [74].

Another robust method for determining the illumination was proposed in Chapter Five. The presented illumination estimation method uses the constraints provided by the dichromatic model based on two Hough transform voting procedures. First, each image pixel votes for every dichromatic plane it could fall

on. This results in a 2D histogram representing the likelihood of each plane. Second, each dichromatic plane votes for each candidate illumination axis that could pass through that plane. The final illumination estimate is determined by intelligently choosing from amongst the most likely candidates. This robust method creates a 2D illumination axis histogram that represents the likelihood of the possible illuminations. Unlike other existing methods trying to solve for Dichromatic model, this method makes no assumption about the number of surfaces or the surface colours, yet performs well in comparison to the other methods tested. The corresponding result has been published in Shi and Funt, CGIV 2008[75].

In Chapter Six, a new approach to illumination estimation for colour constancy and automatic white balancing, by separating the image into illumination and reflectance components, was proposed based on the technique of nonnegative matrix factorization (NMF) with sparseness constraints (NMFsc). The problem of separating an image into its illumination and reflectance components was expressed in terms non-negative matrix factorization. The advantage of a non-negative factorization over other possible factorizations lies in the fact that reflectance and illumination are physically constrained to be non-negative. Sparseness constraints are imposed on the factorization so that it finds a component with little variation in its values, and a second component with significant variation. The sparseness constraints encode the assumptions that the scene illumination is roughly constant throughout the image, and that the surface reflectance is not constant. Tests of the NMFsc approach to illumination

estimation show that its performance is comparable to that of other methods. Although it does not improve upon the state-of-the art in terms of accuracy, the NMFsc method is interesting in that it approaches the problem from an entirely different mathematical perspective and exploits a slightly different set of assumptions. Like Retinex, but unlike most other methods, NMFsc also provides a pixel-by-pixel estimate of the illumination colour. The corresponding result has been published in Shi et al. AIC 2007[77].

In Chapter Seven, I formulated the problem of estimating the chromaticity of the overall scene illumination in terms of interpolation over a non-uniformly sampled data set. The chromaticity is viewed as a function of the image and the set of training images is non-uniformly spaced. Thin-plate-spline interpolation is an excellent interpolation technique for these conditions and has been shown to work well for illumination estimation in particular. TPS calculates its result based on a weighted combination of the entire set of training data. Overall, tests of the thin-plate spline method on a large set of real images demonstrate that the method estimates the colour of the incident illumination quite accurately. This work has been published in CIC 2007 [94] as a result of the team effort of Weihua Xiong, myself and Brian Funt.

Finally, Chapter Eight provided a quaternion-based colour curvature measure as an aid to existing methods that use information from spatial edges. Here, the Hessian matrix is used to estimate curvature and so provides a good foundation for identifying interesting image features such as tubular vessels and blobs. I employ the quaternion representation of colour, which encodes an RGB

colour in a single quaternion number. Information loss during colour-to-gray conversion is avoided by extracting the eigenvalues from the quaternion-valued colour Hessian matrix via QSVD. The quaternion-based method demonstrates improved performance in terms of the resulting vessel map, which is important for vessel segmentation and enhancement. As well it is useful for curvature/edge detection that can be potentially helpful in improving the results of edge-based colour constancy algorithms. The corresponding result has been published in Shi et al. CIC 2008 [76].

In Table 9.1 - Table 9.3, I have summarized the performance of the existing and the proposed methods based on three datasets:

| **Methods** | | **Training Required** | **Median Angular Error** |
|---|---|---|---|
| existing | Grey World | no | 7.0 |
| | Max-RGB | | 6.5 |
| | Multilinear Constraint | | 5.8 |
| | Shades of Grey (n=6) | | 3.7 |
| | Grey Edge | | 3.2 |
| | 2nd order Grey Edge | | 2.7 |
| | Colour by Correlation | yes | 3.2 |
| | Neural Networks | | 7.8 |
| | 2D SVR | | 4.7 |
| | 3D SVR | | 2.2 |
| Chapter 3 | GSI | no | 3.9 |
| Chapter 5 | 3D Hough Transform | | 1.7 |
| Chapter 6 | Multi-NMFsc | | 5.5 |
| | Single-NMFsc | | 3.6 |
| Chapter 7 | TPS | yes | 0.6 |

**Table 9.1 Comparison of methods proposed in this thesis and existing methods in terms of errors based on 321 image dataset.**

| Method | | Train | Angle Distance | | | L2-Distance(x$10^2$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Median | RMS | Max | Median | Mean | RMS | Max |
| existing | SVR 2D | yes | 2.40 | 4.47 | 20.43 | 1.74 | 2.40 | 3.27 | 18.40 |
| | SVR 3D | | 2.02 | 3.94 | 17.46 | 1.40 | 2.09 | 2.94 | 15.42 |
| | C-by-C | | - | - | - | - | 2.92 | 3.89 | - |
| | NN | | - | - | - | - | 2.26 | 2.76 | - |
| | SoG(p=6) | no | 3.02 | 4.99 | 19.71 | 2.19 | 2.96 | 3.80 | 15.96 |
| | Max RGB | | 2.96 | 6.39 | 27.16 | 2.17 | 3.36 | 4.75 | 22.79 |
| | GW | | 4.34 | 6.65 | 31.44 | 3.17 | 4.12 | 5.26 | 29.99 |
| | Edge-Based | | 3.16 | 5.41 | 29.90 | 2.37 | 4.11 | 26.50 | 3.16 |
| Chapter 6 | Multi-NMFsc | | 3.95 | 6.27 | 24.40 | 2.85 | 4.63 | 19.73 | 3.95 |
| | Single-NMFsc | | 3.43 | 5.44 | 22.86 | 2.59 | 4.10 | 19.45 | 3.43 |
| Chapter 7 | TPS | yes | 2.26 | 3.86 | 22.23 | 1.72 | 2.22 | 2.92 | 18.29 |

**Table 9.2 Comparison of methods proposed in this thesis and existing methods in terms of errors based on 900 image dataset.**

| Method | | Train | Angular Degrees | | | Distance(x$10^2$) | | |
|---|---|---|---|---|---|---|---|---|
| | | | Median | RMS | Max | Median | RMS | Max |
| existing | GW | no | 6.07 | 8.14 | 39.98 | 4.44 | 6.39 | 36.89 |
| | SoG(p=6) | | 4.06 | 6.16 | 30.37 | 3.05 | 4.63 | 28.57 |
| | MAX RGB | | 4.61 | 8.05 | 27.41 | 3.33 | 5.96 | 21.62 |
| | Edge-Based | | 4.27 | 7.89 | 30.67 | 3.22 | 5.81 | 24.10 |
| Chapter 3 | GSI | | 5.46 | 7.95 | 38.71 | 4.15 | 6.23 | 31.93 |
| Chapter 6 | Multi-NMFsc | | 5.41 | 7.80 | 31.80 | 3.76 | 5.68 | 24.34 |
| | Single-NMFsc | | 4.50 | 6.68 | 32.13 | 3.25 | 4.93 | 27.40 |
| Chapter 7 | TPS | yes | 4.56 | 6.93 | 34.18 | 3.35 | 5.09 | 25.78 |

**Table 9.3 Comparison of methods proposed in this thesis and existing methods in terms of errors based on 7661 image dataset.**

Which method is to be preferred depends on the situation. For example, TPS may be the best option in terms of accuracy when training and memory are not concerns. As a non-training approach, GSI achieves good estimation accuracy provided the camera sensor responses of the testing images are known. The method based on solving the dichromatic model is a good choice when the testing images are taken under laboratory conditions and speed is not a concern. The NMFsc method can generate a separate illumination estimate for every pixel location, and therefore is preferred when the illumination colour in the scene is not constant. Finally, if pixels representing human skin can be detected

in the testing image, the skin-illumination model may be used for accurate illumination estimation as well as skin tone correction.

Overall, these novel methods proposed in this thesis aim to overcome drawbacks in existing approaches for better performance, increased robustness, and improved efficiency. The corresponding results of tests based on real world image data sets have proven their success in the field of Colour Constancy and Illumination Estimation.

# APPENDICES

## Appendix A – Error Measure

Given and illumination estimate [$R_e$, $G_e$, $B_e$], its corresponding chromaticity values are

$$r_e = \frac{R_e}{(R_e + G_e + B_e)}, g_e = \frac{G_e}{(R_e + G_e + B_e)}, b_e = \frac{B_e}{(R_e + G_e + B_e)}$$

Given illumination chromaticity $r$ and $g$, the other component can be obtained as $b = 1 - r - g$ . Let [$r_r$ $g_r$ $b_r$] be the true illumination chromaticity. The distance error in 2D chromaticity space and 3D angular space are defined as:

$$E_{i-dist} = \sqrt{(r_r - r_e)^2 + (g_r - g_e)^2}$$

$$E_{i-angular} = COS^{-1}\left[\frac{(r_r, g_r, b_r) \circ (r_e, g_e, b_e)}{\sqrt{r_r^2 + g_r^2 + b_r^2} \times \sqrt{r_e^2 + g_e^2 + b_e^2}}\right] \times \frac{2\pi}{360}$$

For test set of $N$ images, the RMS is defined in the standard way as:

$$RMS_{dist} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} E_{i\_dist}^2}$$

$$RMS_{angular} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} E_{i\_angular}^2}$$

## Appendix B – Main Datasets

### 321 Image Dataset (Barnard 2002)

Barnard's 321 images [5] of scenes in a laboratory setting are captured using a calibrated SONY DXC-930 camera. These images are from 33 different scenes under 11 different lights that represent a cross-section of common lights.

### 900 Image Dataset (Cardei 2002)

Cardei's [10] set of 900 uncalibrated images are taken using a variety of different digital cameras manufactured by Kodak, Olympus, HP, Fuji Polaroid, PDC, Canon, Ricoh and Toshiba. The illumination RGB values for these images were measured from a gray card placed in each scene.

### 11346 Image Dataset (Ciurea 2003)

Ciurea's dataset [13] includes a wide variety of indoor and outdoor scenes, including many with people in them. He measured the illumination via a grayball attached to a digital video camera. This was made to appear at a fixed location near the right-bottom corner of each video frame. The average chromaticity value of the pixels in the brightest region is assumed to reflect the RGB of the true scene illumination. The grayball appears in the lower right-hand quadrant of every original image, so for testing the bottom corner or the right half of each image should be cropped. As shown in the figure below, one test image is cropped to remove the gray ball, which is located at a fixed location in the lower right quadrant.


(a) (b)

**Figure 9.1 Example image from the dataset. (a) Original image containing gray ball from which the colour of the scene illumination is determined. (b) Cropped image to be used for algorithm testing with gray ball removed.**

### Reduced subset of 11346 Images

However, many of 11346 images have very good colour balance (i.e., RGB of the gray ball is gray) which could bias the testing of the illumination estimation methods. Therefore, the majority of the correctly balanced images are eliminated from the data set so that the overall distribution of the illumination colour is more uniform, as can be seen in the figure below. The resulting data set contains 7661 images.

The reduced database can be further divided into two independent sets based on geographical location: Subset A includes 3581 images, and subset B includes 4080. Subset A contains images from the Apache Trail, Burnaby Mountain, Camelback Mountain, CIC 2002 and Deer Lake. Subset B contains images from completely different locations including False Creek, Granville Island Market, Marine Drive, Metrotown shopping centre, Scottsdale, Simon Fraser University and Whiteclyff Park.

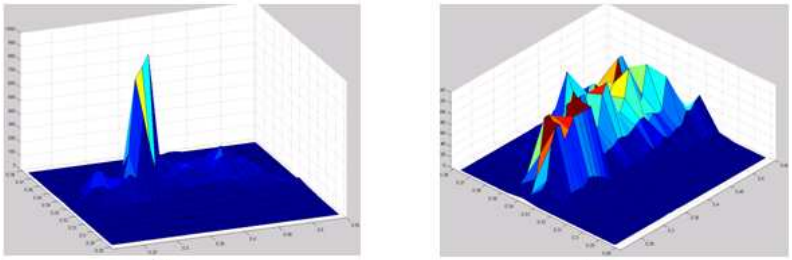



**Figure 9.2 Illumination distribution of full dataset and selected subset. (a) The original data set contains 11,346 images, but the illumination chromaticities cluster around gray (0.33, 0.33). (b) The reduced data set contains 7661 images with a more uniform distribution of illumination chromaticity.**

# REFEREFNCES


[1] E. Angelopoulou, R. Molana, and K. Daniilidis, “Multispectral skin colour modeling,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2, 2001, pp. 635-642.

[2] A.O. Barel, P. Clarys, K. Alewaeters, C. Duez, J. L. Hubinon, and M. Mommaerts, “The Visi-Chroma VC-100A: A New Imaging Colorimeter for Dermatocosmetic Research,” *Skin Res. Technol*, 7, 2001, pp. 24-31.

[3] K. Barnard, L. Martin, A. Coath, and B. Funt, “A Comparison of Computational Colour Constancy Algorithms. Part Two: Experiments on Image Data,” *IEEE Transactions on Image Processing*, 11, 2002, pp. 985-996.

[4] K. Barnard, L. Martin, and B. Funt, “Colour by correlation in a three dimensional colour space,” in *6th European Conference on Computer Vision*, Springer, 2000, pp. 375-389.

[5] K. Barnard, L. Martin, B. Funt, and A. Coath, “A data set for colour research,” *Color Research and Application*, *27(3)*, 2002, pp. 147-151.

[6] H. Bay, T. H. Tuytelaars, and L. J. Van Gool, “SURF: Speeded Up Robust Features,” in *Proc. of the 9th European Conference on Computer Vision*, 2006, pp. 404-417.

[7] A. M. Bazen and S. H. Gerez, “Elastic minutiae matching by means of thin-plate spline models,” in *International Conference on Pattern Recognition*, Aug 2002.

[8] N. L. Bihan and S. J. Sangwine, “Quaternion principal component analysis of color images,” in *IEEE Int. Conf. Image Process(ICIP)*, 1, 2003, pp. 809-812.

[9] B. Buchsbaum, “A spatial processor model for object color perception,” *Journal of the Franklin Institute*, 31,1980, pp.1-26.

[10] V. Cardei, B. Funt, and K. Barnard, “Estimating the Scene Illumination Chromaticity Using a Neural Network,” *Journal of the Optical Society of America A,* 19(12), Dec 2002.

[11] J. F. Cardosoand and A. Souloumiac, “Blind Beamforming for Non Gaussian Signals,” *IEEE Transactions on Signal Processing*, *46(7)*, 1998, pp. 1878-1885.

[12] B. E. Chapman and D. L. Parker, “3D multi-scale vessel enhancement filtering based on curvature measurements: application to time-of-flight MRA,” *Med Image Anal.*, *9(3)*, 2005, pp.191-208.

[13] F. Ciurea and B. Funt, “A Large Image Database for colour Constancy Research,” in *Proc. IS&T.SID Eleventh colour Imaging Conference, Society for Imaging Science and Technology*, Nov. 2003, pp. 160-163.

[14] A. P. Condurache, T. Aach, S. Grzybowski, and H. G. Machens, “Vessel Segmentation and Analysis in Laboratory Skin Transplant Micro-angiogram,” in *Proceedings of the Eighteenth IEEE Symposium on Computer-Based Medical Systems*, Dublin Ireland, June 2005, pp. 21-26.

[15] A. P. Condurache, T. Aach, K. Eck, J. Bredno, S. Grzybowsky, and H. G. Machens, “Vessel Segmentation for Angiographic Enhancement and Analysis,” in *Bildverarbeitung für die Medizin 2005 (Algorithmen, Systeme, Anwendungen)*, Heidelberg, March 2005, pp.173-177.

[16] H. Deng, W. Zhang, E. Mortensen, and T. Dietterich, “Principal Curvature-based Region Detector for Object Recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Minneapolis MN, 2007.

[17] S. Di Zenzo, “Note: A note on the gradient of a multi-image,” *Computer Vision, Graphics, and Image Processing, 33(1)*, 1986, pp. 116-125.

[18] Y. Du, D. L. Parker, and W. Davis, “Vessel Enhancement Filtering in Three-dimensional MR Angiography,” *J. Magn. Reson. Imag.*, *5*, 1995, pp. 151-157.

[19] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1988.

[20] M. D. Fairchild, *Color Appearance Models*, Wiley & Sons. Ltd, 2005, pp. 146-150

[21] O. D. Faugeras, “Digital image color processing within the framework of a human visual system,” *IEEE transactions on ASSP*, *27*, 1979, pp. 380-393.

[22] G. D. Finlayson, “Retinex viewed as a gamut mapping theory of color constancy,” in *Proc. AIC International Color Association Color 97*, 2, Kyoto Japan, 1997, pp. 527-530.

[23] G. D. Finlayson and S. D. Hordley, “Color Constancy at a pixel”, *Journal of the Optical Society of America A*, *18(2)*, 2001, pp. 253-64.

[24] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, “Color by Correlation: A Simple, Unifying Framework for Color Constancy,” *IEEE Trans. on Pattern Analysis And Machine Intelligence*, *23(11)*, 2001, pp. 1209-1221.

[25] G. D. Finlayson and G. Schaefer, “Solving for colour constancy using a constrained dichromatic reflection model,” *Int. J. of Computer Vision*, *42(3)*, May 2001, pp. 127-144.

[26] G. D. Finlayson and E. Trezzi, “Shades of Gray and Colour Constancy,” in *Proc. IS&T/SID Twelfth Color Imaging Conference: Color Science, Systems and Applications*, *Society for Imaging Science and Technology*, Scottsdale, Nov. 2004. pp. 37-41.

[27] P. Földiák and D. Endres, “Sparse coding,” *Scholarpedia*, *3(1)*, 2008.

[28] D. H. Foster and S. M. C. Nascimento, “Four Issues Concerning Colour Constancy and Relational Colour Constancy,” *Vision Research*, *37(10)*, 1997, pp. 341-350.

[29] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, “Multiscale Vessel Enhancement Filtering,” in *Proceedings of the MICCAI’98 Lecture Notes in Computer Science*, Springer-Verlag, 1496, 1998, pp. 130-137.

[30] J. Frankle and J. McCann, “Method and Apparatus for Lightness Imaging,” US Patent #4,384,336, May 17, 1983.

[31] B. Funt and V. C. Cardei, “Bootstrapping colour constancy,” in *Proc. of SPIE*, 3644,1999, pp. 421-428.

[32] B. Funt and F. Ciurea, “Parameters for Retinex,” in *Proc. 9th Congress of the International Color Association*, Rochester, June 2001.

[33] B. Funt, F. Ciurea, and J. McCann, “Retinex in Matlab,” *Journal of the Electronic Imaging*, Jan. 2004, pp 48-57.

[34] B. Funt and W. Xiong, “Estimating Illumination Chromaticity via Support Vector Regression,” in *Proceedings of the IS&T/SID Twelfth Color Imaging Conference: Color Science, Systems and Applications,* 2004, pp. 47-52.

[35] C. Gatta, A. Rizzi, and D. Marini, “Ace: An automatic color equalization algorithm,” in *Proceedings of the CGIV02 the First European Conference on Color in Graphics Image and Vision*, Poitiers, France, 2000.

[36] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey Vision Conf.*, 1998, pp. 147-151.

[37] M. Hauta-Kasari, J. Parkkinen and T. Jääskeläinen, “Unsupervised spectral image segmentation,” in *Proc. 1st International Conference on colour in Graphics and Image Processing*, Saint-Etienne, France, 2000, pp. 84-88

[38] K. Hidai, H. Mizoguchi, K. Hiraoka, M. Tanaka, T. Shigehara, and T. Mishima, “Robust face detection against brightness fluctuation and size variation,” In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems,* Takamatsu, Japan, 2, 2000, pp. 1379-1384.

[39] M. Hiraoka, M. Fibrank, M. Essenpreis, M. Cope, S.R. Arridge, E. Van Der Zee, and D. T. Delpy, “A Monte Carlo investigation of optical pathlength in homogeneous tissue and its application to near-infrared spectroscopy,” *Phys. Med. Biol., 38,* 1993, pp. 1859-1876.

[40] S. D. Hordley, “Scene Illumination Estimation: Past, Present, and Future,” *Color Research and Application*, *31(4)*, 2006, pp. 303-314.

[41] S. D. Hordley and G. D. Finlayson, “Re-evaluating colour constancy algorithms,” in *Proc. 17th International Conference on Pattern Recognition*, 1, 2004, pp. 76-79.

[42] S. Hordley and G. Finlayson, “Re-evaluation of color constancy algorithm performance,” *Journal of the Optical Society of America A.*, *23(5)*, 2006, pp. 1008-1020.

[43] B. K. P. Horn, “Determining lightness from an image,” *Computer Graphics and Image Processing*, *3*, 1974, pp. 277-299.

[44] P. O. Hoyer, “Non-negative Matrix Factorization with sparseness constraints,” *Journal of Machine Learning Research*, *5*, 2004, pp. 1457-1469.

[45] M. Hunke and A. Waibel, “Face location and tracking for human-computer interaction,” In *Twenty-Eight Asilomar Conf. on Signals, Systems and computers*, Monterey, CA, USA, 2, 1994, pp. 1277-1281.

[46] M. Kass and A. Witkin, “Analyzing oriented patterns,” *Computer Vision, Graphics, and Image Processing*, *37*, 1987, pp. 362-385.

[47] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, “A Variational Framework for Retinex,” *International Journal of Computer Vision*, *52(1)*, 2003, pp. 7-23.

[48] G. D. Knott, *Interpolating Cubic Splines*, Birkhauser Inc, 2000.

[49] V. Kries, “Influence of adaptation on the effects produced by luminous stimuli,” *Sources of Color Science*, The MIT Press, Cambridge MA, 1970, pp. 109-119.

[50] K. M. Lam, “Metamersim and Colour Constancy.” *PhD thesis*, University of Bradford, 1985.

[51] E. Land and J. McCann, “Lightness and Retinex Theory,” *Journal of the Optical Society of America A,* 61, January 1971, pp. 1-11.

[52] T. M. Lehmann and C. Palm, “Color Line Search for Illuminant Estimation in Real-World Scenes,” *J. Opt. Soc. Am. A*, *18*, 2001, pp. 2679-2691.

[53] H. C. Lee, “Illuminant color from shading,” in *Proc. SPIE: Perceiving, Measuring, and Using Color*, 1250, 1990, pp. 236-244.

[54] H. C. Lee, E. J. Breneman, and C. P. Schulte, “Modeling light reflection for computer color vision,” *IEEE Trans. Pattern Anal. Machine Intell.*, *12(4)*, 1990, pp. 402–409.

[55] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” In *Advances in Neural Information Processing Systems*, 13, 2000, pp. 556-562.

[56] R. Lenz, L. Tran and P. Meer, “Moment based normalization of colour images,” in *Proc. of the IEEE Workshop on Multimedia Signal Processing*, 1999, pp. 129-132.

[57] M. R. Luo and R. W. G. Hunt. “A chromatic Adaptation Transform and a Colour Inconstancy Index*,” Color Res. Appl.*, *23*, 1998, pp.154-158.

[58] L. T. Maloney and B. A. Wandell, “Color Constancy: A Method for Recovering Surface Spectral Reflectance,” *Journal of the Optical Society of America A*, *3(1)*, 1986, pp. 29-33.

[59] J. Marguier, N. Bhatti, H. Baker, M. Harville, and S. Süsstrunk, “Color Correction of Uncalibrated Images for the Classification of Human Skin Color,” In *Proc. 15th Color Imaging Conference*, 2007, pp. 331-335.

[60] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietikäinen, “A physics-based face database for colour research,” *Journal of Electronic Imaging*, 2000, 9(1), pp. 32-38.

[61] B. Martinkauppi, “Face colour under varying illumination-analysis and applications,” *Ph.D. Dissertation*, University of Oulu, 2002.

[62] J. McCann, “Lessons Learned from Mondrians Applied to Real Images and Color Gamuts,” in *Proc. IS&T/SID Seventh Color Imaging Conference*, 1999, pp. 1-8.

[63] K. Mikolajczyk and C. Schmid, “An Affine Invariant Interest Point Detector,” in *Proceedings of the 7th European Conference on Computer Vision-Part I*, 2002, pp.128-14.

[64] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors*,” International Journal of Computer Vision*, *60(1)*, 2004, pp. 63-86.

[65] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, “A comparison of affine region detectors,” *International Journal of Computer Vision*, 65(1/2), 2005, pp. 43-72.

[66] A. Ming and H. Ma, “A blob detector in color images,” in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 364-370.

[67] K. Okada, S. Kagami, M. Inaba, and H. Inoue, “Plane segment finder: algorithm, implementation and applications,” in *Proc. of Int. Conf. on Robotics and Automation*, 2001, pp. 2120–2125.

[68] J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen, “Characteristic spectra of Munsell colors,” *Journal of the Optical society of America A*, *6*, 1989, pp. 725-730.

[69] S. C. Pei, J. H. Chang, and J. J. Ding, “Quaternion matrix singular value decomposition and its applications for color image processing,” in *International Conference on Image Processing*, 1, 2003, pp. 805-808.

[70] C. Rosenberg, M. Hebert, and S. Thrun, “Image Color Constancy Using KL-Divergence,” In *IEEE International Conference on Computer Vision*, 2001, pp. 239-246.

[71] T. K. Sarkar, “History of Wireless,” *Wiley-Interscience*, 2006, pp. 20-66.

[72] G. Schaefer, “Robust dichromatic colour constancy,” In *Int. Conf. Image Analysis Recogn.*, vol 2, 2004, pp. 257-264.

[73] S. A. Shafer, “Using color to separate reflection components”, *Color Res. App.*, 10(4), 1985.

[74] L. Shi and B. Funt, “Skin colour imaging that is insensitive to lighting conditions,” in *Proc. AIC 2008 Color for Science and Industry*, Interim Meeting in Stockholm, June 15-18, 2008.

[75] L. Shi and B. Funt, “Dichromatic Illumination Estimation via Hough Transforms in 3D,” in *the Fourth European Conference on Colour in Graphics, Imaging, and Vision(CGIV08)*, Barcelona Spain, June 9-13, 2008.

[76] L. Shi, B. Funt and G. Hamarneh, “Quaternion Color Curvature,” in *Proc. Imaging Science and Technology Fourteenth Color Imaging Conference*, Portland, Oregon, November 10-15, 2008.

[77] L. Shi, B. Funt, W. Xiong, S. S. Kim, B. H. Kang, S. D. Lee, and C. Y. Ki, “Illumination estimation via non-negative matrix factorization,” in *Proc. AIC 2007 Color for Science and Industry, Midterm Meeting of the International Color Association*, Hangzhou, July 2007.

[78] M. Shimada, Y. Yamada, M. Itoh, and T. Yatagai, “Melanin and blood concentration in a human skin model studied by multiple regression analysis: assessment by Monte Carlo simulation,” *Phys. Med. Biol*., *46(9)*, September 2001, pp. 2397-2406.

[79] M. Soriano, E. Marszalec, and M. Pietikäinen, “Color correction of face images under different illuminants by RGB eigenfaces,” in *Proc. 2nd Audio- and Video-Based Biometric Person Authentication Conference (AVBPA99)*, 1999, pp. 148-153.

[80] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, 23, 2004, pp. 501-509.

[81] Jr. T. G. Stockham, “Image processing in the context of a visual model,” in *Proc. of IEEE*., 60, 1972, pp. 828-842.

[82] M. Störring, H. J. Andersen, and E. Granum, “Physics-based modelling of human skin under mixed illuminants,” *J. of Robotics and Autonomous Systems*, 35(3-4), June 2001, pp. 131-142.

[83] R. T. Tan, K. Nishino, and K. Ikeuchi, “Color constancy through inverse-intensity chromaticity space,” *J. Opt. Soc. Am. A,* 21, 2004, pp. 321-334.

[84] J. Toro and B. Funt, “A Multilinear Constraint on Dichromatic Planes for Illumination Estimation,” in *IEEE Transactions on Image Processing*, vol 16, 2007, pp. 92-97.

[85] N. Tsumura, H. Haneishi, and Y. Miyake, “Independent-component analysis of skin colour image,” *Journal of the Optical Society of America A*, *16(9)*, September 1999, pp. 2169-2176.

[86] P. Viola and M. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, *57(2)*, 2004, pp. 137-154.

[87] J. J. Vos, O. Estévez, and P. L. Walraven, “Improved colour fundamentals offer a new view on photometric additivity,” *Vision Research*, 30, 1990, pp. 936-943.

[88] J. Weijer and T. Gevers, “Color constancy based on the grey-edge hypothesis,” in *IEEE Conf. Image Processing*, Sep. 2005, pp.722-725.

[89] J. van de Weijer, T. Gevers, and A. Gijsenij, “Edge-Based Color Constancy,” in *IEEE Transactions on Image Processing*, *16(9)*, 2007, pp. 2207-2214.

[90] G. Wyszecki and W.S. Stiles, “Color Science: Concepts and Methods,” *Quantitative Data and Formulas*, 2nd edition, John Wiley & Sons, New York, 1982, pp. 74-103.

[91] W. Xiong and B. Funt, “Color Channels for Stereo Retinex," in *Third International Conference on colour in Graphics, Imaging, and Vision*, Leeds, June 2006.

[92] W. Xiong and B. Funt, “Estimating Illumination Chromaticity via Support Vector Regression,” *Journal of Imaging Science and Technology*, *50(4)*, 2006, pp. 341-348.

[93] W. Xiong, B. Funt, L. Shi, S. Kim, B. Kang, and S. D. Lee, “Automatic White Balancing via Gray Surface Identification,” in *Proc. of the Fifteenth IS&T Color Imaging Conference*, Albuquerque, 2007, pp. 143-146.

[94] W. Xiong, L. Shi, B. Funt, S. Kim, B. Kang, and S. D. Lee, “Illumination Estimation via Thin-Plate Spline Interpolation,” in *Proc. of the Fifteenth IS&T Color Imaging Conference,* Albuquerque, Nov. 2007.

[95] http://geology.com/world-cities/, accessed November, 2008

[96] http://www.funny-potato.com/jellyfish.html, accessed November, 2008

[97] http://www.nikonsmallworld.com, accessed November, 2008