ESTIMATING THE COLOUR OF THE ILLUMINANT USING SPECULAR REFLECTION AND EXEMPLAR-BASED METHOD

by

Hamid Reza Vaezi Joze

B.Sc., Sharif University of Technology, 2006

M.Sc., Sharif University of Technology, 2008

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

Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Spring 2013

APPROVAL

Name:	Hamid Reza Vaezi Joze
Degree:	Doctor of Philosophy
Title of Thesis:	Estimating The Colour of The Illuminant Using Specular Reflection and Exemplar-Based Method
Examining Committee:	Dr. Greg Mori, Associate Professor Chair
	Dr. Mark S. Drew, Senior Supervisor, Professor
	Dr. Ze-Nian Li, Supervisor, Professor
	Dr. Graham Finlayson, Supervisor, Professor, Computing Science, The University of East Anglia, UK Adjunct Professor, Computing Science, SFU
	Dr. Tim Lee, SFU Examiner Adjunct Professor, Computing Science
	Dr. Maria Vanrell, External Examiner, Associate Professor, Computer Science Universitat Autònoma de Barcelona, Spain
Date Approved:	March 11th, 2013

Partial Copyright Licence

Abstract

In this thesis, we propose methods for estimation of the colour of the illuminant. First, we investigate the effects of bright pixels on several current colour constancy algorithms. Then we use bright pixels to extend the seminal Gamut Mapping Colour Constancy algorithm. Here we define the White-Patch Gamut as a new extension to this method, comprising the bright pixels of the image. This approach adds new constraints to the standard constraints and improved estimates. Motivated by the effect of bright pixels in illumination estimation, we go on to incorporate consideration of specular reflection per se, which tends to generate bright pixels. To this effect we present a new and effective physics-based colour constancy representation, called the Zeta-Image, which makes use of a novel log-relative-chromaticity planar constraint. This method is fast and requires no training or tunable parameters; moreover, and importantly, it can be useful for removing highlights. We then go on to present a new camera calibration method aimed at finding a straight-line locus, in a special colour feature space, that is traversed by daylights and approximately by specular points. The aim of the calibration is to enable recovering the colour of the illuminant. Finally, we address colour constancy in a novel approach by utilizing unsupervised learning of a model for each training surface in training images. We call this new method Exemplar-Based Colour Constancy. In this method, we find nearest-neighbour models for each test surface and estimate its illumination based on comparing the statistics of nearest-neighbour surfaces and the target surface. We also extend our method to overcome the multiple illuminant problem.

Acknowledgments

I would like to express my sincere gratitude to my senior supervisor Dr. Mark S. Drew for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me during all the time of research and writing of this thesis as well as our published papers. I was truly lucky to have you, Mark, as my supervisor and I wish you the best.

Besides my senior supervisor, I would like to thank the rest of my thesis committee members: Dr. Ze-Nian Li and Dr. Graham Finlayson for their encouragement, insightful comments, and productive questions especially Graham with whom I collaborated on four papers and from whom I learned a lot.

I would like to thank Dr. Greg Mori and all my colleagues and friends at the SFU Vision and Media Laboratory during these four years because of providing such a good research environment.

More importantly, heartfelt thanks go to my parents; Dr. Mahmoud Vaezi and Dr. Parin Dadandish thousands miles away in Iran for their support and encouragement for so many years since the first day I started my education in elementary school, these degrees would not have been possible without their support and encouragement. I also would like to thanks my three lovely sisters and my amazing brother all of whom I miss.

Finally, but foremost, I express my deepest love and thanks to my wife, Zahra Mozaffari, who was my everything during these four years while we were far from our beloved home country.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The human visual system perceives the colour of objects to some extent consistently under different illumination conditions [7, 44]. Therefore humans perceive the same colour for an object despite variations in the colour of the light [81]. This ability, which is called colour constancy, helps us to recognize objects. It is known that brain and eye are both involved in this processing. The human visual system receives the light stimulus in the eye and transfers that signal to the brain. The brain recognizes familiar objects in part by their consistent colours, regardless of the illumination condition. The detailed mechanism of colour constancy in human vision is not yet known but it has been shown that colour constancy operates already even at only 4 to 5 months of age [199] while on the other hand 9-week-olds respond equally to changes either of the illuminant or of surface reflectance [43]. For an example of colour constancy, a yellow banana looks yellow to us under white sunlight, red sunset light, and also indoor fluorescent light. It is likely that all animals with colour vision have colour constancy [187] and this property is indeed demonstrated for some species, such as monkeys [54] and goldfish [46].

Fig. 1 shows some images captured under different light conditions. Because of the colour constancy ability in our visual system, we know to some extent the actual colour of objects in these scenes despite the extreme lighting conditions in some of these images. However, it is not easy for computer vision systems to discover the actual colour of objects in these scenes.

In computer vision, colour constancy refers to computational approaches to recovering the actual colour of surface objects. As such, it is usually called computational colour constancy. At the early stage of work in computational colour constancy, computational models based on human vision perceptional theory were used to solve the colour constancy problem (mainly Retinex theory [117]). However, recent colour constancy methods have mostly been based on the statistical or physical

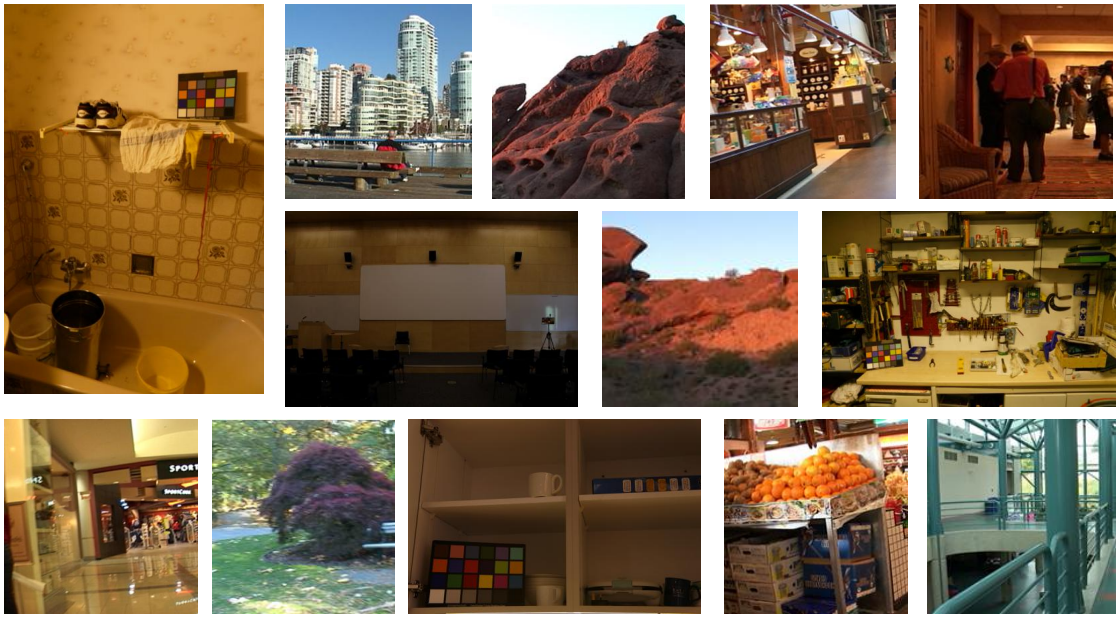


Figure 1.1: Some images captured under different lighting conditions.

aspects of captured images, or have tried to learn a model based on set of training images, and therefore are completely different than human colour constancy.

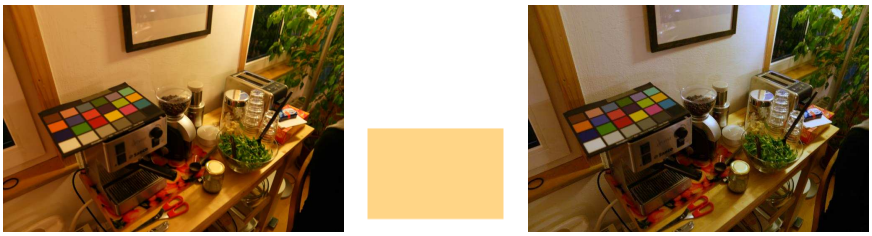
Colour constancy serves an important role in human or animal colour vision system in that it helps them identify objects notwithstanding large differences in illumination. Therefore, it could be essential for any intelligent visual system. In computer vision, colour constancy is essential for many applications such as image retrieval, colour reproduction and object recognition. Therefore many colour constancy algorithms have been proposed by researchers.

On the other hand, in photography white balancing refers to the process of adjusting for the colour of the light, and can typically be used to correct the pixel colours for extreme changes in the colour of light. This process can be done manually or automatically in the camera or using image enhancement software tools. Since automatic white balancing tries to estimate the colour of the illuminant, white balance is the colour constancy method most widely used in still cameras, digital video and image enhancing software tools. However, note that in the white balancing process there is no need to completely *discount* (remove the influence of) the colour of the illuminant since the illuminant is also part of the scene and photographers may prefer to somewhat keep its influence in

the captured image. Current digital cameras as well as video recorders are almost always equipped with some white balancing feature to enable photographers to capture desirable images.

The colour of surface objects are compounded of the actual colour of surfaces, the colour of illuminant, and the camera characteristics, so a computer vision task such as recovering the actual surface colour requires the capability to discount the colour of the illuminant. Colour constancy processing usually includes two main steps: estimating the colour of the illuminant and discounting the effect of the estimated illumination. Fig. 1.2 shows an image captured under an indoor illuminant while the output is the same scene, render under white light.

Specular reflection is the mirror-like reflection of light from a surface, and can be composed of very bright highlights or a more broad sheen on surfaces. In the theoretical case the light reflection from a single incoming direction is reflected into a single outgoing direction. The colour of specular reflections is the same as the colour of the reflected light, within a Neutral Interface Reflection (NIR) [121] condition, which mostly obtains for the surfaces of optically inhomogeneous objects (such as ceramics, plastics, paints, etc.) but not for metals. This property make specular reflection, which is usually in bright areas of image, an appropriate tool for estimating illumination.



(a) (b) (c)

Figure 1.2: (a) Input image captured under indoor illuminant; (b) estimated colour of illuminant; (c) the output, which is the same scene rendered under white light.

1.1 Motivation

Despite extensive research into the computational colour constancy problem, none of the proposed methods is both computationally inexpensive and sufficiently high-performing — most methods fail in some cases. Most simple colour constancy algorithms are based on some assumptions which are not necessarily true in all cases, producing some failure cases [12, 13]; and also in general they do not perform well enough overall compared to recent more complex methods [99]. Complex colour

constancy methods, which mainly derive from learning a model based on training data sets, are computationally expensive and so they are not suitable for real-time application in devices such as digital cameras, video recorders and robots.

In another aspect of the problem, most colour constancy algorithms assume that the spectral distribution of light source is spatially uniform in the image and that the colour of the illuminant is constant across the whole image, Therefore, estimation of the colour of this constant illuminant is the main goal of most of the colour constancy methods we discuss. Notwithstanding the fact that this assumption works well in most cases and is widely used in commercial cameras, there exist common cases for which this assumption is violated in real images. These include: daytime sky-light from windows together with additional indoor light; in-shadow plus non-shadowed lights (say, sunlight plus sky-light producing shadowed and non-shadowed pixels); or two different light sources in an indoor room. This situation, which is called the case of multiple illuminants or multiple light sources, including sources with different colours, is a common failure for current colour constancy methods [55, 105]. For example, inter-reflection produces locally varying light intensity and colour. Therefore, colour constancy in the multiple illuminant situation is still an open problem despite much research for the uniform illuminant situation [105]

As mentioned above, the colour of specular reflections is usually or at least is taken to be the same as the colour of the reflected light [121]. While there are some algorithms which do try to estimate the colour of the illuminant using specular reflection [120, 122, 168, 165], we believe that the information included in the specular reflection component has not been fully used in algorithms to date. Algorithms explicitly using specular reflection have been usually difficult to implement, computationally expensive (require segmentation [120, 168, 165]) or not sufficiently accurate [104]. In this thesis we try to use specular reflection as a main source of information in the problem of estimating illumination, with the aim of proposing simple and fast algorithms that take advantage of the extra information in specularities.

1.2 Thesis Organization

This thesis is organized as follows: In Chapter 2, we discuss background material and previous works related to the topic of this thesis. We describe the physical aspects of specular reflection compared to diffuse reflection and present the simplified model used in computer vision called the dichromatic model. We review algorithms which separate diffuse and specular reflection components. Some of them use single images and some use multiple images. A group of these algorithms

make use of polarization. And another group develops illuminant-dependent colour spaces.

We then describe the computational colour constancy problem and categorize illumination estimation methods into four general groups: statistical, physics-based, gamut based and learning-based methods and review several colour constancy methods in each of these categories. We go into more detail for physics-based methods since these tend more toward considering specular reflection as one main source of information in the problem of estimating illumination.

In the next five chapters, we propose five distinct methods to estimate the colour of illuminant; a statistical method, two physical-based methods, a gamut based method and a learning-based method which include different approaches to this well known problem.

In Chapter 3, we investigate the effects of bright pixels on several current colour constancy algorithms such as the White-Patch, Grey-World, Grey-Edge and the Shades-of-Grey method. Moreover, we describe a simple framework for an illumination estimation method based on bright pixels and compare its accuracy to well-known colour constancy algorithms applied to four standard datasets, and show the beneficial effect of our deliberations. We also investigate failure and success cases using bright pixels, and propose desiderata on input images with regard to the proposed method.

In Chapter 4, we apply the bright-pixel understanding by extending the Gamut Mapping colour constancy algorithm, one of the most important and accurate illumination estimation methods. We define the White Patch Gamut as a new extension to the Gamut Mapping Colour Constancy method, comprising the bright pixels of the image. Adding new constraints based on the possible White Patch Gamut to the standard gamut mapping constraints, a new combined method outperforms gamut mapping methods as well as all other well-known colour constancy methods.

In Chapter 5, we present a new and effective physics-based colour constancy algorithm using specular reflection which makes use of a novel log-relative-chromaticity planar constraint. We call the new feature the Zeta-image. We show that this new feature is tied to a novel application of the Kullback-Leibler Divergence, here applied to chromaticity values instead of probabilities. The new method requires no training data or tunable parameters. Moreover it is simple to implement and very fast. Our experimental results across datasets of real images show the proposed method significantly outperforms other unsupervised methods while its estimation accuracy is comparable with more complex, supervised methods. The addition of new constraints means that every colour constancy method we tried is improved by the new feature. Moreover, and importantly, this new feature can be useful for the purpose of removing specular reflection and highlights.

In Chapter 6, we present a new camera calibration method aimed at finding a straight-line locus, in a special colour feature space, that is traversed by daylights and as well also approximately

followed by specular points. The aim of the calibration is to enable recovering the colour of the illuminant in a scene, using the calibrated camera. Experimental result shows that using such a calibrated locus and the Zeta-image feature we can find the illuminant, with performance competitive with complex methods notwithstanding the fact that many of the images used are not captured under daylight conditions.

In Chapter 7, we focus on surfaces in the image, addressing the colour constancy problem as unsupervised learning of an appropriate model for each training surface in training images. We call this method Exemplar-Based Colour Constancy. We find nearest neighbour models for each surface in a test image and estimate its illumination based on comparing the statistics of pixels belonging to nearest neighbour surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate. The proposed method has the advantage of overcoming multi-illuminant situations, which is not possible for most current methods. We also show a technique to overcome the multiple illuminant situation using the proposed method and test our technique on images with two distinct sources of illumination using a multiple-illuminant colour constancy dataset. The concept proposed here is a completely new approach to the colour constancy problem. We show that it performs very well, for standard datasets, compared to current colour constancy algorithms.

Finally, in the last chapter, we conclude the thesis with a comparison of our five proposed methods for estimating the colour of the illuminant and their criteria such as running time, accuracy, complexity and parameters. Additionally, we finish the thesis with a discussion of future work.

Chapter 2

Background and Related Work

Colour Constancy refers to computational approaches to recover the actual colour of surface objects. The colour of surface objects are compounded of the actual colour of surface and the colour of the illuminant, so recovering the actual surface colour requires the capability to discount the colour of the illuminant. Therefore the colour constancy usually reduces to identification of the illumination since discounting the colour of illuminant is then a much easier task. As we discussed in the first chapter, colour constancy plays an important role in human or animal colour vision systems and helps them identify objects despite large differences in illumination. Therefore, it could be essential for any intelligent visual system. In computer vision colour constancy is essential for many applications such as image retrieval, colour reproduction and object detection [174]. Therefore many colour constancy algorithms have been proposed by researchers [110, 104].

As an example of application of colour constancy, we have shown the utilization of simple colour constancy methods in a standard object classification problem in [174]. We apply the colour constancy algorithms as preprocessing of the bag-of-words learning methods, and compare the performance of these methods with colour invariant descriptors such as C-SIFT [179] and Opponent-SIFT [33] in order to classify 20 different object categories, e.g. bird, bottle, car, dining table, motorbike and people. Although colour constancy as preprocessing does not outperform colour invariant descriptors by itself, we determine that a combination of these local colour constancy methods and colour invariant descriptors improves the performance of object classification substantially, by as much as more than 10 percent, a significant improvement.

Generally colour constancy is achieved by two steps: (1) estimating the colour of the illuminant which is the main part and then (2) discounting the effect of estimated illumination by a transformation of the image, which is called image colour correction. Figure 2.1 shows the process of colour

constancy including its two steps.

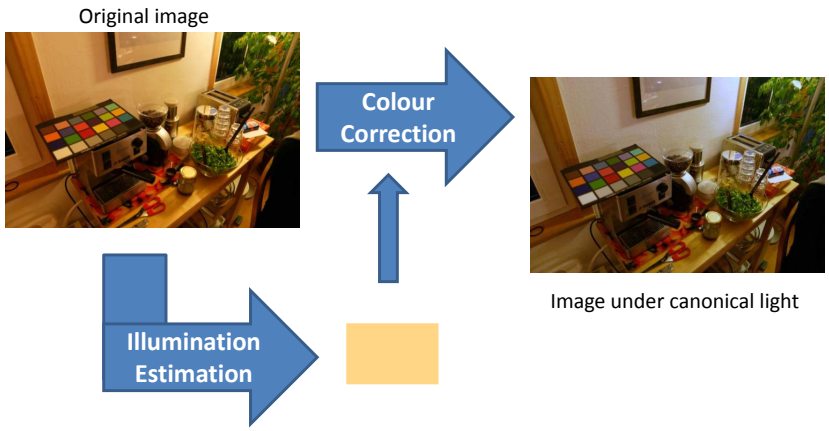


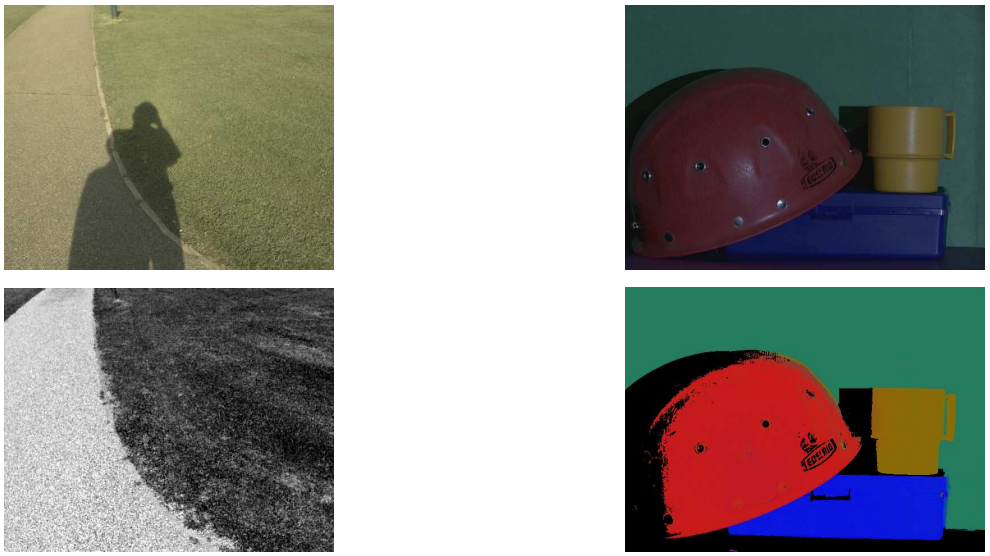
Figure 2.1: The procedure of colour constancy, which includes estimating the colour of the illuminant and colour correction.

In this thesis, we focus on methods that estimate the colour of illuminant using a single image captured by a regular camera. These methods then can be applied to any arbitrary images which have already been captured or images from the Internet, and can also be used in-camera in built-in white balancing techniques. There are some other illumination estimation methods that use additional information from special hardware systems; these are not applicable to previously captured images. Such methods using additional information include be polarized images such as WhitebalPR [77, 153], a new set of camera sensors [201, 140], near-infrared information [83], stereo images [197], extra images [119, 161], or image sequence information such as video [129, 147].

Generally the terms **illumination estimation** or **estimating the illuminant** may refer to estimating the geometry or direction of light [154, 16], estimating the full spectral power distribution of light [45] or estimating of the colour of light. As mentioned already, the main topic of this thesis is the colour of the light and therefore we restrict reference to this specific concept when we refer to illumination estimation or estimating the illuminant in this thesis.

There are also some other kinds of methods that may also be referred to as colour constancy but which however do not recover the actual colour of surface objects and therefore may not include the standard two colour constancy steps of illumination estimation and colour correction. These methods generate a greyscale [68] or multi-channel images [86, 96] as an output which are invariant

to the illuminant [68], specularity [60] or shadowing [64, 51]. Their outputs do not necessarily represent the colour or intensity of the original image but instead are invariant to lighting conditions. Therefore these outputs are usually used as input to high level computer vision tasks such as recognition [86, 97]. Although we discuss some of these methods in Chapter 6 within the broad goal of illumination estimation, these methods are not reviewed as colour constancy methods in this chapter because they are not related to the main topic of this thesis. Figure 2.2 illustrate example outputs of these methods, which produce images invariant to illuminant, shadow or specularity.



(a) (b)

Figure 2.2: (a) The original image and its invariant to the effect of shadowing, from [51]; (b) the original image and its invariant to specularity, from [49].

There are different taxonomies for categorizing methods that estimate the colour of the illuminant such as: supervised methods vs. unsupervised methods; static methods vs. learning-based methods; and physics-based methods vs. non-physics-based methods. Here we categorize illumination estimation methods into four general groups: (1) statistical methods, which try to estimate the illuminant for each image based on its statistical properties, (2) physics-based methods which try to estimate the illuminant using physical models of image formation, (3) gamut based methods, which compare a canonical gamut and image gamut to estimate the illuminant, and (4) learning-based methods which try to estimate illuminants by a model that is learned from training images. However sometimes these categories overlap, such as in the use of statistical and physic-based methods united as static methods, or gamut based and learning based methods united as supervised methods.

In the rest of this chapter we first discuss image colour correction, which is somewhat similar in most colour constancy methods; and then we review statistical, physics-based, gamut based and learning based illumination estimation methods in the following sections. Finally we describe the techniques to evaluate performance of illumination estimation algorithms, and describe metric measures as well as standard data sets that are used for this purpose.

2.1 Image Colour Correction

As mentioned above, after estimating the colour of the illuminant for an input image we need to discount the effect of the estimated illumination by a transformation which generates an output image, one that appears as if it were taken under a reference light source, called the canonical light. That is, we need to transform all colours of the input image to new colours as they appear under the canonical light.

Most colour constancy methods excluding physics-based methods simply follow a Lambertian model that ignores specular reflection. Therefore the image formation model would be similar to the dichromatic model, in eq. (2.11) but without specular reflection. We are interested in also computing the colour of illuminant, $e = (e_1,\ e_2,\ e_3)^T$, in

$$\begin{aligned} I_k &= m_b(\Theta) \int S(\lambda)E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \\ e_k &= \int E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \quad k = 1 \cdots 3 \end{aligned} \tag{2.1}$$

Assuming the colour of illuminant is uniform over the entire image, we can see this transformation as an example of chromatic adaptation [58]. It usually modelled by a 3×3 linear transformation. It means that we compute the colour of each pixel under the canonical light $(R_c,\ G_c,\ B_c)^T$ by transforming the colour of that pixel in the input image, $(R_i,\ G_i,\ B_i)^T$, by multiplication by the 3×3 transformation matrix $D_{3\times3}$:

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = D_{3\times3} \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \tag{2.2}$$

If we assume that either the camera sensors or light are delta functions (narrowband), this linear transformation simplifies to a diagonal transformation, also known as the von Kries Model [186]. The exact criteria in which the linear transformation can replaced by a diagonal transformation and the strength of these conditions is detailed in [189, 62, 88].

Let us assume theoretical narrowband camera sensors such as in Fig. 2.4(b), where $Q_k(\lambda) = A_k\delta(\lambda-\lambda_k)$; then we will have $I_k = A_k m_b(\Theta)\int S(\lambda_k)E(\lambda_k)Q_k(\lambda_k)$ or equally $I_k = m_s(\Theta)s_k e_k$. Doing the same for the canonical light $\mathbf{e^c}$, we will have RGB values under canonical lights equal to $I_k^c = m_b(\Theta)s_k e_k^c$ since $m_s(\Theta)$ and s_k do not change under different light conditions. Therefore we can infer that the ratio between pixel values under the canonical light to pixel values under input lights is exactly equal the ratio between canonical lights to input light separately for each channel, or equally $I_k^c = e_k^c/e_k I_k$. We can encapsulate this in a diagonal transformation model as:

$$D_{diag} = \begin{pmatrix} d_1\ 0\ 0 \\ 0\ d_2\ 0 \\ 0\ 0\ d_3 \end{pmatrix} \qquad d_k = e_k^c/e_k \quad k = 1\cdots 3 \tag{2.3}$$

Although the canonical light could be an arbitrary light, it is usually considered to be a pure white light defined as $(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{3})^T$; therefore $d_k = 1/e_k$.

Although colour correction via a diagonal transform is just approximate, and also it is based on the assumption that the colour of light is uniform, it is widely used in the colour constancy field because of its simplicity and good level of approximation [192, 61, 30]. Real camera sensor curves are not in fact narrowband: Fig. 2.4(a) shows measured sensor sensitivity curves for a Sony DXC-930 3-CCD camera. It has however been shown that if the camera sensors are sufficiently broad-band that have significant overlap, then the diagonal approximation is quite poor, and the sensor responses should be transformed by a particular 3×3 linear transform in order to make camera sensors as narrow-band as possible [9]. This transformation, called spectral sharpening [63], was shown to improve colour constancy, especially for broad-band camera sensors [10, 9]. If we have the sharpening transform matrix T for a pixel of an input image, first we sharpen the sensor using $T(R_i,\ G_i,\ B_i)^T$, then we correct the image by a diagonal image colour correction transform D_{diag}, and then finally we need to return to the actual camera sensor colour space by a transform using T^{-1}.

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = T^{-1}D_{diag}T \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \tag{2.4}$$

The diagonal transformation colour correction model is common for most of the colour constancy methods reviewed in this thesis, usually not using spectral sharpening. This includes even

those methods that use the dichromatic model to estimate the illuminant, in this chapter as well as in the proposed colour constancy methods in this thesis.

Notwithstanding the fact that the diagonal transformation model is widely used in colour correction, other models have also been introduced. In colour correction using a linear transformation, in eq. (2.2), we have 9 unknown matrix entries whereas if we simply measure the colour of a white surface under two illuminants, that provides only three equations, which is insufficient information. Of course, for a diagonal transform this would suffice because of there being only 3 unknown matrix entries. Funt and Jiang [87] select three 3×3 basis matrices using principal component analysis over many possible non-diagonal colour correction matrix transforms and thus find their coefficients from those three equations to form a non-diagonal colour correction matrix transform.

2.2 Statistical Colour Constancy

The Retinex theory [117] presented by Land was one of the first computational theories that attempted to explain human color constancy. This theory is based on the observation that the light incident on a white patch is unchanged after reflection. Additionally, a white reflectance must induce maximal camera responses. Then, the foundational colour constancy method, the White-Patch or Max-RGB method, estimates the illuminant colour from the maximum response of the three colour channels based on [117].

$$e_k = \max_x(I_k(x)) \quad k = 1 \cdots 3 \tag{2.5}$$

where $(e_1,\ e_2,\ e_3)^T$ is the estimated colour of the light and $I(x)$ is the RGB value of a pixel at position x. Since in illumination estimation we are dealing with the colour of the light rather than its intensity, we may restrict our considerations to the chromaticity of an estimated light, which is $(e_1,\ e_2,\ e_3)^T / \sum_k(e_k)$ instead of its actual value. Note that the White-Patch method usually deals with the single brightest pixel in the image for each channel, so it could be noisy and non-robust. Recent research, such as [91, 90, 57], has suggested that carrying out a local mean calculation preprocessing step can significantly improve performance. Some methods [198, 125] try to identify these white-patches or generally grey surfaces in the image. If there is a grey surface in the image which can be correctly identified then the colour of that surface is a good estimate of the colour of the light source since it should be unchanged after reflection.

Another well-known colour constancy method is based on the Grey-World hypothesis [32], which assumes that the average reflectance in the scene is achromatic. Thus it estimates the colour

of the illuminant as the average colour of all pixels in the image.

$$e_k = \frac{\int I_k(x)dx}{\int dx} \quad k = 1 \cdots 3 \tag{2.6}$$

Another alternate method for the Grey-World method is to segment the image and then compute the average colour of all segments of the image [94]. Using segments as surfaces in the image instead of pixels usually improves the result since large coloured segments can dominate the estimates. A recent extension of the Grey-World hypothesis [182, 146] assumes that the average reflectance of semantic classes (or specific visual objects such as sky, grass etc.) in an image is equal to a constant colour, rather than just being grey. Therefore the main problem becomes finding the semantic classes and their constant colour. We shall discuss these methods further when we come to discuss learning based methods later in this Chapter.

Finlayson and Trezzi [75] introduced the Shades of Grey method, using the Minkowski p-norm instead of averaging:

$$e_k = \left(\frac{\int I_k^p(x)dx}{\int dx} \right)^{\frac{1}{p}} \quad k = 1 \cdots 3 \tag{2.7}$$

For $p = 1$ the equation is equal to the grey-world assumption and for $p \to \infty$ it is equal to colour constancy by White-Patch; and it is Shades of Grey for $1 < p < \infty$.

The Grey-Edge method is a recent and important version of the Grey-World hypothesis that states: the average of reflectance differences in a scene is achromatic [180]. Hence, the estimated illuminant is the average over gradients of an image instead of its RGB values themselves.

All grey-based methods, include White-Patch, Grey-World, Shade of Grey and Grey Edge methods which form the main part of statistical methods category have been formalized into a single framework:

$$e_k = \left(\int \left\| \frac{\partial^n I_k(x)}{\partial x^n} \right\|^p dx \right)^{\frac{1}{p}} \quad k = 1 \cdots 3 \tag{2.8}$$

where n is grey-edge order. If $n = 0$, for $p = 1$ the equation is equal to the grey-world assumption, for $p \to \infty$ it is equal to colour constancy by White-Patch and it is Shades of Grey and of Grey for $1 < p < \infty$. For higher n it is Grey-Edge.

The main advantages of statistical methods are their simplicity and speed, whereas they are usually not very accurate in and of themselves. However, the Grey-Edge method performs very well compared to other statistical colour constancy methods and because of its simplicity it is quite popular amongst such non-complex colour constancy methods. The Grey-Edge method has been

extended by considering different kinds of edges (material, shadow, specular, colour shadow and inter-reflection edges) and the use of different weights for them in the process of averaging [103]. However, although these considerations make the estimate more accurate, they also make the method more complex.

2.3 Physics-Based Colour Constancy

Most physics-based colour constancy methods use specularity to estimate the illuminant, via the dichromatic model. We begin this section with describing Dichromatic Reflection Model and then we continue with presenting different physics-based colour constancy methods. For more detail refer to appendix A where we describe the physical aspect of specular reflection compared to diffuse reflection and reviewed algorithms which use or remove specular reflection components.

2.3.1 Dichromatic Reflection Model

Shafer [157] introduced the dichromatic reflection model in 1985 to describe the separation between specular and diffuse reflection components. He assumed a single light source in the scene, without ambient light or inter-reflection between objects. The Dichromatic Reflection Model [157] describes the light, $L(\lambda, \Theta)$, which is reflected from a point on a dielectric, nonuniform material as a mixture of the light $L_s(\lambda, \Theta)$ reflected at the material surface and the light $L_b(\lambda, \Theta)$ reflected from the material body. The vector $\Theta = (\theta_i, \phi_i, \theta_r, \phi_r)$, describes the angles of the incident and emitted light and the phase angle relative to the surface normal; λ is the wavelength parameter. L_s is called the surface reflection component or the specular reflection component. It generally has approximately the same spectral power distribution as the illumination and appears as a highlight or as gloss on the object. L_b is called the body reflection component or diffuse reflection component. It provides the characteristic *object colour* of the surface, and it also includes the properties of object shading:

$$L(\lambda, \Theta) = L_b(\lambda, \Theta) + L_s(\lambda, \Theta) \tag{2.9}$$

The model separates the spectral reflection properties of L_s and L_b from their geometric reflection properties, modelling them as products of spectral power distributions, $c_s(\lambda)$ or $c_b(\lambda)$, and geometric scale factors, $m_s(\Theta)$ or $m_b(\Theta)$, which describe the intensity of the reflected light. Substituting these terms into equation (2.9), we finally obtain the Dichromatic Reflection Model equation:

$$L(\lambda, \Theta) = m_b(\Theta)c_b(\lambda) + m_s(\Theta)c_s(\lambda) \tag{2.10}$$

The model thus describes the light that is reflected from an object point as a mixture of two distinct spectral power distributions, $c_s(\lambda)$ and $c_b(\lambda)$, each of which is scaled according to the geometric reflection properties of surface and body reflection. In the infinite dimensional vector space of spectral power distributions (i.e., real-valued functions of wavelength), the reflected light can be described as a linear combination of the two vectors $c_s(\lambda)$ and $c_b(\lambda)$.

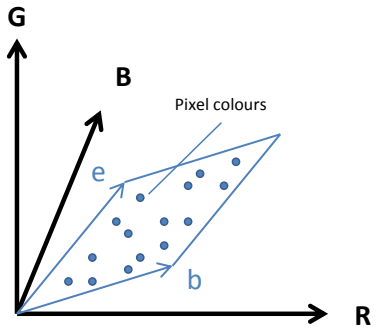


Figure 2.3: Pixel values on a surface lie on a parallelogram in colour space.

Taking into account the spectral power distribution of light source $E(\lambda)$, which equals $c_s(\lambda)$ in the neutral interface reflection condition, a camera sensitivity function $Q_k(\lambda)$ and surface spectral reflection $S(\lambda)$ which is $c_b(\lambda)$, an image formation equation for RGB colour I_k, with $k = 1 \cdots 3$ referring to the three R,G,B channels, can thus be written as follows:

$$\begin{aligned} I_k &= m_b(\Theta) \int S(\lambda)E(\lambda)Q_k(\lambda)\mathrm{d}\lambda + m_s(\Theta) \int E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \\ &= m_b(\Theta)b_k + m_s(\Theta)e_k\,, \qquad k = 1 \cdots 3 \end{aligned} \tag{2.11}$$

where

$$b_k = \int S(\lambda)E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \qquad \text{and} \qquad e_k = \int E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \tag{2.12}$$

Under this assumption, the colours of all pixels from a uniformly-coloured patch of an object are linear combinations of the same interface and body reflection colours $\mathbf{b} = (b_1, b_2, b_3)^T$ and $\mathbf{e} = (e_1,\ e_2,\ e_3)^T$. Colour variation within a uniform object area thus depends only on the geometric scale factors m_s and m_b while $\mathbf{b}$ and $\mathbf{e}$ are constant colour 3-vectors. Accordingly, $\mathbf{b}$ and $\mathbf{e}$ span a *dichromatic plane* in the colour space, and the colours of all pixels from one object patch lie in this

plane. Moreover, pixel values from a surface patch lie within a parallelogram in this dichromatic plane bounded by colour vectors $\mathbf{b}$ and $\mathbf{e}$ (Fig. 2.3).

We can also express the dichromatic model as a special case of the bidirectional reflectance distribution function (BRDF) model. The BRDF, $f(\lambda, \Theta)$, is a four-dimensional function that defines how light is reflected at an opaque surface. The function takes an incoming light and outgoing directions, both defined with respect to the surface normal $\hat{n}$, and returns the ratio of reflected radiance exiting in the outgoing direction to the irradiance incident on the surface from the incoming direction. Each direction is parameterized by azimuth and zenith angles.

Under the NIR model, we then have

$$f(\lambda, \Theta) = c_b(\lambda) f_d + f_s(\Theta) \tag{2.13}$$

where $c_b(\lambda)$ is surface spectral reflection as described above and functions f_d and f_s are the diffuse and specular BRDFs, respectively. For a Lambertian surface the function f_d is constant. Again, taking into account the spectral power distribution of light source $E(\lambda)$ and a camera sensitivity function $Q_k(\lambda)$, the RGB image formation equation for a surface element with surface normal $\hat{n}$, illuminated by a light from direction $\hat{l}$, is written

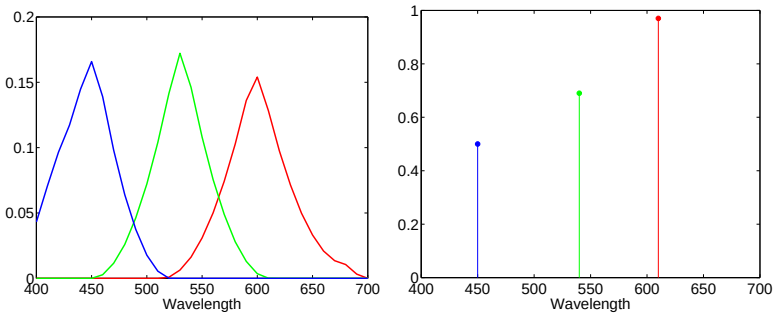
$$I_k = (f_d b_k + f_s(\Theta) e_k) \hat{n} \cdot \hat{l} \tag{2.14}$$

where $\mathbf{b}$ and $\mathbf{e}$ are as defined in eq. (2.12).

In order to simplify the image formation model, some works [68, 60] assume theoretical infinitely-narrowband camera sensors such as Fig. 2.4(b); the Figure compares such theoretical sensors to real camera sensors such as those for the Sony DXC-930 3-CCD camera, shown in Fig. 2.4(a). Here, let us assume theoretical narrowband camera sensors such as those shown in Fig. 2.4(b) where the curves are a discrete version of Dirac delta functions, i.e., sensors which are sensitive only at discrete single wavelengths: $Q_k(\lambda) = A_k \delta(\lambda - \lambda_k)$; then we can simplify the definition of interface and body reflection colours (eq. (2.12)) in the dichromatic equation (eq. (2.11)) as follows:

$$\begin{aligned}
e_k &= \int E(\lambda) Q_k(\lambda) \mathrm{d}\lambda = E(\lambda_k) Q_k(\lambda_k) \\
s_k &= S(\lambda_k) \\
b_k &= \int S(\lambda) E(\lambda) Q_k(\lambda) \mathrm{d}\lambda = S(\lambda_k) E(\lambda_k) Q_k(\lambda_k) = s_k e_k \\
I_k &= m_b(\Theta) s_k e_k + m_s(\Theta) e_k \quad k = 1 \cdots 3
\end{aligned} \tag{2.15}$$

Therefore, knowing the colour of light $\mathbf{e}$ and colour of body surface under uniform white light s_k, the body reflection colour under $\mathbf{e}$ is equal to $b_k = s_k e_k$.



(a) (b)

Figure 2.4: (a): Measured RGB camera sensors — Sony DXC930 camera. (b): Theoretical narrow-band RGB camera sensors.

Even though the dichromatic model was originally introduced by Shafer [157] to describe the reflection of dielectric materials, it has been applied successfully as an approximation of the reflection of many different materials such as human skin [162, 133].

The specular and diffuse components of reflection are usually distinguished by their colour, but there are some conditions where these two components are seen by the cameras as the same. Considering the definition of specular and diffuse reflection in the dichromatic model in eq. (2.12), these two are equal if $S(\lambda)$ is uniform, which means that the surface reflects at all wavelengths equally — it is a uniform grey (or white) surface.

For an object with uniform surface reflection function ($S(\lambda)$), the reflected light is proportional to the incoming light distribution function ($E(\lambda)$). Therefore we can distinguish the colour of the light (e in eq. (2.12)) using a white surface. On the other hand, using equi-energy light which means uniform $E(\lambda)$, the reflected light is proportional to the surface reflection function and we can distinguish the actual colour of the surface (c in eq. (2.12)).

2.3.2 Physics-Based Colour Constancy Methods

Lee [120] proposed a method which uses specularity to compute illumination and this method constitutes one of the first attempts. He used the fact that in the CIE chromaticity diagram [194] the coordinates of the colours from different locations on the same surface will fall on a straight line

connecting the illuminant point and the surface colour point. This is true because the light reflected from a uniform surface is an additive mixture of the specular component and the diffuse component, as discussed in the discussion of the dichromatic model eq. (2.11). Therefore, the coordinates of the chromaticity value of each surface in the CIE chromaticity diagram [194] forms a straight line which passes through the illuminant chromaticity. Hence if there are more surfaces of different colours in the scene then more lines can be determined, and the intersection of these lines is the illuminant chromaticity result, as in Fig. 2.5.

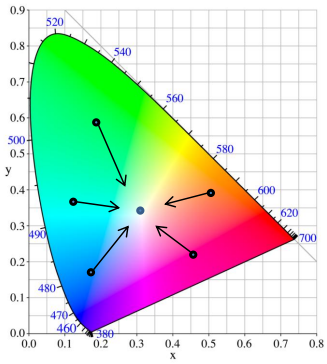


Figure 2.5: The coordinates of the colours from different locations of the same surface will fall on the straight line connecting the illuminant point and the surface colour point.

To implement his idea, Lee used steepest descent and steepest ascent values for each edge point to form a straight line. If steepest descent and steepest ascent values describe the same line, this means they belong to the same surface colour. Finally, the illuminant chromaticity can be computed by finding the intersection of these straight lines by majority voting or a variety of methods. Note that finding these lines in chromaticity space is not in fact easy for real images. Although this algorithm could in principle make good use of a segmented image for finding each straight line, nevertheless for real textured surfaces the segmentation itself is also a difficult task.

Recently, a few methods have been proposed that extend Lee's algorithm [120]. Lee's algorithm is not robust to noise and is limited in the handling of textured surfaces. Lehmann and Palm [122] introduced a method to estimate the colour of a single illuminant for noisy and microtextured images, based on Lee's algorithm. Their approach, named the Colour Line Search, use dichromatic regions of different coloured surfaces. Reliable colour lines are determined directly in the domain of the colour diagram by three steps. First, regions of interest are automatically detected around specular

highlights, and local colour diagrams are computed. Second, colour lines are determined according to the dichromatic reflection model by Hough transform in the corresponding colour diagrams. Third, a consistency check is applied by a corresponding path search in the image domain. The success of this technique depends on an assumption that, in each highlight region, the surface colour is uniform. As a consequence, the technique fails when dealing with complex textured surfaces, which usually have more than one surface colour in their highlight regions.

Another approach to extend Lee's algorithm is to define constraints on the colours of illumination, which makes estimation more robust. Finlayson and Schaefer [73] proposed imposing a constraint on the colours of illumination in 2D chromaticity space. This constraint is based on the statistics of natural illumination colours, and it improves stability in obtaining the intersection. They proposed a convex illuminant constraint and alpha shape [6] constraint. Considering a set of points S and and parameter α, the 2-dimensional alpha shape does not include points when there exists a circle of radius α that does not contain any point in S. Therefore, when α is zero the alpha shape is same as a convex hull and when α is infinite, it is identical to S. They show that using alpha shape constraint, they can remove non-likely lights from the convex hull of natural lights. However because of the non-convexity of the alpha shape the problem needs to proceed by dividing the problem into smaller convex problems.

Furthermore, Finlayson and Schaefer [74] also proposed the use of the Planckian locus as a constraint to accomplish illumination estimation from uniformly coloured surfaces. This Planckian constraint on the illumination chromaticity makes the estimation more robust, especially for natural scene images.

Tominaga and Wandell [168] suggested that the SPDs of all possible observed colours of a dielectric object with highlights are on the hyperplane spanned by the SPD of the diffuse reflection component and the specular reflection component. They called this hyperplane the "colour signal plane". Each object colour forms its own colour signal plane. The SPD of the specular reflection component, which is taken to be the same as the SPD of the illuminant, can be obtained by taking the intersection of these colour signal planes. They use a singular value decomposition technique to determine the intersection of colour signal planes. Fundamentally, within their finite-dimensional model restriction, their method is equivalent to Lee's method. Therefore, Tominaga and Wandell's method has the same limitations that we described for the latter.

As discussed in §A.3, Klinker et al. [113] show that the colours of all pixels from an surface object form a colour cluster in its particular dichromatic plane. And these colour pixels clusters are often in the shape of a *skewed-T*, where the two limbs of the skewed-T correspond to diffuse and

specular reflection. They also proposed a method to determine the colour of the illumination by finding the intersection of these dichromatic planes. If several glossy objects of different colours are illuminated by the same light source, each object produces a dichromatic plane. Because all of these dichromatic planes contain the same interface reflection vector, they intersect along a single line which is the colour of the illumination. However in real images, finding dichromatic planes and also T-shape clusters is quite difficult in the face of noise and multiple coloured surfaces.

Schaefer and et al. [156] try to robustly estimate the illuminant using intersection of dichromatic planes. They find several blocks in an image as surfaces and form the dichromatic plane for each. Therefore intersection of any two dichromatic planes will be an estimate of the light. Then constraints are applied to remove estimates which belong to planes with similar orientations as well as estimates which fall outside the convex hull of s pre-defined set of common lights. Furthermore they form a likelihood for each of these estimates based on their angular distance to other estimates. The final illuminant estimate is then computed by maximum likelihood. They also show that they can integrate these likelihoods with the likelihood from the color-by-correlation method [69], using weighted averaging to increase the accuracy of the method.

Tan et al. [165] introduced a two-dimensional space denoted the "inverse-intensity chromaticity space" to estimate the illumination colour based on the dichromatic model. They use specular chromaticity Γ and diffuse chromaticity Λ as defined in eq. (A.3) to rewrite the dichromatic equation as

$$I_k = m_d(x)\Lambda_k(x) + m_s(x)\Gamma_k \tag{2.16}$$

Then by replacing each channel's image intensity to compute chromaticity ρ equation as defined in eq. (A.2), with its definition in this equation we have following equation for each pixel (note that this assumes uniform illuminant, so Λ is constant while Γ, m_d and m_s vary for each pixel):

$$\rho_k = \frac{m_d\Lambda_k + m_s\Gamma_k}{m_d\Sigma\Lambda_i + m_s\Sigma\Gamma_i} \tag{2.17}$$

Then we will have

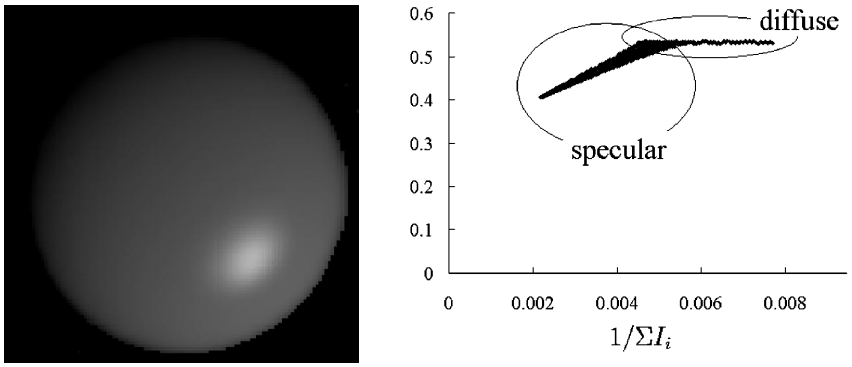
$$I_k = m_d(\Lambda_k - \Gamma_k)\left(\frac{\rho_k}{\rho_k - \Gamma_k}\right) \tag{2.18}$$

By introducing p, which is defined as $p = m_d(\Lambda_k - \Gamma_k)$, we can then derive

$$\rho_k = p\frac{1}{\Sigma I_i} + \Gamma_k \tag{2.19}$$

which forms the core of their method. This shows that by solely calculating the value of p, we are able to determine the illumination chromaticity (Γ), since image chromaticity and total image

intensity can be directly observed from the input image. Observation shows that diffuse and specular pixels of a unique surface form a reliable property in inverse-intensity chromaticity space (see Fig. 2.6).



(a) (b)

Figure 2.6: (a) Synthetic image with a single surface colour. (b) Diffuse and specular points of a synthetic image in inverse-intensity chromaticity space. The figures are taken from [165].

They use a Hough transform in order to estimate the illumination chromaticity from inverse-intensity chromaticity space. Using the Hough transform does not seem to give any solution, because the values of p are not constant throughout the image, and this causes the intersection point of lines not to be located at a single location. Fortunately, even if the values of p vary, the values of Γ_k are constant. Thus, in principle, all intersections will be concentrated at a single value of Γ_k, with a small range of p values. Therefore, if we focus on the intersections in the Hough space, we should find a larger number of intersections at a certain value of Γ_k compared with other values of Γ_k.

2.4 Gamut-Based Colour Constancy

Forsyth's [79] gamut mapping algorithm was at its time and still is one of the most successful colour constancy algorithms, and one of the first colour constancy methods which estimates the illuminant by a model that is learned on training images. Its strength lies in the fact that it makes the uncertainty in the illuminant explicit. It is based on the assumption that in real-world images, for a reference illuminant, one observes only a limited number of colours. Consequently, any variations in the colours of an image are due to variation in the colour of the light. The convex hull of limited set of RGBs that can occur under a reference illuminant is denoted the canonical gamut, denoted $\mathcal{C}$. The

canonical gamut is found in a training phase by observing surfaces under a canonical light source and under known light sources and mapping to canonical light sources [8].

In general, a gamut mapping algorithm accepts an image captured under an unknown light source and generates the input gamut $\mathcal{I}$ using the set of RGB values in the input image, and then determines the set of feasible mappings $\mathcal{M}$, consisting of all possible mappings that can map the input gamut to lie completely within the canonical gamut. Finally an estimator is needed to select one mapping from the set of feasible mappings. The selected mapping can be applied to the canonical illuminant to obtain an estimate of the unknown illuminant. Assuming RGB values are linear responses to the camera sensors, these mappings mostly have been modelled by a diagonal mapping, which is a diagonal matrix that maps image colours under an unknown light source i to their corresponding colours under the canonical illuminant c as discussed in §2.1. The estimator to select one mapping answer could be the diagonal matrix with the largest trace, as originally suggested [79], or the average or weighted average of the feasible set [8].

Although gamut mapping set out a strong novel assumption and acceptable performance for illumination estimation at the time of its proposal, its complexity makes it difficult to implement [59]. Because of these difficulties, a 2-dimensional version of gamut mapping, which uses 2D chromaticities $(R/B \;\; G/B)$ instead of RGB values in order to generate the gamut, was introduced [59]. Of course, it is much weaker than 3D gamut mapping because of of the loss of information in projecting 3D to 2D. Finlayson and Hordley [66, 67] use the 2D version of gamut mapping to find feasible sets of mappings and then transform feasible mappings back to 3 dimensions to select the best mapping which improves performance while reducing the complexity by the use of the 2D version.

The gamut mapping algorithm fails when there is no feasible mapping to map the input gamut to the canonical gamut; in that case the algorithm has no result. To overcome this problem different approaches have been proposed. A simple approach is to increase the size of canonical gamut uniformly in all directions [59]. Another approach is to simulate specularities when computing the canonical gamut for training images [11]. As mentioned above in §A.2, pixel values from a surface patch lie within a parallelogram in the dichromatic plane bounded by colour vectors $\mathbf{b}$ and $\mathbf{e}$ (Fig. 2.3). These parallelograms are usually incomplete because there are not enough specularities; therefore adding specularities means completing these parallelograms [11]. Alternatively, a diagonal-offset model has been proposed proposed [71] for mapping the input gamut to the canonical gamut instead of the diagonal model, by including an extra additive offset $(o_1, o_2, o_3)^T$ to the

usual linear, diagonal transform:

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = \begin{pmatrix} d_1\ 0\ 0 \\ 0\ d_2\ 0 \\ 0\ 0\ d_3 \end{pmatrix} \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix} \tag{2.20}$$

where the offset term is assumed to be small relative to the diagonal terms and is used to overcome the cases with no solution. Therefore the best mapping is the feasible mapping that maximizes the diagonal terms and minimizes the offset term (ideally the offset should be zero); this amounts to minimizing $-d_1 - d_2 - d_3 + o_1 + o_2 + o_3$, as in [71].

Finlayson and Xu [76] introduced an efficient implementation of gamut mapping using convex programming and this is now the common way of implementing gamut mapping. In this implementation the canonical gamut is rewritten in terms of N inequalities which represent the 3-dimensional convex hull of the set of RGBs in the canonical gamut:

$$\mathcal{C}(\rho): \begin{array}{c} a_1R + b_1G + c_1B \geq e_1 \\ a_2R + b_2G + c_2B \geq e_2 \\ \vdots \\ a_NR + b_NG + c_NB \geq e_N \end{array} \tag{2.21}$$

Defining an $N \times 3$ matrix A with ith row equal to $(a_i,\ b_i,\ c_i)$ and an $N \times 1$ vector e, with ith component equal to e_i, we can rewrite this formulation as follows:

$$\mathcal{C}(\rho) : A\rho \geq e \tag{2.22}$$

Now using the diagonal transform we have $\rho = q\ diag(d)$ or equally $\rho = diag(q)\ d$, where d is $(d_1,\ d_2,\ d_3)^T$ and q is the RGB colours of the image under the unknown light source. If we have K 3-dimensional points in the convex hull of input gamut $\mathcal{I}$ for each image, there will be $N \times K$ linear constraints for each image, which defines possible mappings:

$$\mathcal{M} : \forall q_i \in \mathcal{I} : A\ diag(q_i) d \geq e \tag{2.23}$$

In order to estimate the illuminant an optimality criterion must be defined, such as maximizing the L_1 norm of the diagonal.

$$\begin{array}{ll} Maximize & d_1 + d_2 + d_3 \\ \text{subject to} & A\ diag(q)\ d \geq e\ ,\ q \in I \end{array} \tag{2.24}$$

Finlayson et al. introduced Color by Correlation [69, 72] as an improvement on the 2D gamut mapping method. The basic idea of Color by Correlation is that of replacing the canonical gamut with a correlation matrix which describes the extent to which proposed illuminants are compatible with the occurrence of image chromaticities. Rows in this matrix correspond to training illuminants and columns correspond to possible chromaticity ranges resulting from a discretization of space, ordered in any convenient manner. The first version of Color by Correlation [72] was simply a different implementation of a discrete version of 2D gamut mapping, wherein matrix entries were boolean. This was subsequently improved [69] by computing the probability that the observed chromaticities are due to each of the training illuminants, as matrix entries. The estimated illuminant then can be selected by the maximum likelihood of these probabilities, as originally presented, or by the Kullback-Leibler divergence [149].

Gijsenij et al. [102] generalized the gamut mapping assumption that for a reference illuminant, one observes only a limited number of pixel values. They analytically showed that the gamut mapping framework is able to include any linear filter output, including derivatives instead of pixel values. Although using neither derivatives nor linear filter outputs does not outperform gamut mapping using pixel values, nevertheless intersection of these feasible mapping sets with the feasible mapping set generated by pixel based gamut mapping can indeed generate more accurate estimations.

2.5 Learning-Based Colour Constancy

Although the gamut mapping algorithm and its extensions estimate the illuminant by a model that is learned on training images and could thus be considered to be learning-based colour constancy methods, we prefer to categorize gamut mapping in a separate category of its own because of its strength and also the number of extensions as mentioned in the previous section.

One of the first attempts to solve the illumination estimation problem using machine learning techniques used neural networks [84, 35], in which a multilayer neural network system with two hidden layers was designed for the purpose of estimating the (r, g) chromaticity of light. They divide (r, g) chromaticity space into discrete bins as input to input-layer neurons (each bin correspond to one neuron). Thus inputs are binary values representing the presence of a pixel in the image with chromaticity falling in the corresponding (r, g) bin (a binarized 2D chromaticity space histograms). The output layer includes two neurons which are real-valued and correspond to r and g, the chromaticity of illuminant. They are computed by a weighted sum of inner layer neurons operated on by a sigmoid function. This neural network is trained using many synthesized images. Support vector

regression (SVR) was used as another learning tool to solve the illumination estimation problem [92, 196]. In this approach again the inputs to the SVR were binarized 2D chromaticity space histograms, as already used for the neural networks approach [84], but with the addition of a binarized histogram of intensity ($R+G+B$) as a third dimension. As an output, the support vector regression finds the function mapping from image histograms to two dimensional illuminant chromaticities. In a similar approach [188] edge information was used instead of pixel information as an input to support vector regression. The structure which was used in the neural networks approach [35], binarized 2D chromaticity space histograms as input and two dimensional illuminant chromaticities as output, is also applied to different machine learning tools such as ridge regression and kernel regression [4, 3].

Another learning-based approach to illumination estimation problem is the Bayesian approach [150, 93], in which the variability of reflectance of illuminant is modelled as independent random variables. These methods estimate illuminant colour from the posterior distribution condition learned from training images. Here the illuminant prior could be uniform over a subset of illuminants [150] or could be an empirical distribution of illuminants in training images [93].

Besides static colour constancy methods such as Max-RGB, Grey-World, Grey-Edge and Shades-of-Grey, which as mentioned before are based on simple assumptions, recently efforts at fusing these algorithms have generated better performance than for the individual algorithms. One of the first attempts in this direction was carried out by Cardei and Funt [34], which applied a weighted committee mechanism over several of these methods.

Another attempt to combine the estimates from other algorithms is by Schaefer et al. [156]. They integrate a likelihood computed using a physics-based approach (discussed in §2.3) with the likelihood from the color-by-correlation method [69] using weighted averaging to increase the accuracy of the method.

More complex methods try to learn to select the best algorithm or combination of algorithms for each image using pixel information as well as spatial information, and hence they do not deal with the image as simply a bag of pixels.

Gijsenij and Gevers [99, 100] clustered the images by a k-means algorithm using natural image statistics to characterize the images on the basis of Weibull distribution parameters. They then correspond each cluster with the best single algorithm for training images for that cluster. To estimate the illuminant of a test image, they select the algorithm according to its cluster or combination of the individual algorithms according to the distances to neighbouring clusters. They call this method which perform very well Natural Image Statistics. In a similar approach [124], Weibull features of

the training images have been extracted and the best single algorithm assigned with that. Therefore, given a test image, we can find out the K nearest neighbour images based on a Weibull feature from the training image, so the labelled algorithms of that K training images decide the best algorithm or find the best combination in order to estimate illumination for the test image.

In a different approach to selecting best algorithms, Wu et al. [193] introduce a multi-resolution texture descriptor based on an integrated Weibull distribution to extract texture information. They used an image similarity measure derived from the Wiccest feature and spatial pyramid matching to find the K most similar images for a test image from training images, and with these neighbouring images they provide a combination for uniting the data-driven strategy and prior knowledge.

Bianco et al. [20] proposed a two-level learning method to find illumination. First, they learn a classifier to determine if an image is in indoor, outdoor or unsure classes and then they learn a different model for estimating illuminant for each classifier. Therefore for any test image it is first classified into one of the classes and then its illumination is estimated using the model learned for that class.

As mentioned in [104], many learning-based colour constancy methods that try to find the best or combination of algorithms for each image using extracted features go through a similar procedure. They extract texture, shape or colour features from sets of training images, and estimate the colour of illuminant for each of these images using several statistical illumination estimation algorithms. They then learn a model based on extracted features as well as the error of these estimates to ground truth, which is known. This model could e.g. learn the set of weights associate with estimates of these illumination estimation algorithms [100, 193] or directly learn the colour of illuminant [21, 124]. Figure 2.7 shows this procedure in both the training and test phases. It could be stated that the main differences amongst this kind of algorithm are in the feature extraction blocks, where the feature could be simple, such as a colour histogram [20, 21], or Edge direction histogram [21], or more complex features such as Weibull features [100, 124, 193], Wiccest features [193], or Wavelet statistics [21].

Van de Weijer et al. [182] extend the grey world hypotheses to say: the average reflectance of semantic classes in an image is equal to a constant colour, rather than being just grey. Therefore, for each of the semantic classes present in an image they compute the illuminant that transforms the pixels assigned to that class into the average reflectance colour of that semantic class in the training images. In this method, semantic classes are assigned to each 20×20 patch of an image based on models learned in the training phase. This approach is a top-down approach, as opposed to bottom-up approaches in many other colour constancy methods. They also make use of high-level

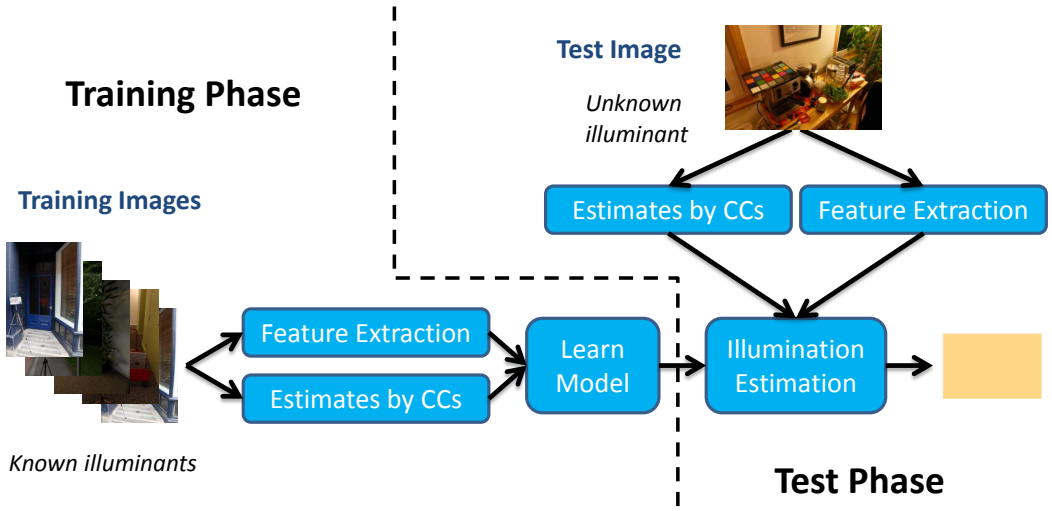


Figure 2.7: The common procedure of learning-based colour constancy methods that try to find the best algorithm or a combination of algorithms for each image using extracted features in both training and test phases.

visual information to select the best illuminant out of a set of possible illuminants generated by other methods.

In a similar approach [146], the special visual object categories (called here memory-colour categories) which have a relatively constant colour such as sky or grass & foliage (which were used in their experiment) are detected using the Bag-of-Features machine learning method. Then the initial estimate provided by a statistical colour constancy method can be adjusted to map the observed colour of the category to its actual colour which is determined in the training phase. The main difference between this work and [182] is that the visual object categories are known and hand labelling and tagging with the category label is required for training images.

There are also methods which assume that skin colours provide enough information for illumination estimation [106, 24]. Bianco and Schettini [24] suggest using faces found by a face detection tool to detect skin areas since many images are captured with faces or people. They show that skin colours tend to form a cluster in various color spaces. They then use the diversity between the gamut of skin pixels and a skin canonical gamut learned using training images to estimate the illuminant.

2.6 Evaluation

In order to evaluate performance of a colour constancy method we need to calculate the error of our estimation. Such error calculations need to be carried out over set of images that are indicative of success or failure, and we need to consider carefully the metric used to compare the overall performance of different methods over those images.

2.6.1 Error of Estimation

Considering the actual chromaticity of illuminant $\mathbf{e}$ and the estimated chromaticity of illuminant $\mathbf{e_{est}}$ by any of of the aforementioned illumination estimation methods, there are different measures used to calculate the error. Two measures commonly used to calculate this error are angular error , which is the angle between vector $\mathbf{e}$ and vector $\mathbf{e_{est}}$ in three dimensional space, and Euclidean error, i.e., Euclidean distance in r, g chromaticity space; note that these two measures are highly correlated to each other [111].

$$err_{angle}(e, e_{est}) = acos(\frac{e \cdot e_{est}}{\|e\|\|e_{est}\|}) \tag{2.25}$$

$$err_{euc}(e, e_{est}) = \sqrt{(e_R - e_{est,R})^2 + (e_G - e_{est,G})^2} \tag{2.26}$$

Gijsennij et al. [101] compare perceptually different measurement for evaluating illumination estimation errors. They show that the angular error is a reasonably good indicator of the human perceptual performance of color constancy methods. They conclude th at the same conclusion holds for Euclidean distance but they show that optimizing the weights for each specific data set can increase the correlation of this measure with perceptual error. They call it the perceptual Euclidean distance.

It is also possible to compute the error of a final colour constancy output, which is the difference between the corrected image and the image of the same scene captured under the canonical illuminant. In this case the error measurement can be the RMS error for all pixels in r, g, b or R, G between the corrected image and the ground truth image. This metric is difficult to obtain since it requires the same images be captured under different lights but with exactly the same geometry. In this thesis, we use angular error as our measurement for computing estimation error because of its frequent use in the literature [111, 12].

2.6.2 Comparison of methods

It is important to compare the performance of illumination estimation algorithms. Having a set of images with known colour of illuminant, we can calculate the error of estimation (using either of the two generally used error metrics) for each image. Then the total metric of performance of an algorithm for that set of images can be the mean of errors. It is shown that the mean by itself is not a good index for evaluating performance of methods [111]. The median or trimean of errors is usually preferred in the literature [111, 101] because the mean is sensitive to outliers. The median indicates the performance of the methods for half of the images or equally the 50^{th} percentile error. The trimean of a distribution is defined as a weighted average of its median and its two quartiles. A more complete way to compare algorithms is to show a plot to visualize the distribution or cumulative distribution of error for a colour constancy method for a specific set of images.

There are some different approaches which try to minimize the worst-case errors or, equally, consider the maximum of errors as the measure. If we require an algorithms which gave us acceptable estimation for 90% or 99% of cases, we need to minimize the 90^{th} or 99^{th} percentile errors [136]. In this thesis we calculate both mean and median as our measurement to compare different illumination estimation algorithms.

2.6.3 Colour Constancy Data sets

In order to evaluate general performance of a colour constancy method we need set of images taken under a known colour of illuminant. Early colour constancy data sets were mostly generated from hyperspectral data. A hyperspectral database by Barnard et al. [14] which is still used to evaluate colour constancy methods under synthetic data includes 1995 surface reflection spectra and 287 illuminant spectra. Therefore, the synthetic data can be the RGB value of each of these surface under any of these illuminants using arbitrary camera sensor functions ($\int S(\lambda)E(\lambda)Q_k(\lambda)\mathrm{d}\lambda$ where $S(\lambda)$ is surface spectra, $E(\lambda)$ illuminant spectra and $Q_k(\lambda)$ are camera sensor functions). Another example of hyperspectral databases which could be used to evaluate colour constancy algorithms are [144, 137, 80], which include hyperspectral information for some scenes from which images under different illuminants can be rendered.

Although hyperspectral data are a good starting point to evaluate a colour constancy method, this is not usually sufficient to reflect the practical amount of failure that is seen for real-world images that include disturbances such as noise, clipping, colour distribution, sensor failure, etc. Therefore in this thesis we place emphasis on colour constancy data sets with real images. Here we introduce

four well known data sets; SFU Laboratory [14], GreyBall [40], ColorChecker [93] and HDR [85] data sets, which are widely used to evaluate colour constancy methods in the literature as well as in this thesis. There are also other colour constancy data sets which are not as popular as these four, such as the Barcelona data set [185, 145] and an extension to ColorChecker data set [36].

SFU Laboratory Data Set : The first data set is Barnard's dataset [14], denoted the SFU Laboratory dataset; this contains images of 31 scenes captured under 11 different measured illuminants (321 images in total). The scenes are divided into two sets as follows: minimal specularities (22 scenes, 223 images – i.e., 19 missing images); and non-negligible dielectric specularities (9 scenes, 98 images – 1 illuminant is missing for 1 scene). The illuminant includes 3 different fluorescent lights, 4 different incandescent lights and those same four incandescent lights with a blue filter. Again this data set is a small dataset which is good for evaluating new methods but because of the limited number of scenes and illuminant as well as the fact that it was taken under laboratory conditions, it is not considered as a practical evaluation for real-world images.

GreyBall Data Set : The next dataset, which contains many low quality real-world images, is the GreyBall dataset of Ciurea and Funt [40]; this contains 11346 images extracted from 15 video clips recorded under a wide variety of imaging conditions (city, mall, indoor, desert, forest, road etc.). The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of each image – and this must be masked off during experiments. For each image, the scene illuminant is measured in terms of mean of RGB values of the pixels on the sphere. However the images have the resolution of only 360×240 pixels and the quality of images is not good because of the movement of the camera while recording clips; nevertheless because of the variation of imaging conditions in this set, it is widely used to evaluate colour constancy methods.

ColorChecker Data Set : Another dataset, which contains out-of-laboratory images, is Gehler's colour constancy dataset [93], denoted as the ColorChecker dataset. This dataset consists of 568 images, both indoor and outdoor. The illuminant ground truth for these images is known because each image has a Macbeth ColorChecker placed in the scene (which must masked off in tests). This data set has higher quality images compared to the above two data sets. Although this set includes 568 images captured from variety of different indoor and outdoor scenes, its variation is not as large as in the GreyBall data set. As well, since images were produced with automatic camera settings, they contain clipped pixels, are non-linear (gamma correction has been applied), are demosaiced,

and are also affected by the camera's white balancing. In order to overcome these problems, Shi and Funt [160] provided a re-processed version of the ColorChecker colour constancy dataset [93] that is based on putative recovered raw data of for the same set of images. Thus usually it is better to apply colour constancy algorithms to the re-processed version instead because it is not already affected by the camera's built-in processes. The ground truth measure of the illumination's RGB colour were computed by dataset providers [93, 160] using six achromatic squares of Colorchecker in each image.They used the median of the RGB channels separately while removing clipped pixels by a threshold.

HDR Data Set : The last colour constancy dataset is the HDR dataset [85] provided by Funt, which contain 105 images constructed in the standard way from multiple exposures of the same scene. The colour of the scene illumination was determined by photographing an extra HDR image of the scene with Macbeth ColorChecker card. Although HDR is a small dataset, it has two advantages compared to other datasets: it has high quality images and no clipped pixels that might have arisen from exceeding the dynamic range. The scene illuminant is determined by manually sampling the RGB values from each of the white patches from the four Colorcheckers which mount in each image. Figure 2.8 shows examples from these datasets.

2.7 Summary

In this chapter we discussed the well known colour constancy problem, which refers to computational approaches to recovering the actual colour of surface objects. We show that this reduces to identification of the illumination. In this thesis we focus on illumination estimation methods using single images captured by a regular camera. Here we categorized illumination estimation methods into four general groups: (1) statistical methods which try to estimate the illuminant for each image based on its statistical properties, (2) physics-based methods which try to estimate the illuminant using physical models of image formation, (3) gamut based methods which compare a canonical gamut with the image gamut in order to estimate the illuminant and (4) learning-based methods which try to estimate the illuminant by a model that is learned on training images. Although we reviewed several colour constancy methods in all of these categories we go into more detail for physics-based methods, which usually consider specular reflection as a main source of information in the problem of estimating illumination. We also show the simplified model used in computer vision as defined in the dichromatic model. We describe the image colour correction as a final step of

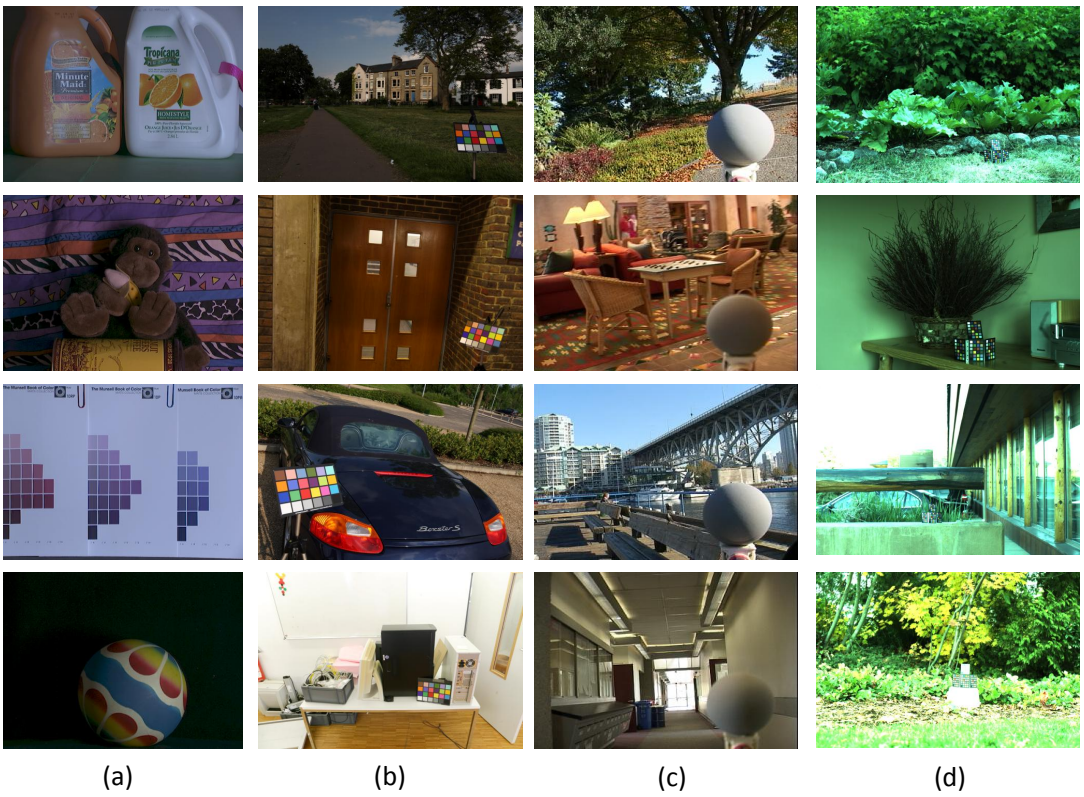


Figure 2.8: Examples from colour constancy datasets: (a) SFU Laboratory (b) ColorChecker (c) GreyBall (d) HDR.

colour constancy, which is to some degree applied similarly in most colour constancy methods. And finally we described techniques for evaluating performance of illumination estimation algorithms, including metric measures as well as describing data sets which useful for this purpose.

Chapter 3

The Role of Bright Pixels in Illumination Estimation

The White-Patch method, one of the very first colour constancy methods, estimates the illuminant colour from the maximum response of three colour channels. However, this simple method has been superseded by advanced physical, statistical and learning based colour constancy methods. Recently, some research works have suggested that the simple idea of using maximum pixel values is not as limited an idea as it seems at first glance. These works show that in several situations some manipulations can indeed make this approach perform very well. Here, we extend the White-Patch assumption to include any of: white patch, highlights or light source; let us refer to these pixels in an image as the "bright" pixels areas. We propose that bright pixels are surprisingly helpful in the illumination estimation process.

In this chapter, we investigate the effects of bright pixels on several current colour constancy algorithms such as White-Patch, Grey-Word, Grey-Edge and Shades-of-Grey methods.

Moreover, we describe a simple framework for an illumination estimation method based on bright pixels and compare its accuracy to well-known colour constancy algorithms applied to four standard datasets. We also investigate failure and success cases, using bright pixels, and propose desiderata on input images with regard to the proposed method.

3.1 Introduction

Illumination estimation, which is the main step in colour constancy processing, is an important pre-requisite for many computer vision applications. One of the first colour constancy methods, the

so-called White-Patch or Max-RGB method estimates the illuminant colour from the maximum response of three colour channels [117]. With the advent of newer and more precise colour constancy methods such as Grey-World [32], Gamut Mapping [79], Grey-Edge [180] and many other complex methods (refer to [110] for an overview), few researchers or commercial cameras use the White-Patch method. On the other hand, recent research such as that on perceptual color contrast enhancement by Choudhury and Medioni [39] or on the "rehabilitation" of MaxRGB by Funt and Shi [91] propose that a local mean calculation such as local blurring as a preprocessing step can significantly improve the performance of this simple method, consisting of simply finding the maximum in each colour channel. Simply put, these works propose it is advantageous to calculate the max of a local mean image.

Recently, Drew et al. [48] found analytically that the geometric mean of bright (generally, specular) pixels is the optimal estimate for the illuminant, based on a standard dichromatic model for image formation (which accounts for the matte and highlight appearance of objects). This work proposes that in the presence of specular highlights the "mean of the max" is the best illuminant estimate, in contradistinction to previous works [91, 39] which say it is the "max of the mean." The analytical approach [48] claims performance comparable with very complex colour constancy methods despite its simplicity.

The bright areas of images can be white surfaces or light sources as well as highlights and specularity, and all are helpful in the illumination estimation process. Highlights and white surfaces both tend to have the colour of light in ideal conditions for dielectric materials such as plastic.

As we discuss in §2.3, the colour of specular reflections is the same as the colour of illumination, within a Neutral Interface Reflection (NIR) [121] condition, which mostly obtains for the surfaces of optically inhomogeneous objects (such as ceramics, plastics, paints, etc.) This property make specular reflection, which is usually in bright areas of image, an appropriate tool for estimating illumination. Many illumination estimation methods derive from the dichromatic model for specular reflection proposed by Shafer [157]. Refer to §2.3 for an overview of these methods.

In this chapter, we investigate the effects of bright pixels on different colour constancy algorithms. We describe a simple framework for an illumination estimation method based on bright pixels and compare its accuracy to well-known colour constancy algorithms applied to four standard datasets. We also investigate failure and success cases, using bright pixels, and draw conclusions on input images with regard to the proposed method.

3.2 Illumination Estimation by Specular reflection

In specular reflection, light from a single incoming direction is reflected into a small cone of outgoing directions. This contrasts with diffuse reflection, where light is partially absorbed and partially scattered within the surface material. Areas of images that are specular tend to be bright. Moreover, the spectral power distribution (SPD) of specular reflections is the same as the illumination's SPD, within a Neutral Interface Reflection (NIR) [121] condition, which mostly obtains for the surfaces of optically inhomogeneous objects (such as ceramics, plastics, paints, etc.); however it does not always hold for the surfaces of optically homogeneous objects (such as gold, bronze, copper, etc.) [107]. These properties make specular reflection, which is usually in bright areas of image, an appropriate tool for estimating illumination. Many illumination estimation methods derive from the dichromatic model for specular reflection proposed by Shafer [157].

Klinker et al. [114] showed that when the diffuse colour is constant over a surface, the colour histogram of its image forms a skewed-T shaped distribution, with the diffuse and specular pixels forming linear clusters. They used this information to estimate a single diffuse colour. Therefore in order to use this principle, their approach needed to segment an image into several regions of homogeneous diffuse colour.

Lee [120] proposed a method which uses specularity to compute illumination by using the fact that in the CIE chromaticity diagram [194] the coordinates of the colours from different points from the same surface will fall on a straight line connected to the specular point. This is the case when the light reflected from a uniform surface is an additive mixture of the specular component and the diffuse component. This seminal work initiated a substantial body of work on identifying specular pixels and using these to attempt to discover the illuminant [122, 164]. Another approach extending these algorithms is to define a constraint on the possible colours of illumination, making estimation more robust [73, 74].

3.3 Extending the White Patch Hypothesis

The White-Patch hypothesis is essentially that there is always a white surface in the image. Let us extend this assumption to include any of: white patch, specularities, or light source (or an effective white, e.g. a bright yellow and red pixel which combined have the same max R, G and B as a white patch). We also use the term *gamut* of bright pixels, in contradistinction to maximum channel response of the White-Patch method, which typically deals only with the brightest pixel in the image.

Obviously, using a single pixel or very small area is noisy and not robust.

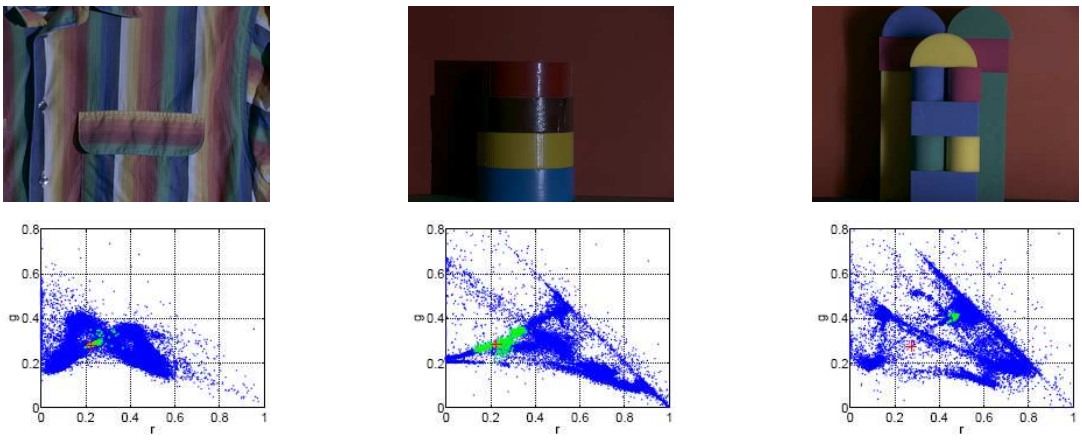
Since are we dealing with bright pixels we need to be very careful about clipped pixels, i.e. pixels where the light reflection exceeds the dynamic range of the camera. Here for each colour channel we remove pixels which exceed 90% of the dynamic range.

Although we can simply define bright pixels as the top $T\%$ of luminance L given by $L = R + G + B$, this is not a precise definition for our application since using this definition we may have set of different pixels tagged as bright pixels for a scene under different illumination conditions. Ideally we to need define bright pixels invariant to illumination, which could e.g. be the top $T\%$ of RGB. Considering the diagonal matrix $diag(d_1, d_2, d_3)$ for transferring the illuminant, the sum of channels will be $d_1R + d_2G + d_3B$ while their product will be $d_1d_2d_3RGB$. Therefore the order of pixels considering the product will be unchanged so it is illuminant invariant while this is not the case for the sum. However since the possible illuminant colours are limited in such that that we rarely see extreme red light, for example, our experiments show than the sum of channels and the product of channels perform similarly. Hence we define bright pixels as the top $T\%$ of $R + G + B$.

To investigate the utility of this assumption, we carry out a simple experiment. We check whether or not the actual illuminant colour falls inside the 2D-chromaticity gamut of bright pixels. We find that the actual illuminant colour falls in the 2D gamut of the top 5% brightness pixels of each image for the SFU Laboratory Dataset [14] for 88.16% of the images, in 74.47% of images for the ColorChecker dataset [160], and in 66.02% of images for the GreyBall Dataset [40]. Fig. 3.1 shows the 2D gamut in chromaticity space $\{r, g\} = \{R, G\}/(R + G + B)$, with the top-5% brightness pixels in green. The actual measured illuminant is shown as a red star. Clearly, as the failure case Fig. 3.1(c) shows, with no supporting evidence it may happen that the illuminant does not fall within the bright region.

When there are no strong highlights, source of light, or white surfaces in the image, the bright pixels are not helpful; in that case there can be areas of an image belonging to the brightest surface which tend towards that particular surface's surface colour. Alternatively this situation may simply arise from a set of single pixels from all over the image.

The fundamental question here is what is the probability of having an image without strong highlights, source of light, or white surface, in the real world? Knowing the answer to this question is vital when we investigate the effect of bright pixels in colour constancy.



(a) (b) (c)

Figure 3.1: Examples of image evidence: top-5% brightness pixels in green, other pixels in blue, and red star showing the correct illuminant point in r, g chromaticity space. (a) Image with white patch; (b) Image with specularity; (c) Image without white patch or specularity.

3.4 The Effect of Bright Pixels in Colour Constancy Algorithms

The foundational colour constancy method, the White-Patch or Max-RGB method, estimates the illuminant colour from the maximum response of the three colour channels [117]. It is based on the assumption that the maximum response in the channels is caused by a white patch. The White-Patch method usually deals with the brightest pixel in the image, so it is noisy and non-robust. Funt and Shi [91, 90] suggested that carrying out a local mean calculation preprocessing step can significantly improve its performance.

Another well-known colour constancy method is based on the Grey-World hypothesis [32], which assumes that the average reflectance in the scene is achromatic. Shade-of-Grey [75] extend Grey-World for higher order using the Minkowski p-norm instead of averaging:

$$e_k = \left(\frac{\int I_k^p(x)dx}{\int dx} \right)^{\frac{1}{p}} \quad k = 1 \cdots 3 \tag{3.1}$$

For $p = 1$ the equation is equal to the grey-world assumption and for $p \to \infty$ it is equal to colour constancy by White-Patch; and it is Shades of Grey for $1 < p < \infty$. At first glance we see no distinction for bright pixels in the Grey-World assumption; However since we use an averaging to compute the illuminant point as shown in eq. (3.1), the higher values, which are brighter pixels, contribute a good deal more compared to dark pixels, especially for higher p.

Grey-Edge is a recent version of the Grey-World hypothesis that states: the average of the reflectance differences in a scene is achromatic [180]. Using the same formulation as for grey-based methods, Grey World, Shades of Grey, and Grey Edge can be combined in a single framework for illuminant estimation methods:

$$e_k = \left(\int \left\| \frac{\partial^n I_k(x)}{\partial x^n} \right\|^p dx \right)^{\frac{1}{p}} \quad k = 1 \cdots 3 \tag{3.2}$$

Where n is the grey-edge "order". Although, the effect of bright pixels is not so essential for the Grey-Edge method, we can use the same reasoning regarding averaging considering that the edge value for brighter pixels are larger than edge value for darker pixels.

Another well known colour constancy method is the Gamut Mapping [79] algorithm, which is more complex and more accurate. The bright pixels are the upper boundaries of the colour gamut for a single image. In chapter 4, we introduce a White-Patch Gamut Colour Constancy [176] algorithm, which includes the top-brightness pixels in a 3D gamut; we will show that adding new constraints based on the white patch gamut to standard Gamut Mapping constraints outperforms the Gamut Mapping method and its extensions.

As a simple experiment in order to investigate the effect of bright pixels, we run White-Patch, Grey-World, Grey-Edge and Shades of Grey methods for the top 20% brightness pixels for each image, and compare to using all image pixels.

We use the standard well-known colour constancy methods: White-Patch, Grey-World, Grey-Edge, and Shades-of-Grey implemented by [180], testing on the re-processed version of the ColorChecker dataset [160], using the dataset's suggested clipping threshold. For those methods which need tunable parameters, we utilize optimal parameters for that dataset.

Table 3.1 shows the accuracy of using top 20% brightness pixels for reprocessed version of the ColorChecker dataset [160], in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. The results indicate that although we only use one fifth of the pixels, performance is better than or equal to using all pixels. This is especially true for Grey-World and Shades-of-Grey (both follow eq. (3.2)), where using top-brightness pixels significantly outperforms using all pixels.

3.5 The Bright-Pixels Framework

Here we propose a simple framework in order to investigate the effect of bright pixels for illumination estimation. First of all, since we dealing with bright pixels we need to be careful about clipped

Table 3.1: Angular errors for several colour constancy algorithms for the linear (no gamma correction applied) ColorChecker dataset [160] using all pixels as compared with using only the top 20% brightness pixels.

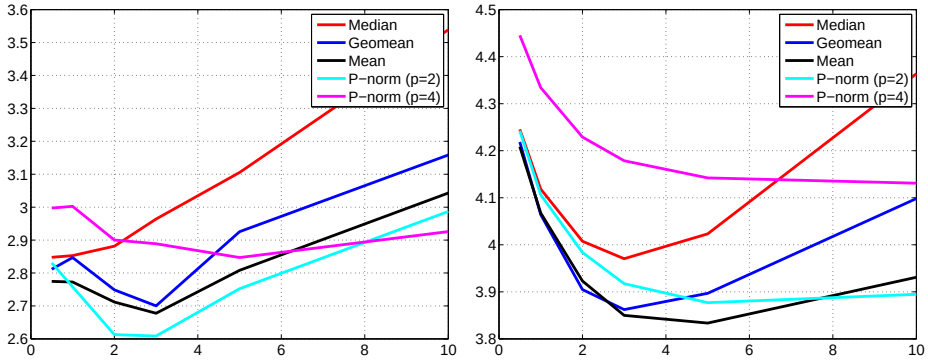
Dataset	All Pixels		20% brightness	
Methods	Median	Mean	Median	Mean
White Patch	6.31°	7.82°	6.31°	7.81°
Grey World	6.33°	6.40°	3.46°	4.23°
Grey Edge ($p = 1, \sigma = 6$)	4.73°	5.56°	4.65°	5.46°
Shades of Grey ($p = 4$)	3.51°	4.45°	3.08°	4.17°

pixels. Therefore we remove pixels exceeding 90% of the dynamic range of the camera for each colour channel. We simply define bright pixels as T percentile of the luminance, taken to be the sum of channels, $R + G + B$.

If these bright pixels represent highlights, a white surface, or a light source, they approximate the colour of the illuminant. Any statistical estimator can be brought to bear for estimating the illuminant, e.g. the median, mean, geometric-mean or the Minkowski p-norm.

Figure 3.2 plots angular errors in terms of mean and median for recovering the illuminant, using T percentile (from 1% to 10%) of brightness pixels, using different statistical estimators: median, geometric-mean, mean and the Minkowski p-norm for $p = 2$ and $p = 4$, for the linear-image ColorChecker dataset [160]. Considering that the best median and mean angular errors in this dataset have been reported as respectively 2.5° using Gamut Mapping in [104] and 3.5° by the complex High Level Visual Information algorithm [182], the achievement is surprisingly good whilst being very simple (refer to Table 3.3 for results for other color constancy methods). We see that optimal performance in terms of the median is for the p-norm estimator, with $p = 2$ for the top-3% brightness pixels; in terms of using the mean, is for the Mean algorithm for top-5% brightness pixels.

Figure 3.3 shows examples of images from the ColorChecker Dataset having angular error more than 13°, using the top-3% brightness pixels and p-norm estimator with $p = 2$. Figure 3.3 indicates that a common failure within a bright pixel framework is when there are multiple illuminants in the scene (we can see the same failures in the GreyBall dataset). Examples are skylight from windows plus indoor light, in-shadow plus non-shadow lights, or two different light sources in an indoor room. Although most color constancy methods assume a single light source, nevertheless in these datasets there are some images with more than one illuminant. Obviously, in the case of more than one illuminant the bright-pixel method finds the brightest illuminant while other methods such as



(a) Median Error (degrees) for T=1% to 10%

(b) Mean Error (degrees) for T=1% to 10%

Figure 3.2: The plots of angular errors in terms of (a) median error and (b) mean error for recovering the illuminant, using T percentile of brightness pixels using different 3-vector statistical estimators: median, geometric-mean, mean and the Minkowski p-norm for $p = 2$ and $p = 4$, for the linear-image ColorChecker dataset [160].

Gamut Mapping find the dominant illuminant or combination of illuminants.

Another failure case happens if bright pixels are not good estimators of the illuminant; or equally there are no highlights, white surfaces, or light sources in the image. Although at first glance this seems to be a common situation, our experiments on current standard color constancy datasets have shown that this happens even less than the multiple-illuminants situation (Figure 3.3 shows a few examples). In this case bright pixels either capture the colour of the brightest surface in the image or a distribution of bright pixels from all over the image. In the former case we can simply check if these pixels are in the possible chromaticity gamut of illuminants; and the latter case can be distinguished based on the distribution of these pixels in chromaticity space.

As we mentioned, a local mean calculation such as local blurring has been shown to improve the performance of simple methods such as White-Patch [91]. Therefore, here we examined applying three different local mean calculations as preprocessing, as follows: (1) resizing to 64×64 pixels by bicubic interpolation; (2) median filtering (inspired by [91]); and (3) a Gaussian blurring filter.

Figure 3.3 shows that the p-norm (and we can consider the mean as p-norm with $p = 1$) is a better estimator than median or geomean. Table 3.2 gives median angular error, with *optimal* parameters (T and p), for the reprocessed ColorChecker dataset using our three local mean preprocessing, for shades of grey and the 1st-order and 2nd-order grey-edge method for top-brightness pixels. For the meaning of "n" the reader is referred to eq. (3.2).

Figure 3.3: Examples of images from ColorChecker Dataset with maximum angular error, using top 3% brightness pixels and p-norm estimator with $p = 2$

Overall, we define illumination estimation by the bright pixel framework by following equation using the grey-based illumination estimation formulation.

$$e_k = \left(\frac{\int_{\phi_T} I_k^p(x) dx}{\int dx} \right)^{\frac{1}{p}} \quad k = 1 \cdots 3 \tag{3.3}$$

where ϕ_T includes the top-T% brightness pixels in the image.

3.6 Further Experiments

We applied the proposed framework to four standard color constancy datasets. The first is Barnard's dataset [14], denoted the SFU Laboratory dataset; this contains 321 measured images under 11 different measured illuminants. The second dataset, which contains out-of-laboratory images, is the re-processed version of the Gehler colour constancy dataset [93], denoted as the ColorChecker

Table 3.2: The median angular errors for the linear-image ColorChecker dataset [160] using top brightness pixels for three variations of eq. (3.2) when different local mean operations are applied as preprocessing. The first value in parentheses for each element is the optimum value of T and the second is the value of p in the p-norm for that experiment.

	Shades of Grey	n=1 grey-edge	n=2 grey-edge
no local mean	2.61° $(2\%, 2)$	4.61° $(5\%, 2)$	4.46° $(5\%, 2)$
64×64 bicubic	2.88° $(3\%, 1)$	4.86° $(5\%, 1)$	4.76° $(5\%, 2)$
Median filter	2.69° $(3\%, 2)$	4.32° $(5\%, 1)$	4.29° $(5\%, 1)$
Gaussian filter	2.72° $(3\%, 2)$	4.37° $(5\%, 1)$	4.13° $(5\%, 1)$

dataset, which was provided by Shi and Funt [160]. This dataset consists of 568 images, both indoor and outdoor. The illuminant ground truth for these images is known because each image has a Macbeth ColorChecker placed in the scene (which must masked off in tests). The third dataset, which contains low quality real-world video frames, is the GreyBall dataset of Ciurea and Funt [40]; this contains 11346 images extracted from video recorded under a wide variety of imaging conditions. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of the image – and this must be masked off during experiments. The last color constancy dataset is the HDR dataset [85] provided by Funt, which contain 105 images constructed in the standard way from multiple exposures of the same scene. The colour of the scene illumination was determined by photographing an extra HDR image of the scene with 4 Gretag Macbeth. Although HDR is a small dataset, it has two advantages compare to other datasets: it has high quality images and no clipped pixels that might have arisen from exceeding the dynamic range.

We search using brute force for optimal parameters: i.e., the value of p in the p-norm, the gradient order n in edge-based p-norm, which local mean method to apply, and finally the top-brightness threshold. Table 3.3 shows the overall optimal performance of a bright-pixels framework for our four standard datasets, compared to the standard methods. The Bright Pixels row represents the optimal value reachable by a bright-pixel framework over all methods White-Patch, Grey-World, and Grey-Edge. For the bright-pixels framework, if the estimated illuminant is not in the possible illuminant gamut for that dataset, meaning that there is no white surface, specularity, or light source in the image, we fall back on the Grey-Edge method instead – this is the row labelled Bright Pixels + grey-edge in Table 3.3. This situation occurs relatively seldom: for 178 out of 11136 images for the GreyBall set, 3 out of 568 for the ColorChecker set, 89 out of 321 for the SFU Laboratory dataset, and 9 out of 105 for the HDR dataset.

Table 3.3: Comparison of the bright-pixels framework with well-known colour constancy methods.

Dataset	SFU Laboratory		Color Checker		Grey Ball		HDR	
Methods	Median	Mean	Median	Mean	Median	Mean	Median	Mean
White Patch	6.5°	9.1°	5.7°	7.4°	5.3°	6.8°	4.3°	6.3°
Grey World	7.0°	9.8°	6.3°	6.4°	7.0°	7.9°	7.3°	7.9°
Grey Edge	3.2°	5.6°	4.5°	5.3°	4.7°	5.9°	3.9°	6.0°
Gamut Mapping	2.3°	3.7°	**2.5°**	4.1°	5.8°	7.1°	-	-
1st-jet Gamut Mapping [102]	2.1°	3.6°	**2.5°**	4.1°	5.8°	6.9°	-	-
Bayesian [93]	-	-	3.5°	4.8°	-	-	-	-
High Level Vis. Info. [182]	-	-	**2.5°**	**3.5°**	-	-	-	-
Natural Image Statistics [100]	-	-	3.1°	4.2°	**3.9°**	**5.2°**	-	-
Rehabilitation of MaxRGB	3.1°	5.6°	-	-	-	-	3.9°	6.3°
Bright Pixels	1.90°	5.84°	2.61°	3.98°	4.71°	5.72°	**3.49°**	**5.77°**
Bright Pixels + grey-edge	**1.62°**	**2.72°**	2.61°	3.96°	4.64°	5.57°	**3.49°**	5.92°

Using eq. (3.2) to test the bright-pixels hypothesis, the optimal parameters for the SFU laboratory dataset are: Gaussian filter as preprocessing plus using the Shades of Grey method with $p = 2$ for the top $.5\%$ brightness pixels. Here we test order n in $\{0, 1, 2\}$, p-norm parameter p in $\{1, 2, 4, 8, 16\}$, brightness threshold T in $\{.5\%, 1\%, 2\%, 3\%, 5\%\}$. The optimal parameters for the ColorChecker dataset are: no preprocessing, and using the Shades of Grey method with $p = 2$ for the top 2% brightness pixels. The optimal parameters for the GreyBall dataset are: no preprocessing, and using the Shades of Grey method with $p = 2$ for the top 1% brightness pixels. The optimal parameters for HDR dataset are: a Gaussian filter as preprocessing, and then the 2nd-order grey-edge method with $p = 8$ for the top 1% brightness pixels.

3.7 Conclusion

In this chapter, we investigate the effects of bright pixels in a variety of standard colour constancy algorithms. Moreover, we describe a simple framework for illumination estimation method based on bright pixels. We have demonstrated that this simple method does very well compared to well-known colour constancy algorithms as well as compared to more complex supervised color constancy methods, over four large standard datasets.

The fundamental question which arises in this section is what is the probability of having an image without strong highlights, source of light, or white surface in the real world? Based on current standard datasets in the field of color constancy we saw that the simple idea of using the p-norm of bright pixels, after a local mean preprocessing step, can perform surprisingly competitively compared to complex methods. Therefore, we conclude that either the probability of having an image without strong highlights, source of light, or white surface in the real world is not overwhelmingly great or the current color constancy datasets are conceivably not good indicators of performance with regard to possible real world images.

Chapter 4

White Patch Gamut Mapping Colour Constancy

As we discussed in the previous chapter, the bright areas of images can be taken to include include highlights and specularity, and also white surfaces or light sources; and indeed all may be helpful in the illumination estimation process. We showed that using top brightness pixels, instead of all pixels, significantly improves the performance of most well known illumination estimation methods.

In this chapter, go on to use bright pixels to extend the Gamut Mapping Colour Constancy algorithm, one of the main illumination estimation methods. We define the White Patch Gamut as a new extension to the Gamut Mapping Colour Constancy method, comprising the bright pixels of the image. Adding new constraints based on the possible White Patch Gamut to the standard gamut mapping constraints, a new combined method outperforms gamut mapping methods as well as other well-known colour constancy methods.

The new constraints that are brought to bear are powerful, and indeed can be more discriminating than those in the original gamut mapping method itself.

4.1 Introduction

Estimation of illumination, the main concern of colour constancy processing, is an important problem in image processing for digital still images or video, forming a prerequisite for many computer vision applications. The foundational colour constancy method, the so-called White-Patch

or Max-RGB method estimates the light source colour from the maximum response of the different colour channels [117]. With the advent of newer colour constancy methods such as Grey-World [32], Gamut Mapping [79], Grey-Edge [180] and many other complex algorithms (see [110] for an overview), few researchers in the field or commercial cameras use the White-Patch method. On the other hand, recent works by Choudhury and Medioni [39] and Funt and Shi [91] proposed that finding the maximum after a local mean calculation such as local blurring can significantly improve the performance of this simple method. Moreover, Drew et al. [48] analytically showed that the geometric mean of pixels in bright (generally, specular) pixel regions is the optimal estimate for the illuminant. This insight says that in the presence of specular highlights illuminant estimation is the mean of the max, while former works [91, 39] argued the converse. The analytical approach [48] claims comparable performance with complex colour constancy methods despite its simple approach.

The Gamut Mapping algorithm, a more complex and more accurate algorithm, was introduced by Forsyth [79]. It is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours. Several extensions have been proposed [8, 67, 59, 102].

The bright areas of images can be highlights and specularity as well as white surfaces or light sources, and all are helpful in the illumination estimation process. Highlights and white surfaces both tend to have the colour of the light in ideal conditions.

In this chapter, we define the White Patch Gamut as a new extension to the Gamut Mapping Colour Constancy method, consisting of the gamut of the bright pixels of an image. New constraints are added as a result. Experiment shows that adding these new constraints to those from standard gamut mapping constraints outperforms gamut mapping methods as well as other well-known colour constancy methods.

4.2 Gamut Mapping

Forsyth's [79] gamut mapping algorithm is one of the most successful colour constancy algorithms. Its strength lies in the fact that it makes the uncertainty in the illuminant explicit. It is based on the assumption that in real-world images, for a reference illuminant, one observes only a limited number of colours. Consequently, any variations in the colours of an image is due to variation in the colour of light . The convex hull of a limited set of RGBs that can occur under a reference illuminant is called the canonical gamut, denoted $\mathcal{C}$. The canonical gamut is found in a training phase by observing surfaces under known light light sources.

In general, a gamut mapping algorithm accepts an image captured under an unknown light source and generates the input gamut $\mathcal{I}$ using the set of RGB values in the input image, and then determines the set of feasible mappings $\mathcal{M}$, consisting of all possible mappings that can map the input gamut to lie completely within the canonical gamut. Finally an estimator is needed to select one mapping from the set of feasible mappings. The selected mapping can be applied to the canonical illuminant to obtain an estimate of the unknown illuminant. Assuming RGB values are linear responses to the camera sensors, these mappings mostly have been modeled by a diagonal mapping, which is a diagonal matrix that maps image colours under an unknown light source i to their corresponding colours under the canonical illuminant c.

$$\rho = \begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = D \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \quad D = \begin{pmatrix} d_1\ 0\ 0 \\ 0\ d_2\ 0 \\ 0\ 0\ d_3 \end{pmatrix} \tag{4.1}$$

The estimator to select one mapping could be the diagonal matrix with the largest trace, as originally suggested [79], or the average or weighted average of the feasible set [8].

Finlayson and Xu [76] introduced an efficient implementation of gamut mapping using convex programming. In this implementation the canonical gamut is rewritten in terms of N inequalities which represent the 3-dimensional convex hull of the set of RGBs in the canonical gamut:

$$\mathcal{C}(\rho): \begin{array}{c} a_1R + b_1G + c_1B \geq e_1 \\ a_2R + b_2G + c_2B \geq e_2 \\ \vdots \\ a_NR + b_NG + c_NB \geq e_N \end{array} \tag{4.2}$$

Defining an $N \times 3$ matrix A with ith row equal to $[a_i b_i c_i]$ and an $N \times 1$ vector e, with ith component equal to e_i, we can rewrite this formulation as follows:

$$\mathcal{C}(\rho) : A\rho \geq e \tag{4.3}$$

Now using the diagonal transform we have $\rho = q\ diag(d)$ or equally $\rho = diag(q)\ d$, where d is $[d_1, d_2, d_3]$ and q is the RGB colours of the image under the unknown light source. If we have K 3-dimensional points in the convex hull of input gamut $\mathcal{I}$ for each image, there will be $N \times K$ linear constraints for each image, which defines possible mappings:

$$\mathcal{M} : \forall q_i \in \mathcal{I} : A\ diag(q_i)d \geq e \tag{4.4}$$

In order to estimate the illuminant an optimality criterion must be defined, such as maximizing the L_1 norm of the diagonal.

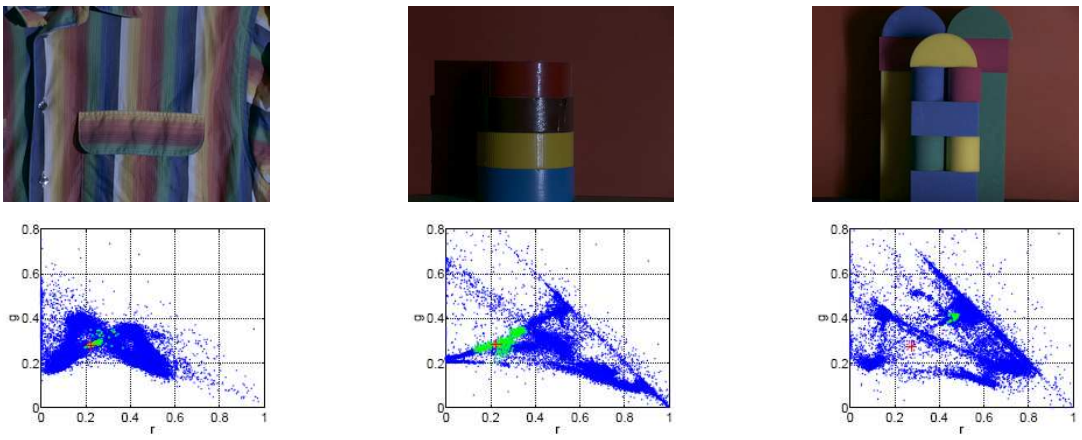
4.3 White Patch Gamut Mapping

The foundational color constancy method, the so-called White Patch method, usually estimates the illuminant by finding a white patch in the scene which has maximum response in all channels. Experiments show that illuminant information is more informative in bright regions than darker region [167, 82]. Let us extend the idea of the white patch by assuming that there is always some subset of any of the following: a white patch, highlights or light source in the image. We also use the *gamut* of bright pixels, in contradistinction to White-Patch, which usually deals only with the brightest pixel in the image. Obviously, using a single pixel can be noisy and is not robust.

Since we are dealing with bright pixels we need to be careful about clipped pixels, which are pixels where the light reflection exceeds the dynamic range of the camera. We remove pixels which exceed 90% of the dynamic range for each channel. Then we simply define bright pixels as the top $T\%$ of the sum of channels $R + G + B$. Here, T could for example be 1, 2 or 5%.

To investigate the utility of this assumption, we carry a simple experiment to check whether the actual illuminant colour falls inside the 2D gamut of bright pixels or not. We find that the actual illuminant colour falls in the 2D gamut of the top 5% brightness pixels of each image for SFU Laboratory Dataset [14] for 88.16% of images, and in 66.02% of images for the GreyBall Dataset [40]. Fig. 4.1 shows 2D gamut in chromaticity space $\{r, g\} = \{R, G\}/(R + G + B)$, with the top 5% brightness pixels in green. The actual measured illuminant is shown as a red star. Note that with no supporting evidence (Fig. 4.1(c)) the illuminant may not fall in the bright region.

Motivated by this experiment, we define the White Patch Gamut, consisting of the set of RGBs in an image for a white patch, highlights, or light source. As well, we define the canonical White Patch Gamut, denoted $\mathcal{C}_w$, comprising the limited set of RGBs that can occur for a white patch, highlights, or light source under a reference illuminant. This can be trained by observing images under known light sources. I.e., we simply identify them in a training phase by collecting the top-brightness pixels which fall near actual values of the illuminant colour, projected onto 2D chromaticity space. Similarly, the input White Patch Gamut $\mathcal{I}_w$ is defined as the gamut of the top-brightness



(a) (b) (c)

Figure 4.1: Examples of image evidence: top 5% bright pixels in green, other pixels in blue, and red star showing the illuminant point in r, g chromaticity space. (a) Image with white patch; (b) Image with specularity; (c) Image without white patch or specularity.

pixels for an input image.

$$\mathcal{C}_w(\rho): \begin{array}{c} a_1^w R + b_1^w G + c_1^w B \geq e_1^w \\ a_2^w R + b_2^w G + c_2^w B \geq e_2^w \\ \vdots \\ a_N^w R + b_N^w G + c_N^w B \geq e_N^w \end{array} \tag{4.5}$$

We can again define A_w and e_w in the same fashion as A and e. Using the principle of gamut mapping set out in §4.2, we can determine the set of feasible mappings $\mathcal{M}_w$, meaning all possible mappings that can transfer the input White Patch Gamut to lie completely within the canonical White Patch Gamut.

$$\mathcal{M}_w : \forall q_i \in \mathcal{I}_w : A_w \; diag(q_i) \; d \geq e_w \tag{4.6}$$

These constraints are only in operation if there is a white patch or specularity in the image. However our experiments show that, if they can indeed be brought into play, these new constraints are even more powerful than the original gamut mapping constraints.

4.3.1 Generating the White Patch Gamut

When there are no strong highlights, light sources, or white surfaces in the image, the bright pixels are not helpful; in such a case they can stem from areas of images belonging to the brightest surface,

which tends to be some particular surface colour or possibly a bag of single pixels from all over the image. In the former case we can simply check if these pixels are in the possible chromaticity gamut of illuminants (this gamut is computed in the training phase) and in the latter we can investigate the distribution of these pixels in chromaticity space. The standard deviation of these pixels in RGB space is a good measure for this purpose. In either case, there is no White Patch Gamut for such an image. Thus, we cannot necessarily define a White Patch Gamut for all images. On the other hand, in order to omit outliers we remove pixels with distance to the mean greater than 2 standard deviations. Fig. 4.2 shows (red) the 2D gamut and the white patch gamut for example images, compared to their canonical gamuts (blue) in chromaticity space.

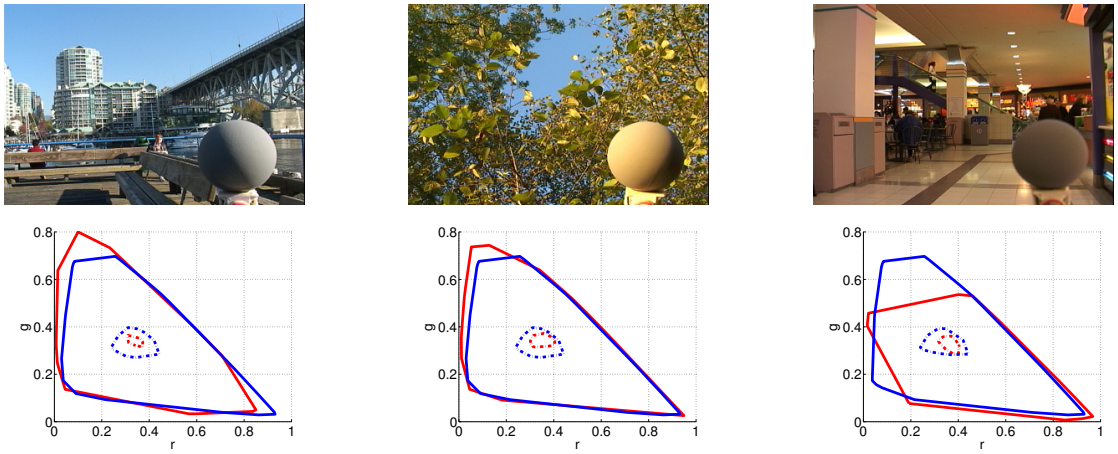


Figure 4.2: Examples of 2D gamut (solid red) and white patch gamut (dotted red) compared to canonical gamut (solid blue) and canonical white patch gamut (dotted blue) in rg chromaticity space.

Fig. 4.3 shows the 3D canonical gamut for the SFU Laboratory dataset and its 3D canonical White Patch Gamut in RGB space. As we expected the White Patch Gamut is considerably smaller.

4.3.2 Combination Method

Until now, we have two sets of feasible mappings: $\mathcal{M}$, which is based on standard gamut mapping and $\mathcal{M}_w$, which is based on white patch gamut mapping. These sets can be used in combination instead of selecting only one mapping per algorithm. Since each feasible set represents all illuminant estimates that are considered possible, an obvious combination is to consider only those estimates that are all feasible sets, which is intersection of these two sets ($\mathcal{M} \cap \mathcal{M}_w$). Another approach

to combine these is to consider all possible estimates in both feasible sets, meaning the union of these two sets ($\mathcal{M} \cup \mathcal{M}_w$). Experiment shows that the intersection approach outperforms the union approach. For the intersection based combination, the optimization becomes the following convex programming problem:

$$\begin{aligned} Maximize \quad & d_1 + d_2 + d_3 \\ \text{subject to} \quad & A\ diag(q)\ d \geq e\,,\ q \in I \\ & A_w\ diag(q)\ d \geq e_w\,,\ q \in I_w \end{aligned} \tag{4.7}$$

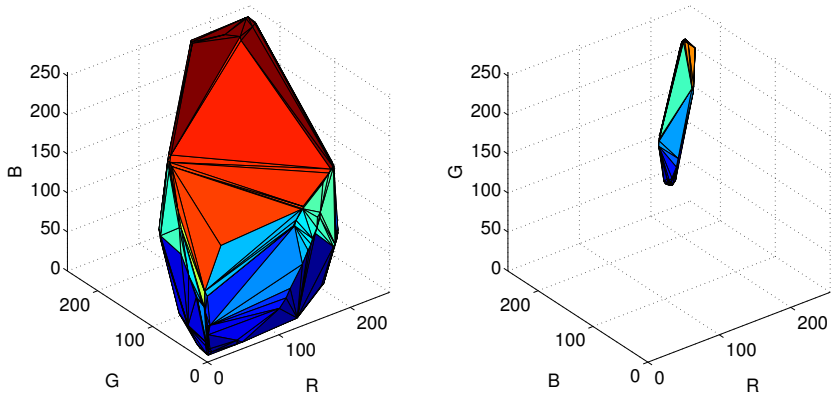


Figure 4.3: (a) 3D canonical gamut (b) 3D canonical white patch gamut for SFU Laboratory dataset in RGB colour.

4.4 Experimental Results

First, consider the Barnard colour constancy dataset [14], which we call the SFU Laboratory dataset; this contains 321 measured images under 11 different measured illuminants. The scenes are divided into two sets as follows: minimal specularities (223 images), and non-negligible dielectric specularities (98 images).

For a more real-world (out of the laboratory) image experiment we also used the Ciurea and Funt [40] GreyBall dataset, which contains 11346 images extracted from video recorded under a wide variety of imaging conditions. The images are divided into 15 different clips taken at different

Table 4.1: Comparison of White Patch Gamut Mapping (WP) with well known colour constancy methods.

Dataset	SFU Laboratory		GreyBall	
Methods	Median	Mean	Median	Mean
White Patch	6.5°	9.1°	5.3°	6.8°
Grey World	7.0°	9.8°	7.0°	7.9°
Grey Edge	3.2°	5.6°	4.7°	5.9°
Gamut Mapping	2.3°	3.9°	5.5°	7.0°
Gamut Mapping 1-jet	2.1°	3.9°	5.5°	6.8°
Natural Image Stat.	-	-	**3.9**°	**5.2**°
WP Gamut Mapping	**1.9**°	**3.4**°	4.7°	6.0°

locations. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of the image. This grey sphere must be masked out during experiments.

We use Matlab code provided by [102] for general gamut mapping, which is based on convex programming. For the SFU Laboratory dataset, 31 images (all images recorded under the *syl-50MR16Q*-illuminant) were used for computation of the canonical gamut, and these were omitted from the test set. For the GreyBall Dataset, the canonical gamuts were computed using 14 out of 15 categories and we tested on the remaining category. We repeated the experiment 15 times. We used the top 5% of the sum of channels to generate the bright gamut.

Table 4.1 gives the accuracy of the proposed methods for the SFU Laboratory dataset [14] as well as the GreyBall dataset [40], in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. For those methods which need tunable parameters, we utilize optimal parameters for each dataset. (In these experiments, we actually accomplished slightly better estimation for pixel based gamut mapping than other reported results for the GreyBall dataset since we mask image margins because of their low quality.)

This experiment shows that white patch gamut mapping improves the performance of gamut mapping by 13% to 18% in terms of angular mean and median error; these results show that the proposed method is in fact the best extension of gamut mapping compared to other extensions such as generalized edge-based gamut mapping [102] (1st-jet is their best) or the 2D version of gamut mapping [59]. To our knowledge, for the SFU Laboratory dataset white patch gamut mapping does best in terms of both mean and median angular error compared to any reported colour constancy method, even very complex ones. For the GreyBall dataset, for which gamut mapping generally

does not work very well (since the images are not gamma corrected), white patch gamut mapping outperforms gamut mapping by 15%; it does better than or equal to all the unsupervised colour constancy methods, but does not surpass those which combine or select other colour constancy methods, such as Natural Image Statistics [100].

4.5 Conclusion

In this chapter, we define the White Patch Gamut as a new extension to the gamut mapping colour constancy method, consisting of the bright pixels of the image. Our experiments show that adding new constraints based on the White Patch Gamut to standard gamut mapping constraints outperforms gamut mapping methods as well as other well-known colour constancy methods. These new constraints are therefore a powerful addition to the field and in fact can be more discriminating than the original set of constraints themselves.

White patch gamut mapping improves the performance of gamut mapping by 13% to 18% in terms of angular mean and median error over the two well-known standard datasets that we tested, thus comprising is the best extension of gamut mapping colour constancy to date.

Chapter 5

The Zeta Image

In this chapter we present a new and effective physics-based colour constancy algorithm which makes use of a novel log-relative-chromaticity planar constraint. We call the new feature the Zeta-image. We show that this new feature is tied to a novel application of the Kullback-Leibler Divergence, here applied to chromaticity values instead of probabilities. The new method requires no training data or tunable parameters. Moreover it is simple to implement and very fast. Our experimental results across datasets of real images show the proposed method significantly outperforms other unsupervised methods while its estimation accuracy is comparable with more complex, supervised, methods. As well, we show that the new planar constraint can be used as a post-processing stage for any candidate colour constancy method in order to improve its accuracy. We show as well that it can be used to identify and remove specularities. Its application in this chapter demonstrates its utility, delivering state of the art performance.

5.1 Introduction

Identification of illumination is an important problem in image processing for digital cameras, for both still images and video, and can form a prerequisite for many computer vision applications. In a scene consisting of dielectric materials (e.g., plastics, and indeed most non-metals) there is typically substantive specular content. This does not necessarily mean extremely bright mirror-like reflection, but can consist for example of the glint of light reflected from blades of grass, or the sheen of light reflected from a desk surface. For non-metals, this very common specular content is an important indicator of the colour of the lighting in a scene, and hence has substantial bearing on determination of a correct white balance for camera images.

Many colour constancy algorithms have been proposed (see [110, 104] for an overview). The foundational colour constancy method, the so-called White-Patch or Max-RGB method, estimates the light source colour from the maximum response of the different colour channels [117]. Another well-known colour constancy method is based on the Grey-World hypothesis [32], which assumes that the average reflectance in the scene is achromatic. Grey-Edge is a recent version of the Grey-World hypothesis that says: the average of the reflectance differences in a scene is achromatic [180]. Finlayson and Trezzi [75] formalize grey-based methods by subsuming them into a single formula using the Minkowski p-norm.

The Gamut Mapping algorithm, a more complex and more accurate algorithm, was introduced by Forsyth [79]. It is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours. Several extensions have been proposed [8, 67, 59, 102].

Lee [120] proposed a method which uses specularity to compute illumination by using the fact that in the CIE chromaticity diagram [194] the coordinates of the colours from different points from the same surface will fall on a straight line connected to the specular point. This is the case when the light reflected from a uniform surface is an additive mixture of the specular component and the diffuse component. This seminal work initiated a substantial body of work on identifying specular pixels and using these to attempt to discover the illuminant [122, 165]. Although these works are theoretically strong, none of them report performance over real world datasets of images with and without specularities.

Gijsenij et al. [104] propose the following desiderata in their survey on colour constancy methods in assessing computational methods: (1) the requirement for training data; (2) the accuracy of the estimation; (3) the computational runtime of the method; (4) transparency of the approach; (5) complexity of the implementation; (6) number of tunable parameters.

In this chapter, we set out a new discovery, consisting of a planar constraint that must be obeyed, in a certain standard model of reflectance, by specular or near-specular pixels in an image. The new feature involved we call the Zeta-image[1], and below we show that this feature is tied to the information-theoretic concept of applying one distribution to generate bitstring codes for another; here we view chromaticity components, which add to 1, in the role of probabilities. We present a novel physics-based colour constancy algorithm based on a log-relative-chromaticity planar constraint (LRCP-Constraint), which requires no training data or tunable parameters. It is easy to implement and very fast compared to more complex colour constancy methods such as gamut mapping.

[1] Patent applied for.

Our experimental results over three large datasets of both laboratory and real world images show that the proposed method significantly outperforms other unsupervised methods while its accuracy of estimation is comparable with more complex methods that need training data and tunable parameters.

The chapter is organized as follows. In §5.2, we describe the fundamental process of the log-relative-chromaticity planar constraint and define the Zeta-Image which we define in this chapter. In §5.3, we explicate aspects of the new planar constraint by making use of synthetic data. §5.4 contains two proposed colour constancy algorithms based on the LRCP-Constraint. We compare our proposed algorithms with previous colour constancy algorithms over three standard datasets in §5.5, demonstrating the efficacy of the new insight for real images. §5.6 examines the role of the Zeta-image in manipulating specularities, and §5.7 concludes the chapter.

5.2 Relative Chromaticity Near Specular Point

5.2.1 Image Formation Model and Relative Chromaticity

Let the RGB 3-vector for the light itself as seen by the camera be e_k, and let the 3-vector for the reflectance at a pixel as seen by the camera, under equi-energy white light, be s_k. Now in a product-of-vectors simple model [29] we approximately have the matte ("body", i.e., non-specular) RGB value at that pixel equal to

$$R_k \simeq \frac{\sigma s_k e_k}{q_k} \tag{5.1}$$

where σ is shading. In the standard Lambertian model for matte shading, σ equals lighting-direction dotted into surface normal. Here, q_k is a triple giving the overall (integrated) camera sensor strength [47].

If we also consider an additional specular component, this equation becomes

$$R_k \simeq \frac{\sigma s_k e_k}{q_k} + \beta e_k \tag{5.2}$$

where β represents the amount of specular component at that pixel. The value of β for a pixel will depend upon the lighting direction, the surface normal, and the viewing geometry [157]. Here, the specular component βe_k is simply assumed to be the same colour as the light itself, in a Neutral Interface Model [120] for dielectrics. For purposes of discovering properties of this equation, let us assume for the time being that β is simply a constant — in actuality it will be a scalar property of each pixel and this issue is further discussed below in §5.2.4. Let us lump values $\sigma s_k/q_k$ into a

single quantity and for convenience call this simply s_k. Now we have

$$R_k \;=\; s_k e_k \;+\; \beta e_k \tag{5.3}$$

The chromaticity [194] ρ_k is colour without magnitude, in an L_1 norm: $\boldsymbol{\rho} = \{R, G, B\}/(R + G + B)$, so here we have

$$\rho_k \;=\; \frac{s_k e_k \;+\; \beta e_k}{\sum_{j=1}^{3}(s_j e_j \;+\; \beta e_j)} \tag{5.4}$$

Let the chromaticity of the light itself be denoted

$$\rho_k^e \;=\; e_k / \sum_j e_j \tag{5.5}$$

Now here we wish to examine the properties of the Relative Chromaticity, which we define to be the chromaticity divided by the chromaticity of the light, ρ_k^e. Let us call this quotient chromaticity χ_k, so that

$$\chi_k \;=\; \frac{\rho_k}{\rho_k^e} \;=\; \frac{s_k e_k \;+\; \beta e_k}{\sum_{j=1}^{3}(s_j e_j) \;+\; \beta \sum_{j=1}^{3} e_j} \cdot \frac{\sum_{j=1}^{3} e_j}{e_k} \tag{5.6}$$

where all divisions are taken to be component-wise. Dividing by the light chromaticity is the main innovative step in this chapter: it is an *ansatz* that we claim will bear fruit by generating a constraint on the illuminant colour.

For convenience, let $E \equiv \sum_{j=1}^{3} e_j = |\boldsymbol{e}|$ where $|\cdot|$ is the L_1 norm. Then we arrive at

$$\chi_k \;=\; \frac{s_k + \beta}{\frac{(\sum_j s_j e_j)}{E} + \beta} \tag{5.7}$$

So, for a pixel with no matte component s_k but only a purely specular component, we would have $\chi_k \equiv 1$ for all 3-vector elements $k = 1..3$.

5.2.2 Log-Relative-Chromaticity and Planar Constraint

Next we show that in fact log values are preferable, in that a simple planarity constraint falls out of the formulation once we move to the log domain.

Let us define a new quantity, $\boldsymbol{\psi}$, which is the logarithm of the ratio $\boldsymbol{\chi}$ defined above: we call this the Log-Relative-Chromaticity:

$$\psi_k \;=\; \log(\chi_k) \;=\; \log\left(\frac{\rho_k}{\rho_k^e}\right) \tag{5.8}$$

Now *near* a specular point, we can take the limit as $(1/\beta) \rightarrow 0$. Let $\alpha = 1/\beta$. Then in the limit as specularity increases, $\boldsymbol{\psi}$ goes to

$$\psi_k \;\; = \;\; \lim_{\alpha \rightarrow 0} \; \log \left\{ (\alpha s_k + 1) \, / \left(\alpha \textstyle\sum_j (s_j e_j)/E + 1 \right) \right\}$$

Using a Maclaurin series,

$$\psi_k = \alpha \left(s_k - \tfrac{\sum_j s_j e_j}{E} \right) + O\left(\alpha^2\right) \tag{5.9}$$

Omitting $O\left(\alpha^2\right)$, we note by inspection that the quantity ψ_k is *orthogonal to the illuminant vector*:

$$\sum_k \psi_k \, e_k \;\equiv\; 0 \;, \text{ so also } \sum_k \psi_k \rho_k^e \;=\; 0 \tag{5.10}$$

Therefore we have a planar constraint on image pixels that are near-specular:

Planar constraint: *For near-specular pixels, Log-Relative-Chromaticity values are orthogonal to the light chromaticity.*

Note that in eq. (5.9) above we have expressed this orthogonality in a different way than the usual, Euclidean-norm based calculation of the part of the vector $\boldsymbol{s}$ that is orthogonal to vector $\boldsymbol{e}$, *viz.* $(s_k - \boldsymbol{s} \cdot \hat{\boldsymbol{e}} \, \hat{e}_k)$ for normalized light vector $\hat{\boldsymbol{e}}$.
Nevertheless, it is easy to verify that eq. (5.10) does indeed hold. The meaning of eq. (5.9) is that we are forming an L_1-based projection onto the plane orthogonal to the light.

Here we define the **Zeta-image** ζ as the dot-product of the log-relative-chromaticity $\boldsymbol{\psi}$, eq. (5.11), with a putative light direction:

$$\zeta \;=\; -\boldsymbol{\psi} \cdot \boldsymbol{\rho}^{\,e} \;=\; -\sum_{k=1}^{3} \log(\boldsymbol{\rho}_{\,k} / \boldsymbol{\rho}_{\,k}^{\,e}) \cdot \boldsymbol{\rho}_{\,k}^{\,e} \tag{5.11}$$

Based on the planar constraint the Zeta-image value is zero for near-specular pixels; and as well we will show that it is a non-negative value. We also know that the zeta-image is invariant to shading since it uses chromaticity instead of the RGB value for pixels. Note that the zeta value might be too noisy for the case of dark pixels since we take the log. In an extreme case when we have zero for any of channels, the zeta value will be ∞. Therefore we do not rely at all on the value of zeta for dark pixels.

In the following sections we will see the application of Zeta-image for illumination estimation as well as specularity manipulation.

5.2.3 Relative Chromaticity Near Specular Point for Incorrect Candidate Illuminant

Having enunciated a plane constraint governing relative chromaticity formed by dividing by the actual light chromaticity, we also investigate the effect of dividing by an incorrect light, one that was not in fact involved in forming the image at hand.

Suppose the correct illumination chromaticity vector is ρ^e but instead we have supposed that it is $\rho^{e'}$. Then in carrying out our division by the putative illuminant chromaticity, we have instead

$$\begin{aligned} \chi'_k &= \rho_k/\rho_k^{e'} = (s_k e_k \ + \ \beta e_k) \sum_j (s_j e_j \ + \ \beta e_j) \cdot \frac{\sum_j e'_j}{e'_k} \\ &= \frac{\alpha s_k + 1}{\alpha \frac{\boldsymbol{s} \cdot \boldsymbol{e}}{E} + 1} \left(\frac{e_k}{e'_k}\right) \left(\frac{E'}{E}\right) \end{aligned} \tag{5.12}$$

so that

$$\lim_{\alpha \to 0} \ \psi'_k = \log\left(\frac{e_k}{e'_k}\frac{E'}{E}\right) + \alpha\left(s_k - \frac{\boldsymbol{s} \ \cdot \boldsymbol{e}}{E}\right) \ + \ O(\alpha^2) \simeq \psi_k^0 \ + \ \log\left(\frac{\rho_k^e}{\rho_k^{e'}}\right) \tag{5.13}$$

where ψ_k^0 is the set of planar values found for the *correct* light, in eq. (5.9).

For this new version of $\boldsymbol{\psi}$, now for the set of image chromaticities relative to the putative light chromaticity, with the log taken, the gamut of values is shifted by the last term on the right hand side of the second line of eq. (5.13).

Comparing eqs.(5.10) and (5.13), we see that the idea of dividing by the illuminant chromaticity produces a gamut for specular pixels, for the case of an incorrect choice of light chromaticity, which is shifted with respect to the original gamut. This property is potentially of value for determining the actual value of the light chromaticity in a scene.

5.2.4 Varying specular factor β

The fact that specular scalar factor β is not a constant makes no difference to the argument: for near-specular pixels any value of β still leads to quantity ψ_k lying in the plane orthogonal to the illuminant, and that plane being shifted as in eq. (5.13) for the incorrect choice of illuminant.

5.3 Illustrations of Plane Constraint

5.3.1 Synthetic Example of Specular Plane

Using synthetic data as a guide, we can investigate the planar constraint.

Suppose s_k is uniform-randomly distributed. Then $\boldsymbol{s}$ uniformly occupies the unit cube in RGB colour space. Let us form $\boldsymbol{\psi}$ values that are consistent with the near-to-specular model (5.9): these values are on a plane orthogonal to the light 3-vector e_k that was utilized to form the gamut.

Fig. 5.1 shows the gamut for $\boldsymbol{\psi}$ in blue, when we divide by the correct illuminant chromaticity, and the shifted gamut $\boldsymbol{\psi}'$ when we have divided by an incorrect chromaticity. The blue plane goes through the origin — the $\boldsymbol{\psi}$ vector for the light itself is $\mathbf{0}$ — and $\boldsymbol{\psi}$ vectors that are on the blue plane are orthogonal to the lighting direction $\boldsymbol{\rho}^e$, shown as a green arrow.

However, for the shifted plane, shown in red, the center of the plane is not at the origin, so $\boldsymbol{\psi}'$ vectors on that plane, such as the cyan arrow shown on the red plane, are actually 3-vectors in the $\boldsymbol{\psi}$ space (shown in black) that are not orthogonal to the green lighting vector. That is, the calculation (5.10) proceeds using vector components which are defined relative to the fixed origin position $\mathbf{0}$, not some point such as the middle of the red plane.

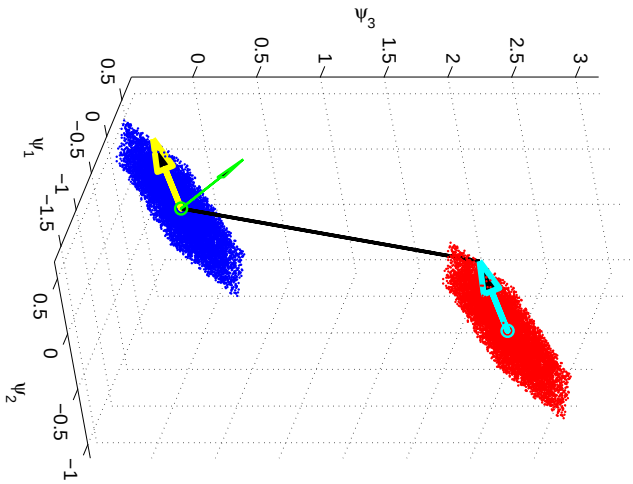


Figure 5.1: Gamuts for Log-Relative-Chromaticity values $\boldsymbol{\psi}$ generated using the correct illuminant chromaticity ρ_k^e, in blue, and generated using an incorrect candidate illuminant chromaticity $\rho_k^{e'}$, in red.

5.3.2 Synthetic Example of Matte plus Specular Contributions

A complete synthetic example consists of accurately modeled matte plus specular components. Here, let us consider a test image consisting of three shaded spheres (as in [60]), with surface colours

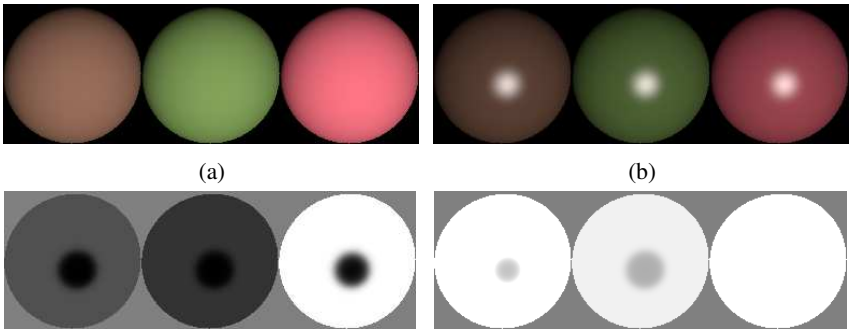


Figure 5.2: (a): Ground-truth synthetic image of Lambertian surfaces under standard illuminant D65. (b): Image with specular reflection added. (c): Dot products eq. (5.10) of Log-Relative-Chromaticity $\boldsymbol{\psi}$ formed using the correct scene illuminant: specular pixels are correctly determined (low values of dot-product). (d): Dot product where now an incorrect light is utilized in generating $\boldsymbol{\psi}$ values: specular pixels are poorly recognized. The ground truth chromaticity for this light is $[.279, .316, .404]$. As a different, incorrect light, we divide by the very different illuminant chromaticity $[.172, .363, .464]$.

equal to patches 1, 4, and 9 of the Macbeth ColourChecker [134] (dark skin, moderate olive green, moderate red), and under standard illuminant D65 (standard daylight with correlated colour temperature 6500K [194]) using the sensor curves for a Kodak DCS420 digital colour camera. If we adopt a Lambertian model then the matte image is as in Fig. 5.2(a). We now add a specular reflectance lobe for each surface reflectance function. We use the Phong illumination model[78], together with underlying matte Lambertian shading. Here, we use a Phong factor of 1 for the magnitude relative to matte. For the Phong power, we use a power of 20, where the inverse is basically roughness, 0.05. The matte image goes over to one with highlights as in Fig. 5.2(b). These surfaces are particular instances of dichromatic surface spectral reflectances [157].

Now suppose we have correctly guessed the lighting chromaticity; then let us go on to form the Log-Relative-Chromaticity $\boldsymbol{\psi}$. Forming the dot-product (5.10) we arrive at values as depicted in Fig. 5.2(c). The $\boldsymbol{\psi}$ values for low dot product values are found to indeed lie nearly on a plane, and the idea of asking that the dot-product (5.10) be small is indeed justified for this synthetic data. Moreover, the perpendicular to the plane found is very close to the actual light used to form the image.

To show the discriminative power of the planar constraint, let us now suppose we have guessed an incorrect light chromaticity. The ground truth chromaticity for this light is $[.279, .316, .404]$. As a

different, incorrect light, we divide by the very different illuminant chromaticity $[.172, .363, 0.464]$. If we now instead divide by the chromaticity of the wrong light, the dot-product appears as in Fig. 5.2(d), a much less discriminative identification of specular pixels than when we in fact use the correct light. Moreover, the perpendicular to the best-fit plane is much poorer an estimate of the correct light chromaticity. That is, suppose we identify as candidate specular pixels those whose absolute value of dot-product eq. (5.10) with the candidate light chromaticity falls in the lowest 10-percentile, say (i.e., nearest to zero). Forming a Singular Value Decomposition of $\boldsymbol{\psi}$ values for those pixels determines the best-fit plane. The third eigenvector is then associated with the normal to that plane, and within the model presented here that normal should be close to the illuminant chromaticity in direction.

5.4 Planar Constraint Method

Here we begin construction of an algorithm by considering first a simple search method as motivation, and then stating an analytic solution.

5.4.1 Global Search

The planar constraint suggests that the dot product for near-light-colour (e.g., specular) pixels is minimized for the correct illuminant. This points to a useful descriptor for finding the specular point.

Suppose we were to assume that for any candidate illuminant the lowest 10-percentile, say, of dot-product values (5.10) could be near-specular pixels. Now, to find the correct illuminant, we need to minimize dot-product values (5.10) over candidate illuminants for those lowest 10-percentile pixels. Thus an optimization can be stated as follows:
We already defined the Zeta-image ζ as the dot-product of the log-relative-chromaticity, $\boldsymbol{\psi}$, with a putative light direction in eq. (5.11).

$$\begin{aligned} Minimize: \qquad & \min_{\boldsymbol{\rho}^{\,e}} \ \mathcal{I} = \sum_{\boldsymbol{\psi} \in \Psi_0} |\zeta| \\ \text{subject to} \quad & \textstyle\sum_{k=1}^{3} \rho_k^e = 1 \,,\; 0 < \rho_k^e < 1 \,,\; k = 1..3 \end{aligned} \tag{5.14}$$

where Ψ_0 is the set of pixel dot-product values (5.10) with the candidate illuminant chromaticity $\boldsymbol{\rho}^{\,e}$ that are in the lowest 10-percentile.

The meaning of eq. (5.14) is that we first carry out a search, over possible illuminant chromaticities $\boldsymbol{\rho}^{\,e}$. This can be phrased as either an optimization-based approach or, as here, a simple

hierarchical grid search. Then we adopt a heuristic that says that the lowest 10-percentile of values of dot-products with the candidate illuminant could be specular or in general illuminant-coloured. For these pixels we calculate the sum of absolute values of dot-products and take as the best candidate light that which delivers the minimum sum.

Fig. 5.3(a) shows an input image, and Fig. 5.3(b) shows a boolean map of the lowest 10-percentile of dot-product values (5.10) with the correct illuminant chromaticity of that image. In contrast, if we show the lowest 10-percentile of dot products with the chromaticity of an incorrect light, where ψ values are constructed using that incorrect light, the boolean map identifying putative specular/illuminant-coloured pixels is as displayed in Fig. 5.3(c). We see that using the correct light produces a much more plausible map of possible pixels that will help identify the light.

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

Figure 5.3: (a): Image taken under measured illuminant. (b): Light-coloured pixels identified using planar constraint, when correct illuminant is chosen. (c): Putative illuminant-coloured pixels when incorrect illuminant is used. (d): Analytic Zeta-image (float, reversed in intensity).

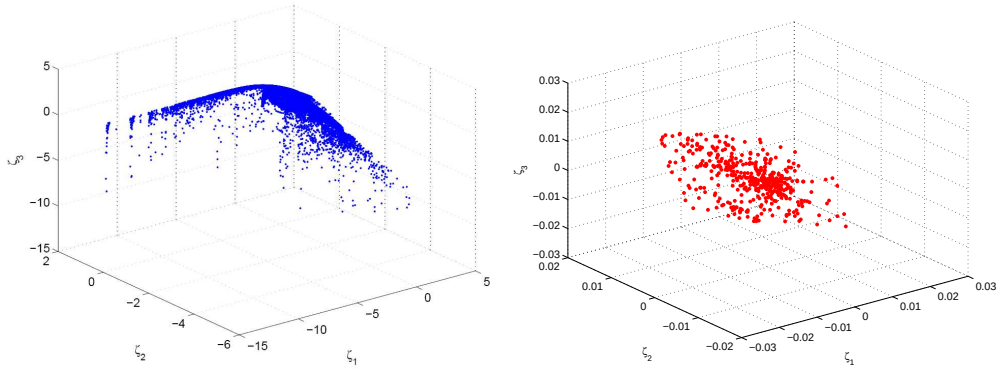
Fig. 5.4(a) shows Fig. 5.3(a) in ψ space computed by eq. (5.8). There is no evidence of a plane while, in contrast, Fig 5.4(b) demonstrates that ψ values of near-specular pixels based on the optimization eq. (5.14) do indeed form a plane orthogonal to the light vector, as can be easily verified numerically.

For visualization, in Fig. 5.5 we display as well a mesh plot of the objective function of eq. (5.14) for Fig. 5.3(a). This function has only one local minimum, which makes it easy to compute.

The float-valued Zeta-image is displayed in Fig. 5.3(d). We show next that we can directly use the Zeta-image to analytically find the correct illuminant chromaticity.

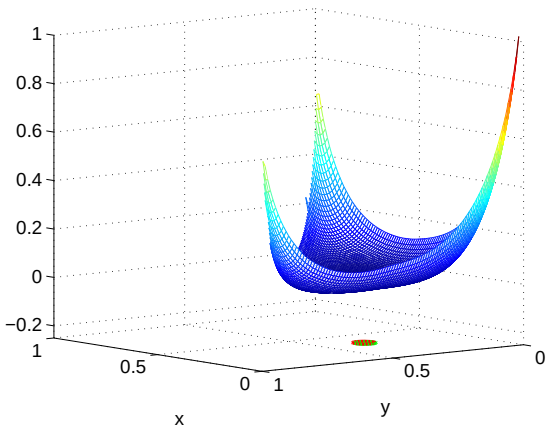
5.4.2 Analytic Solution

Having motivated the method, we now state an analytic solution that in fact produces excellent results and is very simple and fast. Suppose we identify a possible set Ψ_0 of specular pixels by any



(a) (b)

Figure 5.4: (a): Fig. 5.3(a) in ψ space computed by eq. (5.8). (b): ψ values of near-specular pixels form a plane.



(a)

Figure 5.5: A mesh plot of the objective function in eq. (5.14), for image Fig. 5.3(a) The red ring shows the estimated illuminant, which is the index achieving the minimum value of the objective function; the green ring (lying almost on top of the red ring) shows the correct illuminant chromaticity location. The angular error between these two values is 1.09°.

convenient method — e.g., we could simply take the top 5% of brightness.

Let the number of bright pixels be N. Then our analytic solution is as follows:

Theorem: Up to an L_1 normalization, **the light colour is given by the geometric mean**

$$\rho_k^e \;=\; g(\boldsymbol{\rho}_{\,k}) \;\equiv\; \left(\prod_{i=1}^{N} \rho_k^i\right)^{1/N} \quad , \; k = 1..3 \tag{5.15}$$

To prove this, let us first solve an auxiliary optimization replacing (5.14). For physical lights, we expect ζ from eq. (5.11) to be non-negative, so [with no absolute-value bars as in $\mathcal{I}$ in eq. (5.14)] we first solve

$$\min_{\boldsymbol{\rho}^{\,e}} \sum_i \zeta_i \;+\; \lambda\,(\sum_{k=1}^{3} \rho_k^e \;-\; 1) \tag{5.16}$$

where $i = 1..N$ ranges over the N pixels in $\boldsymbol{\psi} \;\in \Psi_0$, and λ is a Lagrange multiplier enforcing that $\boldsymbol{\rho}^{\,e}$ is a chromaticity. Taking partial derivatives with respect to ρ_k^e we have the normal equations

$$-\sum_i \left[\left(\log(\rho_k^i - \log \rho_k^e\right) - 1\right] + \lambda = 0 \tag{5.17}$$

$$\text{with solution } \rho_k^e \;=\; g(\boldsymbol{\rho}_{\,k}) \;\cdot\; \exp(-(N+\lambda)/N) \tag{5.18}$$

Also, taking the derivative of (5.16) with respect to λ enforces $\sum_{k=1}^{3} \rho_k^e \;=\; 1$, so that scaling the solution (5.18) to obey L_1-norm=1 solves for λ as well if we like.

The meaning of (5.16) then is: **The planarity constraint yields the geometric mean of the chromaticities as the solution for the light**, up to trivial scaling of the L_1-norm.

However so far we have omitted absolute value bars, with a full optimization minimizing $\mathcal{I} = \sum_i |\zeta_i|$. We now observe that the form (5.11) formally has the structure of the Kullback-Leibler Divergence from information theory, in that chromaticities for image, ρ_k^i, and light, ρ_k^e all add to unity: $\sum_{k=1}^{3} \rho_k \;=\; 1$. We are minimizing $\sum_{i=1}^{N}[\sum_{k=1}^{3} -\rho_k^e \,\log(\rho_k^i/\rho_k^e)]$, which has the K-L structure except that instead of summing over probabilities for symbols we are here summing over colour-channels. Thus each dot product (the sum over k at each pixel) is necessarily nonnegative since it represents the extra bits required to code samples from ρ_k^e when using a code based on ρ_k^i. Hence we can simply consider the minimization (5.16), with solution (5.15) up to scaling. As a final step, we calculate a final value for ρ_k^e by trimming pixels to the least-10% values of the Zeta-image ζ and recalculating the geometric mean (5.15). Thus we propose an iterative-analytic solution algorithm by an iterative method which finds the illuminant using an initial specular candidate set and then iteratively the zeta value is calculated using the estimated illuminant and also the illuminant using

the geometric mean of pixels with the least-10% values of the zeta. The initial specular candidate is set selected via top-$T\%$ brightness pixels for different T values and the final estimates is the one that provides the minimum value of zeta after two or three iterations. The detailed algorithm is expressed in algorithm 5.1.

Algorithm 5.1 Zeta Analytic Solution for Illumination Estimation

1: **for all** threshold in $\{5\%, 3\%, 2\%, 1\%, .5\%\}$ **do**
2: Find top bright pixels by that threshold on $R + G + B$ value as specular candidate set
3: Initially estimate the specular point (e) by geomean of pixels in specular candidate set
4: Calculate ζ by e using eq. (5.11)
5: Recalculate e by geomean of pixels with 10% low ζ
6: Iterate to step 4 twice
7: **end for**
8: Choose the estimate by threshold that minimizes average ζ for final set of specular candidate set

Areas of images that are specular tend to be bright. According to our theory the geometric mean of pixels in these bright (generally, specular) regions is the optimal estimate for the illuminant. Our insight is in contradistinction to the work of Choudhury and Medioni [39] and Funt and Shi [91] which proposed finding the max after a local mean calculation (e.g. after local blurring). In the presence of specular highlights illuminant estimation is the mean of the max but not the converse. Indeed, using the correct ordering is crucial (a fact borne out by our experiments reported below). We have also try a grid based local search around the result of algorithm 5.1 to make it more precise. Experiment shows that it improve performance slightly while make it much slower so we decide not to use it.

5.4.3 2nd Algorithm: Planar Constraint Applied as a Post-Processing Step

Suppose we have an estimate $\boldsymbol{\rho}^{e\star}$ of the correct illuminant, from any colour constancy algorithm. If our estimate is indeed near the correct illuminant we can then identify as near-specular pixels those whose absolute value of dot-product eq. (5.10) with the candidate light chromaticity falls in the lowest 10-percentile, say, (i.e., nearest to zero). Forming a Singular Value Decomposition of $\boldsymbol{\psi}$ values for those pixels determines the best-fit plane, with the third eigenvector associated with the normal to that plane. Within the model presented here that normal should be close to the illuminant chromaticity in direction. Because of the additional evidence brought to bear by eq. (5.10) we expect the estimate to improve. If instead the illuminant estimate is wrong, then we have found that the above SVD step will almost always not change it much and no harm is done by carrying out this

post-processing step.

We carry out

$$\psi_k \;=\; \log\left(\frac{\rho_k}{\rho_k^{e*}}\right)\;,\; k=1..3; \quad \Psi = \boldsymbol{\psi}^{\,T}\boldsymbol{\rho}^{\,e\star}\;;$$

$$\boldsymbol{\psi}\,(\Psi_0) = SVD(\boldsymbol{\psi}\,(\Psi_0)) = U\,diag(\boldsymbol{d}\,)\,V^T\;; \tag{5.19}$$

$$\boldsymbol{\rho}^{\,e} = \boldsymbol{V}_{\,3}/|\boldsymbol{V}_{\,3}|\;; \quad \text{success} = (d_3 \text{ small})\&(\boldsymbol{\rho}^{\,e} \simeq \boldsymbol{\rho}^{\,e\star})$$

where $\boldsymbol{\rho}^{\,e\star}$ is the estimate of the illumination provided by a colour constancy method, Ψ_0 is the lowest 10-percentile of Ψ, and $\boldsymbol{\rho}^{\,e}$ is the estimate of the illumination based on the planar constraint.

The meaning of eq. (5.19) is that, for any estimate $\boldsymbol{\rho}^{\,e\star}$ of the light chromaticity $\boldsymbol{\rho}^{\,e}$, if the model (5.10) is obeyed around the light point then SVD should produce an estimate of the light that agrees with $\boldsymbol{\rho}^{\,e\star}$.

We demonstrate below that this planar constraint does indeed improve the estimate of $\boldsymbol{\rho}^{\,e}$, verifying the suitability of the plane constraint applied as a post-processing step, for any candidate colour constancy algorithm. In the next section we will demonstrate the substantial improvement delivered by this simple planar constraint when added to each of several well-known colour constancy algorithms as a post-processing step.

5.5 Experiment Results

5.5.1 Datasets

We apply our proposed method to three different real-image datasets [14, 160, 40] and compare our results to other colour constancy algorithms.

Our first experiment uses the Barnard dataset [14], denoted the SFU Laboratory dataset, which contains 321 measured images under 11 different measured illuminants. The scenes are divided into two sets as follows: minimal specularities (22 scenes, 223 images – i.e., 19 missing images); and non-negligible dielectric specularities (9 scenes, 98 images – 1 illuminant is missing for 1 scene).

For a more real-world (out of the laboratory) image experiment we used the re-processed version of the Gehler colour constancy dataset [93], denoted the ColorChecker dataset, provided by Shi and Funt [160]. This dataset consists of 568 images, both indoor and outdoor. The illuminant ground truth for these images is known because each image has a Macbeth ColorChecker placed in the scene. The ColorChecker is masked off in tests.

Ciurea and Funt [40] introduced the GreyBall dataset, which contains 11346 images extracted from video recorded under a wide variety of imaging conditions. The images are divided into 15 different clips taken at different locations. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the corner of the image. This grey sphere must be masked during experiments.

5.5.2 Previous Methods

To compare, we use the standard well-known colour constancy methods: White-Patch, Grey-World, and Grey-Edge implemented by [180]. For Grey-Edge we use optimal settings, which differ per dataset [98] ($p = 7$, $\sigma = 4$ for the SFU Laboratory dataset and $p = 1$, $\sigma = 6$ for the ColorChecker dataset). We also use the result provided by Gijsenij and et al. [102] for pixel-based gamut mapping, using the best general gamut mapping setting, which is for *1st-jet* as reported in [102] (although we could not precisely match their exact results using the code they released). For other methods we use results as provided by Gijsenij [104, 98].

For methods which need training data, such as the gamut mapping methods, in the SFU Laboratory dataset 31 images (all images recorded under the *syl-50MR16Q*-illuminant) were used for computation of the canonical gamut, and subsequently these were omitted from the test set. For the ColorChecker dataset, three-fold cross-validation was used to learn the canonical gamut (with the folds as well as the ground truth supplied with the original dataset). Testing for supervised methods is as described in [104], §VII-A.

5.5.3 Post-Processing

Table 5.1 shows the accuracy of the plane constraint eq. (5.19) in §5.4.3 as a post-processing step applied to the results of each of the well-known colour constancy algorithms (White-Patch, Grey-World and Grey-Edge), in order to improve the estimate. The errors are in terms of the median of angular errors for 3-D illuminant chromaticity.

As expected, applying the plane constraint can significantly improve the estimate of illuminants while being very fast and easy to implement.

Experimental results in Table 5.1 indicate that estimate errors can be reduced by some 15 to 30 percent by this simple and very fast mechanism.

Table 5.1: Median of angular errors for well-known colour constancy algorithms for the SFU Laboratory [14] dataset and ColorChecker dataset [160], plus result after post-processing with planar constraint eq. (5.19) for each colour constancy algorithm.

Method	SFU Lab.	ColorChecker
White-Patch	6.5°	5.7°
White-Patch + Planar Con.	5.1°	4.4°
Grey-World	7.0°	6.3°
Grey-World + Planar Con.	5.0°	4.3°
Grey-Edge	3.2°	4.3°
Grey-Edge + Planar Con.	2.7°	3.8°

5.5.4 Global Search and Analytic Solution Experiment

Table 5.2 indicates the accuracy of the proposed methods for the SFU Laboratory dataset [14], the ColorChecker dataset [160] and the GreyBall dataset [40], in terms of the mean and median of angular errors, for several colour constancy algorithms applied to these datasets. For those methods which need data-dependent tunable parameters, we utilize optimal parameters for their dataset. For an overview of results of different algorithms on these dataset refer to [104, 98].

To our knowledge, for the SFU Laboratory dataset the Planar Constraint Search eq. (5.14) does best in terms of median angular error compared to any reported colour constancy method, even those needing training data. We do note that for this dataset the Planar Constraint Search eq. (5.14) is not the best for the ColorChecker dataset, with Gamut Mapping methods performing better. However, both Planar Search and the Analytic method of §5.4.2 (Geomean) do as well or better than the other relatively fast methods for the GreyBall dataset, and are only bested by the much more complex method [100].

Run-times average 5.2s for Planar-Constraint search and 415ms for the Analytic method, compared to 617ms for the GreyEdge algorithm and 63.2s for 1st-Jet Gamut Mapping, operating on the SFU Laboratory dataset using (unoptimized) Matlab.

Figure 5.8 shows examples of corrected images from GreyBall dataset based on proposed method using geometric mean (called the Zeta-Image method) compared to White-Patch, Grey-World, Grey-Edge and Gamut Mapping methods and their angular error compare to ground truth which was

Table 5.2: Angular errors for several colour constancy algorithms for SFU Laboratory dataset [14], ColorChecker dataset [160] and GreyBall dataset [40].

Dataset	SFU Laboratory		Color Checker		Gray Ball	
Methods	Median	Mean	Median	Mean	Median	Mean
White Patch	6.5°	9.1°	5.7°	7.4°	5.3°	6.8°
Gray World	7.0°	9.8°	6.3°	6.4°	7.0°	7.9°
Gray Edge	3.2°	5.6°	4.5°	5.3°	4.7°	5.9°
Bayesian [93]	-	-	3.5°	4.8°	-	-
Gamut Mapping	2.3°	3.7°	2.5°	4.1°	5.8°	7.1°
Gamut Mapping 1jet [102]	2.1°	3.6°	2.5°	4.1°	5.8°	6.9°
Natural Image Statistics [100]	-	-	3.1°	4.2°	3.9°	5.2°
Planar Constraint Search	1.9°	4.3°	2.8°	4.1°	4.6°	5.9°
Geomean	2.1°	6.2°	2.7°	4.2°	4.7°	5.8°

obtained using a grey sphere mounted onto the video camera.

5.6 Specularity Manipulation

The Zeta-image should identify specular pixels. Here we are interested in seeing whether in general the Zeta-image provides a guide to manipulating specularities. We begin with an illuminant-dependent colour space in order to separate the effect of specular reflection. Here we use SUV colour space proposed by Mallick et al. [132], defined by two rotations making the first dimension, S, the known illuminant colour 3-vector. In SUV colour space, the other axes U and V are then orthogonal to the illuminant and free from illumination colour (although for clipped pixels this condition is not valid).

Generally, modifying specularities is reduced to modifying the S channel, which we aim to do by manipulating the Zeta-image. Thus we maintain the UV components unchanged. Since smaller zeta indicates more specular contribution, or specular reflection, we can simply deprecate this amount by multiplying the S channel by a monotonic function $f(\zeta)$ of zeta. We could use regression on non-bright pixels to generate f, but in Fig. 5.7 we simply use following equation:

$$S_{modified} = S(\zeta/\alpha)^{\beta} \tag{5.20}$$

where α normalizes the ζ value and determines the pixels that are affected by ζ; and β, with range

$[-1, 1]$, determines the shape of the transfer-function curve. Positive β reduces specularity since it reduces the modified S; and negative β enhances the specularity. This simple transform generates arguably excellent specular-free images and justifies the applicability of the Zeta-image approach. Fig. 5.6 shows the procedure for specularity manipulation using the Zeta-image.

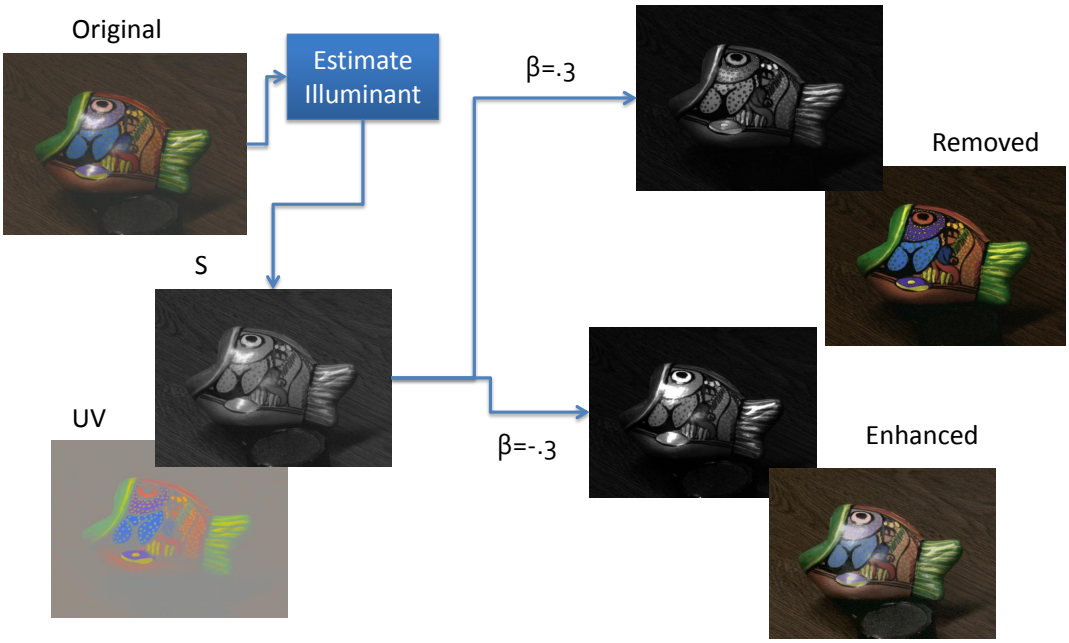
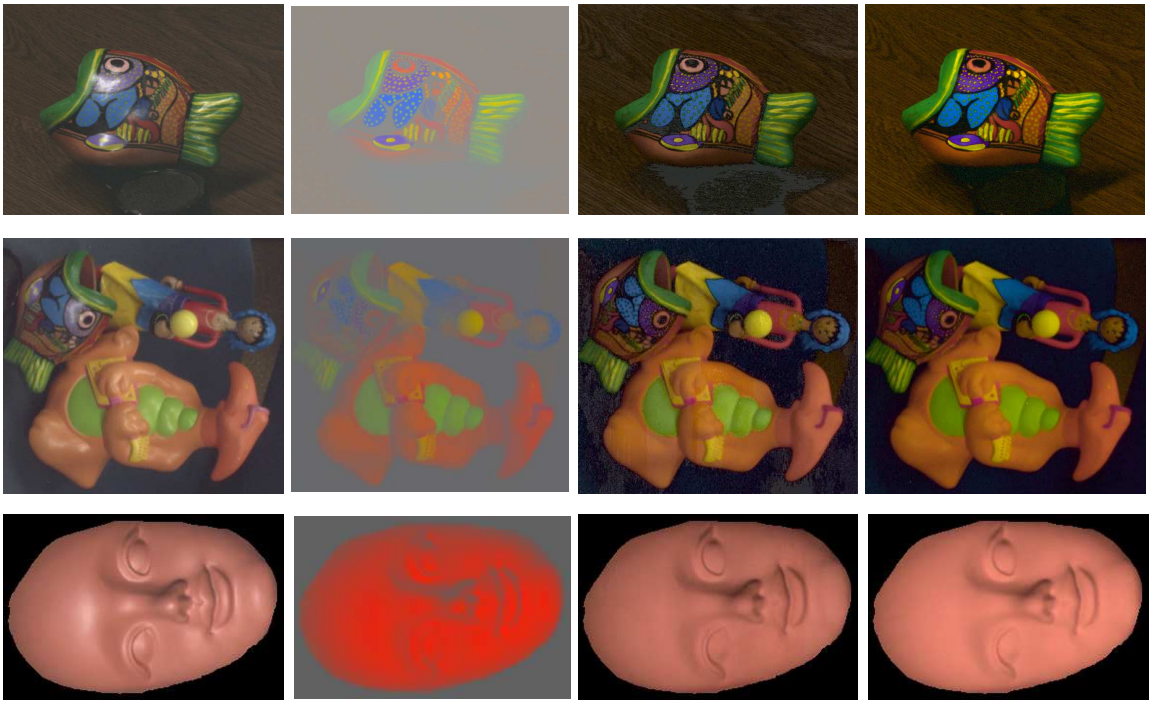


Figure 5.6: The procedure of specularity manipulation using Zeta-image.

Fig. 5.7 shows the result of this test compared to Tan and Ikeuchi's method [164] as well as the UV channels from [132]. Comparison (particularly zoomed-in) with Fig. 5.7(c) shows the higher quality output of the new image; and moreover this advantage is gained without having to know the illuminant. While this is only a brief test of course, it does indicate that, as claimed, the Zeta-image carries additional information regarding specular highlights.



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

Figure 5.7: (a): Original images. (b): Specularity-free UV channels by [132]. (c): Specularity-free images by [164]. (d): Specularity-free images by modifying SUV by the Zeta-image.

5.7 Conclusions

In this chapter we present a novel physics-based insight regarding a plane constraint that obtains for log-relative-chromaticity values near the illuminant point (for white surfaces, or specularities in the neutral-interface model). This insight provides a useful and very simple method for identifying the illuminant chromaticity that requires no training data or tunable parameters. It is easy to implement and very fast compared to complex colour constancy methods such as gamut mapping, while retaining comparable performance. Experiment results over datasets consisting of laboratory images and of real-world images demonstrate that the proposed method significantly outperforms other unsupervised methods while its accuracy of illuminant estimation is comparable with the best (supervised) methods but much faster.

As well, the plane constraint can also be brought to bear to improve estimates provided by

other illuminant estimation algorithms. Since the feature vectors corresponding to specular or near-specular pixels in an image must lie on a plane (within the simple and straightforward model employed here), the planar constraint can be used as a post-processing step; experimental results indicate that estimate errors can be reduced by some 15 percent by this simple and very fast mechanism.

Here we have concentrated on a simple and fast algorithm, with a fixed threshold; but of course the 10% threshold we have used should in fact be a free parameter subject to optimization. Although results show that the new algorithm significantly outperforms other methods in terms of median error, the method does occasionally fail and this brings down mean performance.

To show that the Zeta-image does indeed carry extra information regarding specularities in images, we also showed in a test using calibrated images that the Zeta-image can be used not only to identify specularities but further to generate highlight-free images retaining natural shading, with excellent results. Using two parameters, a simple mapping can be shown to increase specularity or decrease it, in a controlled manner.

Future work includes identifying failure either by heuristics or by integrating with other methods [22, 34, 100]. Nonetheless, performance achieved by this simple approach shows the scientific merit of the new insight obtained. The most interesting contribution is the development of the Zeta-image itself. The use of entropy measures over chromaticity values has not been utilized before, we believe, and further investigation of its implications will be studied.

Figure 5.8: Examples of corrected images from GreyBall dataset based on proposed method using geometric mean (called the Zeta-Image method) compared to White-Patch, Grey-World, Grey-Edge and Gamut Mapping methods, and their angular error compared to ground truth which was obtained by having a grey sphere mounted onto the video camera.

Chapter 6

Camera Calibration for Daylight Specular-Point Locus

In this chapter we present a new camera calibration method aimed at finding a straight-line locus, in a special colour feature space, that is traversed by daylights and as well also approximately followed by specular points. The aim of the calibration is to enable recovering the colour of the illuminant in a scene, using the calibrated camera. First we prove theoretically that any candidate specular points, for an image that is generated by a specific camera and taken under a daylight, must lie on a straight line in log-chromaticity space, for a chromaticity that is generated using a geometric-mean denominator. Use is made of the assumptions that daylight illuminants can be approximated using Planckians and that camera sensors are narrowband or can be made so by spectral sharpening. Then we show how a particular camera can be calibrated so as to discover this locus. As applications we use this curve for illuminant detection, and also for re-lighting of images to show they would appear under lighting having a different colour temperature.

6.1 Introduction

The objective of this chapter is to show that natural lights must necessarily follow a straight-line locus, in a special 2-D chromaticity feature space generated using a geometric-mean denominator to remove the effect of magnitude from colour, and that this locus can be derived from a camera calibration. Transformed back into non-log coordinates, the straight line in log colour space means that in terms of ordinary L_1-norm based chromaticity $\{R, G\}/(R+G+B)$ lights follow a particular

curve. The locus determined is camera-dependent. Derivation of the parameters of this locus via a camera calibration means that then one can use the path to help identify the illuminant in the scene, and also to transform from one illuminant to another.

In this chapter, the main use we make of the above observation regarding the path following by illuminants is to apply this additional constraint to colour constancy algorithms as extra information that can be brought to bear. We show that the specular-locus thus found does help in discovering the lighting in a scene. Moreover, since we know the path that illuminants would take depending on the colour temperature T, we can re-light a scene simply by changing T and thus moving along the locus. Using measured data for changing lights for static scenes we show below that this shift in lighting is indeed accurate.

The history of using specularities to discover the illuminant is lengthy, and here we simply highlight some key contributions used in this chapter. Shafer [157] introduced the widely used and quite effective dichromatic model of reflectance for dielectric materials, wherein surface reflectance consists of (i) a diffuse ('body') component that depends on subsurface material properties of a reflecting surface and (ii) a specular ('surface') component that depends on the air-surface interface layer and not the body-reflectance properties. The diffuse component is responsible for generating the colour and shading for an object and the specular component is responsible for highlights. For a dielectric (e.g., plastics) the neutral-interface model [121] states that the colour of the specular contribution is approximately the same as the colour of the illuminant itself. However, simply taking specular colour as identical with light colour is insufficient: typically, specular reflection looks white to the viewer (for dielectric materials), but in fact a careful inspection of specular pixels shows that the body colour is still present to some degree.

Klinker et al. [114] showed that when the diffuse colour is constant over a surface, the colour histogram of its image forms a T-shaped distribution, with the diffuse and specular pixels forming linear clusters. They used this information to estimate a single diffuse colour. Therefore in order to use this principle, their approach needed to segment an image into several regions of homogeneous diffuse colour. Morever, Lee [120] proposed a method which uses specularity to compute illumination by using the fact that in the CIE chromaticity diagram [194] the coordinates of the colours from different points from the same surface will fall on a straight line connected to the specular point. This is the case when the light reflected from a uniform surface is a additive mixture of the specular component and the diffuse component. This seminal work initiated a substantial body of work on identifying specular pixels and using these to attempt to discover the illuminant [122, 164]. Another approach extending these algorithms is to define a constraint on the possible colours of illumination,

making estimation more robust [73, 74].

Finlayson and Drew [60] used 4-dimensional images (more colours than R,G,B) formed by a special 4-sensor camera. They first formed colour ratios to reduce the dimensionality to 3 and to eliminate light intensity and shading; then projecting log values into the plane orthogonal to the direction in the 3-D space corresponding to a lighting change direction they arrived at generalized colour 2-vectors independent of lighting. They noted that in the 2-space, specularities are approximately linear streaks pointing to a single specular point. Therefore they could remove specularities by the simple expedient of replacing each 2-D colour by the maximum 2-vector position at its particular direction from the specular point. Note, however, that in [60] the authors were constrained to using a *four*-sensor camera. Here we relax that necessity by adding a more complete camera calibration phase.

In [128], Lu and Drew carried out an analysis again based on the formulation in [60], but in 3-D rather than 4-D and using an additional image generated by imaging a with-flash exposure in addition to an image with no flash. The addition of an extra image means that by subtracting the images an estimate of illuminant colour temperature can be established based on closeness to a predetermined set of clusters for different lights in a log-chromaticity space, using the mean over the image in that space compared to the clusters.

In this chapter we present a new camera calibration method aimed at finding a specular-point locus in the log-chromaticity colour feature space, for daylight illuminants. We prove that, in a simplifying model for image formation under non-fluorescent illumination, any candidate illuminants for an image generated by a specific camera must lie on a line in log-log chromaticity space if we use a geometric mean to normalize colour. This has the consequence that ordinary r, g chromaticities formed by dividing by the sum $R + G + B$ must lie on a specific curve. To support these theoretical considerations, we demonstrate the applicability of the line in log chromaticity space for several different datasets and, as applications, we use the resulting curve for illumination recovery and re-lighting with a different illumination.

In essence, we are proposing a type of new colour constancy algorithm, one that uses a camera calibration. Many colour constancy algorithms have been proposed (see [110, 104] for an overview). The foundational colour constancy method, the so-called White-Patch or Max-RGB method, estimates the light source colour from the maximum response of the different colour channels [117]. Another well-known colour constancy method is based on the Grey-World hypothesis [32], which assumes that the average reflectance in the scene is achromatic. Grey-Edge is a recent version of the Grey-World hypothesis that says: the average of the reflectance differences in a scene is achromatic

[180]. Finlayson and Trezzi [75] formalize grey-based methods by subsuming them into a single formula using the Minkowski p-norm. The Gamut Mapping algorithm, a more complex and more accurate algorithm, was introduced by Forsyth [79]. It is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours. Several extensions have been proposed [8, 67, 59, 102].

The chapter is organized as follows: To begin, in §6.2 we discuss the underlying assumptions that allow us to create a simplified model of colour image formation. Then in §6.3 we examine how the simplified model plus an offline calibration of the camera can be used to analyze the specular highlights. We propose a specular-point locus in chromaticity space in §6.4 based on the calibration for each camera. In §6.5 and §6.6 we use the proposed illuminant locus to demonstrate its applicability in two application areas: illuminant identification, and image re-lighting. In §6.7 we introduce a method to generate a matte image using our estimated illuminant, giving a specular-free image. Finally, we conclude the chapter in §6.8.

6.2 Image Formation

To generate a simplified image formation model we apply the following set of simplifying assumptions (cf. [70]): (1) illumination is Planckian or is sufficiently near the Planckian locus that a blackbody radiator forms a reasonable approximation for this use [68]; (2) surfaces are dichromatic [157]; and (3) RGB camera sensors are narrowband or can be made sufficiently narrowband by a spectral-sharpening colour-space transform [63].

Thus we begin by considering a narrowband camera, with three sensors. Note again that in [60] the authors were constrained to using a *four*-sensor camera. Here we relax that necessity by adding a more complete camera calibration phase for a camera with only three sensors.

Real camera sensor curves are not in fact narrowband: Below, we investigate how the assumption of Planckian lighting impacts models of image formation by making use of a 3-sensor delta-function sensitivity camera. It is evident that real sensors are far from idealized delta functions: each is typically sensitive to a wavelength interval over 100nm in extent. Nevertheless, as we shall see, they behave sufficiently like narrowband sensors for our theory to work and moreover this behaviour could be promoted by carrying out calculations in an intermediate spectrally sharpened colour space [63].

Now let us briefly examine image formation in general for a dichromatic reflectance function comprising Lambertian and specular parts. For the Lambertian component, suppose there are $i =$

$1..L$ lights, each with the same SPD $E^i(\lambda)$ (e.g., an area source) given by Wien's approximation of a Planckian source [194]:

$$E^i(\lambda) \;=\; I^i c_1 \lambda^{-5} e^{-c_2/(\lambda T_i)} \;,\; c_1 = 3.74183 \times 10^{16} \,,\; c_2 = 1.4388 \times 10^{-2} \tag{6.1}$$

with distant lighting from lights in normalized directions $\boldsymbol{a}^{\,i}$ with intensities I^i (the constant c_1 determines the units). If the surface projecting to retinal point $\boldsymbol{x}$ has spectral surface reflectance $S(\lambda)$ and normal $\boldsymbol{n}$ then, for a delta-function narrowband sensor camera with spike sensor sensitivities $Q_k(\lambda)$=$q_k \delta(\lambda - \lambda_k)$, $k = 1..3$, the 3-vector RGB response R_k is

$$\begin{aligned}
R_k \;&=\; \textstyle\sum_{i=1}^{L} \boldsymbol{a}^{\,i} \cdot \boldsymbol{n} \int E^i(\lambda) S(\lambda) Q_k(\lambda) d\lambda \\
&=\; \textstyle\sum_{i=1}^{L} c_1 \boldsymbol{a}^{\,i} \cdot \boldsymbol{n}\; S(\lambda_k) I^i(\lambda_k)^{-5} e^{-c_2/(\lambda_k T_i)}\; q_k \\
&=\; \left[\textstyle\sum_{i=1}^{L} (c_1 I^i \boldsymbol{a}^{\,i})\right] \cdot \boldsymbol{n}\; S(\lambda_k) (\lambda_k)^{-5} e^{-c_2/(\lambda_k T)}\; q_k \qquad \text{if all } T_i = T \\
&\equiv\; \widetilde{\boldsymbol{a}} \cdot \boldsymbol{n}\;\; S(\lambda_k) \lambda^{-5} e^{-c_2/(\lambda_k T)}\; q_k \;,\; k = 1..3
\end{aligned} \tag{6.2}$$

The above is the matte model employed. For the specular part, let us assume a specular model dependent on the half–way vector $\boldsymbol{n_S}$ between the illuminant direction and the viewer:

$$\boldsymbol{R}^{\,\text{Specular}} \;=\; \sum_{i=1}^{L} \boldsymbol{b}^{\,i}_{S} \Phi(\boldsymbol{n}^{\,i}_{S} \cdot \boldsymbol{n}) \,, \tag{6.3}$$

where $\boldsymbol{b}^{\,i}_{S}$ is the colour of the specularity for the i^{th} light. E.g., in the Phong specular model [78],

$$\Phi(\boldsymbol{n_S}^{i} \cdot \boldsymbol{n}) = (\boldsymbol{n_S}^{i} \cdot \boldsymbol{n})^p \quad , \tag{6.4}$$

where a high power p makes a more focussed highlight.

Now in a neutral interface model [120], the colour of the specular term is approximated as:

$$\boldsymbol{b}^{\,i}_{S} \;\equiv\; \text{colour of the light.} \tag{6.5}$$

Hence for Lambertian plus Specular reflectance, we arrive at a simple model:

$$R_k \;=\; \left[\widetilde{\boldsymbol{a}} \cdot \boldsymbol{n}\; S(\lambda_k) \;+\; \textstyle\sum_{i=1}^{L} c_1 I^i \Phi(\boldsymbol{n}^{\,i\,T}_{S} \boldsymbol{n})\right] \lambda_k^{-5} e^{-c_2/(\lambda_k T)}\; q_k \quad . \tag{6.6}$$

For each pixel at a retinal position $\boldsymbol{x}$, the second term in the brackets is a constant, β say, that depends only on geometry and not on the light colour. Therefore we have

$$R_k \;=\; [\widetilde{\boldsymbol{a}} \cdot \boldsymbol{n}\, S(\lambda_k) \;+\; \beta]\, \lambda_k^{-5} e^{-c_2/(\lambda_k T)}\, q_k \tag{6.7}$$

with possibly several specular highlights on any surface ($\beta = \beta(\boldsymbol{x})$).

If we define

$$\alpha \;=\; \beta/(\widetilde{\boldsymbol{a}} \cdot \boldsymbol{n}) \quad , \tag{6.8}$$

then our expression simplifies to:

$$R_k \;=\; (\widetilde{\boldsymbol{a}} \cdot \boldsymbol{n})\, [S(\lambda_k) + \alpha]\, \lambda_k^{-5} e^{-c_2/(\lambda_k T)}\, q_k \tag{6.9}$$

6.3 Specular-Point Line in Log Chromaticity Space

We note that dividing by a colour channel (green, say) removes the initial factor in eq. (6.9). We can divide instead by the geometric mean (cf. [60]) so as not to be forced to choose a particular normalizing channel. Define the mean R_M by

$$R_M \;=\; \sqrt[3]{\Pi_{k=1}^{3}\, R_k} \quad . \tag{6.10}$$

Then we can remove light intensity and shading by forming a chromaticity 3-vector $\boldsymbol{r}$ via

$$r_k = R_k/R_M,\; k = 1..3. \tag{6.11}$$

Thus from eq. (6.9) we have

$$\log r_k \;=\; \log\left(\frac{s_k + \alpha}{s_M + \alpha}\right) \;+\; w_k \;+\; (e_k - e_M)\,\frac{1}{T}\,, \quad k = 1..3, \tag{6.12}$$

where we simplify the expressions by defining some short-hand notations as follows:

$$\begin{aligned} &s_k \;=\; S(\lambda_k);\; v_k \;=\; \lambda_k^{-5} q_k;\; v_M = \left\{\textstyle\prod_{j=1}^{3} \lambda_j^{-5}\, q_j\right\}^{1/3} \;, w_k \;=\; \log\left(v_k/v_M\right) \\ &e_k \;=\; -c_2/\lambda_k;\; e_M \;=\; (-c_2/3)\, \textstyle\sum_{j=1}^{3}(1/\lambda_j)\,, \end{aligned} \tag{6.13}$$

and we define an effective geometric-mean-respecting value s_M by setting

$$(s_M + \alpha) \;\equiv\; \left\{\prod_{j=1}^{3}(s_j + \alpha)\right\}^{1/3}$$

In the case of broad-band sensors we replace some of the definitions in eq. (6.12) above by values that are equivalent for delta-function cameras but are appropriate for real sensors (extending definitions in [60]):

$$\begin{aligned} \sigma_k &= \int q_k(\lambda) d\lambda \,, \\ e_k &= (1/\sigma_k) \int -(c_2/\lambda) q_k(\lambda) d\lambda \,, \\ e_M &= (1/3) \textstyle\sum_{j=1}^{3} e_k \,, \\ s_k &= (1/\sigma_k) \int S(\lambda) q_k(\lambda) d\lambda \,, \\ v_k &= \int \lambda^{-5} \, q_k(\lambda) d\lambda \end{aligned} \tag{6.14}$$

The meaning of eq. (6.12) is that the log of the chromaticity is given by: (i) A term consisting of the matte-surface term s_k combined with a term α, a scalar at each pixel that is the specular contribution; (ii) a constant 3-vector offset term, w_k, which is a characteristic of the particular camera; and (iii) a term equal to the product of a "lighting-change" 3-vector $(e_k - e_M)$, also characterizing the camera, times the inverse of the correlated colour temperature T encapsulating the colour of the light.

Thus as the light colour (i.e., T) changes, say into a shadow or because of inter-reflection, the log-chromaticity at a pixel $\boldsymbol{x}$ simply follows a straight line in 3-space (as temperature T changes), along the light-change direction $(e_k - e_M)$, even including the specular term α. For a fixed T, if α changes on a patch with reflectance vector s_k, then the plot of $\log \boldsymbol{r}$ will be a curved line.

In this chapter, we mean to calibrate the camera so as to recover (a projection of) both this light-change vector as well as the constant additive term w_k. The difference from previous work [60] is as follows.

In the method [60], going over to a chromaticity space meant that 4 dimensions were reduced to 3. Then in that 3-space, light-change vector $(e_k - e_M)$ was obtained as the first eigenvector of mean-subtracted colour-patch values. To then go over to a 2-space, log-chromaticity values were then projected onto the subspace orthogonal to 3-D light-change vector. This meant that all lighting colour and strength were projected away. In that plane, the illuminant, and consequently the specular point as well, were always located in precisely the same spot. It was argued that, at a highly specular point in an input image, the pixel values would essentially consist of the specular point and thus one could derive that point from training images. Then forming radii from that specular spot out to the least-specular pixel position effectively removed specularities.

Here, in contrast, we start with 3-D colour values, rather than 4-D ones, and so chromaticity vectors are effectively 2-D. Now calibration of the camera is used to provide both a value of the

offset term w_k in eq. (6.12) as well as of the lighting-colour-change vector $(e_k - e_M)$.

For specular pixels, there is no surface term s_k above in eq. (6.12), and $\log(\alpha/\alpha) = 0$, so the value of this log-geometric-mean chromaticity at a purely specular pixel becomes the simpler form

$$\log r_k \;=\; w_k \;+\; (e_k - e_M)\,\frac{1}{T}\;, \quad k = 1..3, \tag{6.15}$$

Thus as T changes we have a line, in a 2-D colour space, whereon any specular point must lie. To determine just where it does lie, we form an objective function measure, which is in fact minimized provided we choose the correct value of T: an example of such a measure is given below in §6.5.1. Hence we recover the temperature T and therefore the light colour. Moreover, since have an illuminant locus we can go on to re-light images by moving the illuminant along the locus obtained during the camera calibration phase. Such re-lit images are shown below in §6.6 where images are shown as they would appear under a different colour temperature.

Note that although we work with 3-vectors, the step of division by the geometric mean creates $\log \boldsymbol{r}$ vectors that lie on a plane: they are all orthogonal to the vector $(1,1,1)^T$ — in fact, each of the three terms in eq. (6.12) lies in this plane. Thus the components are not independent.

6.4 Recovery of Specular-Point Locus

To find the vector $(e_k - e_M)$, $k = 1..3$ we image matte Lambertian colour patches. Here we use the 18 non-grey patches of the Macbeth ColourChecker [134]. We form $\log r_k$ values using temperatures T from $5500°$K to $10500°$K.

According to eq. (6.12) (with no specular contribution), for each surface we should see a set of points in 3-space that falls on a straight line along $(e_k - e_M)$. Thus for each surface, if we then subtract the mean, in each channel k of $\log r_k$, we see a set of nearly coincident lines through the origin.

Therefore, as pointed out in [68] (in a 2-D setting like eq.(6.12) but with $k = 1..2$), we can find vector $(e_k - e_M)$ by forming the covariance matrix of mean-subtracted $\log r_k$ values and calculating eigenvectors. The first eigenvector is the desired approximation of direction $(e_k - e_M)$.

To derive the offset term w_k, we utilize the recovered normalized version of vector $(e_k - e_M)$ and image two lights (below) to determine the scaling along the inverse-temperature line.

Since we know that our colour features lie on the plane perpendicular to the unit vector $\boldsymbol{u} = 1/\sqrt{3}(1,1,1)^T$, to simplify the geometry we first rotate all our log-chromaticity vector coordinates into that plane by forming the projector P_u onto the $\boldsymbol{u}$ direction. 2-D coordinates $\boldsymbol{\chi}$ are formed by

multiplication of the rotation matrix U from the eigenvector decomposition of the projector $P_u^{\perp} = I - P_u$ onto the plane:

$$P_u^{\perp} \;=\; U^T \, diag(1,1) \, U, \qquad U \text{ is } 2 \times 3. \tag{6.16}$$

We denote 2-vectors in this 2-D space as $\boldsymbol{\chi}$. And, explicitly, we form 2-vectors in the plane by

$$\boldsymbol{\chi} \;=\; U \, \log \boldsymbol{r} \tag{6.17}$$

Now suppose that in the 2-D coordinates $\boldsymbol{\chi}$, two lights E_1 and E_2 produce vectors $\boldsymbol{\chi}^{E_1}$ and $\boldsymbol{\chi}^{E_2}$: for each light we form chromaticity (6.11), take logs, and then project via (6.17). Consider the recovered *normalized* light-change direction vector, projected into this plane: define the 3-vector $\boldsymbol{e}$ as having components $(e_k - e_M)$, and denote its unit 2-vector projection as $\hat{\boldsymbol{\xi}}$. Note that we recover only a *normalized* version of $\boldsymbol{e}$ from our SVD analysis of imaged colour patches, with the norm unknown. That is, we work in the plane by rotating with U, and further normalize that projected 2-vector, giving a known, normalized, 2-vector $\hat{\boldsymbol{\xi}}$ from our calibration.

Also, denote by $\boldsymbol{\eta}$ the projected vector w_k: this is what we aim to recover.

$$\begin{aligned} \hat{\boldsymbol{\xi}} \;&=\; (U\boldsymbol{e}\,)/\nu\,, \;\; \text{where} \;\; \nu \;\equiv\; \|U\boldsymbol{e}\,\|, \\ \boldsymbol{\eta} \;&=\; U\boldsymbol{w}\;. \end{aligned}$$

Then the 2-vector coordinates for the two lights E_i, $i = 1..2$ are

$$\chi_\mu^{E_i} \;=\; \eta_\mu \,+\, \nu\hat{\xi}_\mu/T_i, \;\; i = 1..2, \; \mu = 1..2 \tag{6.18}$$

where ν is an unknown scale, and T_i are known colour-temperatures. Note that since we are imaging lights, not surfaces, the surface term s_k in eq. (6.12) is not present.

Forming the difference 2-vector $(\boldsymbol{\chi}^{E_2} - \boldsymbol{\chi}^{E_1})$, we obtain a result involving only the normalized direction $\hat{\boldsymbol{\xi}}$. So we can determine the norm ν if we know T_1 and T_2. For consider the difference 2-vector

$$\chi_\mu^{E_1} \,-\, \chi_\mu^{E_2} \;=\; \nu\hat{\xi}_\mu \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \tag{6.19}$$

Even from these two data points we can easily determine the normalized vector $\hat{\boldsymbol{\xi}}$ since it is simply given by the direction of the difference in $\boldsymbol{\chi}$. Since we know T_1, T_2, the norm ν thus falls out of eq. (6.19).

Finally, subtracting the term $\nu\hat{\xi}_\mu/T_i$, $i = 1, 2$, from each of the two $\boldsymbol{\chi}$ vectors and taking the mean, we recover the offset term $\boldsymbol{\eta}$.

Let us denote by $\boldsymbol{\xi}$ the product $\nu\hat{\boldsymbol{\xi}}$, so using this vector and the offset $\boldsymbol{\eta}$ we arrive at a line (for this particular camera calibrated as above) parametrized by temperature T that must necessarily be traversed by any candidate specular point:

$$\boldsymbol{\chi} \;=\; \boldsymbol{\eta} \;+\; (1/T)\boldsymbol{\xi} \tag{6.20}$$

In summary, the calibration algorithm proposed is expressed in algorithm 6.1.

Algorithm 6.1 Proposed Camera Calibration

Colour target:
Record RGB responses R_k, $k = 1..3$ *(reflected from colour target)*
for several lights $\longrightarrow$ *each pixel follows a parallel straight line;*
calculate geometric mean at each pixel from eq. (6.10).
Derive geometric-mean-based chromaticity 3-vector $\boldsymbol{r}$ *from eq. (6.11), and take logarithms.*
Find 3-vector $(e_k - e_M)$ *as first eigenvector for* $\log \boldsymbol{r}$ *values, mean-subtracted for each colour patch.*

For illuminants $E(\lambda)$, characterized by their known temperatures T
(in a light-box, for example):
Derive $\log \boldsymbol{r}$ *as above, for light reflected from a grey patch.*
Project $\log \boldsymbol{r}$ *onto plane orthogonal to* $(1, 1, 1)$ *via eq. (6.17), forming 2-D coordinates* $\boldsymbol{\chi}$.
Subtracting pairs of $\boldsymbol{\chi}$ *values for known values* T*, find 2-D projected light-change vector* $\boldsymbol{\xi}$ *via eq. (6.19).*
Using $\boldsymbol{\xi}$*, find mean value of camera offset vector* $\boldsymbol{\eta}$ *over* $\boldsymbol{\chi}$ *vectors used, eq. (6.20).*

As set out in algorithm 6.1, a more accurate way to recover the offset term $\boldsymbol{\eta}$ and the vector $\boldsymbol{\xi}$ is to utilize several different known illuminants and capture them using the camera to be calibrated: lights should approximately lie on a straight line in $\boldsymbol{\chi}$ space. Then line parameters $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, as well as outliers, can be recovered using a robust regression method such as the Least Median of Squares (LMS) [151].

We shall find in the following sections that the offset $\boldsymbol{\eta}$ and the vector $\boldsymbol{\xi}$ are all the calibration information that we need for different applications such as illuminant identification and re-lighting.

6.4.1 Real Images

The image formation theory used is based on three idealized assumptions: (1) Planckian illumination, (2) dichromatic surfaces; and (3) narrowband camera sensors. To determine if real images

stand up under these constraints and generate the needed straight line in 2-D colour space, we make use of datasets of measured images [93, 14]. Fig. 6.1 displays measured illuminant points in χ space for 86 scenes captured by a high-quality Canon DSLR camera for 86 different lighting conditions [93]. Notwithstanding the fact that the camera sensors are not narrowband and illuminants are not perfectly Planckian, we can see that these illuminants do indeed approximately form a straight line, thus justifying the suitability of the theoretical formulation.

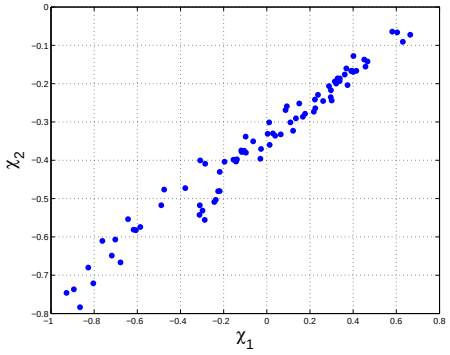
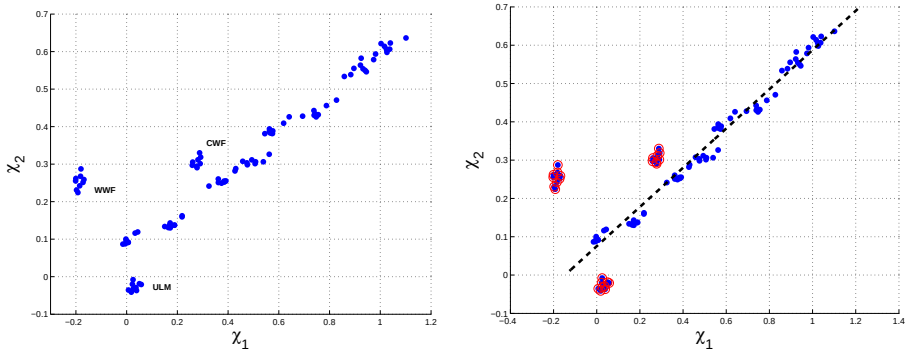


Figure 6.1: 86 illuminants for Canon camera in χ space [93]. Note that these illuminants do approximately follow a straight line.

Since we assume that lights can be characterized as Planckian, we expect that severely non-Planckian lights will form outliers to the straight-line path determined. Figs. 6.2(a,b) demonstrate that this is indeed that case. Here we show illuminant points transformed to 2-D χ space for 98 images consisting of measured images of 9 objects that are specifically selected to include substantial specular content, under different illumination conditions [14]. In this dataset, illuminants for 26 of the images are fluorescent (Sylvania Warm White Fluorescent(WWF), Sylvania Cool White Fluorescent (CWF) and Philips Ultralume Fluorescent(PUF)). These show up in Fig. 6.2(a) as outlier points. Fig. 6.2(b) shows that the robust LMS method correctly identifies these points as outliers and thus does not include them in calculating line parameters.

6.5 Illuminant Identification

Our camera calibration process has generated a locus in chromaticity space that candidate natural daylight illuminations will follow. In this section we show how for a new image we can identify a point on this locus as an estimate of the illuminant.



(a) (b)

Figure 6.2: (a): 98 illuminants for images containing significant specular content [14], plotted in 2-D $\boldsymbol{\chi}$ colour space. Note clusters of points that arise from fluorescent illuminants (WWF, CWF, ULM). (b): Outliers automatically determined by LMS regression are shown using a red circle, and the regression line is shown as black dashed.

Recently, Drew et al. [48] presented an illuminant estimation method based on a planar constraint. This stated that for near-specular pixels, Log-Relative-Chromaticity (LRC) values are orthogonal to the light chromaticity: they showed that if one divides image chromaticity by illuminant chromaticity, then in a log space the resulting set of 3-vectors are approximately planar, for near-specular pixels, and orthogonal to the lighting — for the correct choice of the illuminant only. Hence they propose an objective function based on this planar constraint which is minimized for the correct illuminant.

Here, we utilize this daylight illuminant planar constraint by further constraining the light to lie on the daylight locus we have derived above. The locus provides an additional constraint on the illuminant and hence improves the estimate.

To begin, we briefly recapitulate below the derivation of this planar constraint.

6.5.1 Planar Constraint

Suppose we rewrite eq. (6.9) for the 3-vector RGB response $\boldsymbol{R}$, here relinquishing the requirements that lighting be Planckian and sensors be narrowband, but instead applying the different simplifying assumption that matte pixel 3-vector RGB triples be a component-wise product of a light 3-vector $epsilon_k$, $k = 1..3$, and a surface triple ς_k [29]. Here, ς_k is the reflectance at a pixel under equi-energy white light.

Adding a Neutral Interface Model term [120] for specular content, as in eq. (6.7), we have approximately

$$R_k \;\simeq\; \frac{\kappa \varsigma_k \epsilon_k}{q_k} \;+\; \beta \epsilon_k \tag{6.21}$$

where κ is shading. E.g., for Lambertian matte shading κ equals lighting-direction dotted into surface normal. Here, q_k is again a triple giving the overall camera sensor strength [47]; β represents the amount of specular component at that pixel. The value of β for a pixel will depend upon the lighting direction, the surface normal, and the viewing geometry [157]. Let us lump values $\kappa\varsigma_k/q_k$ into a single quantity and for convenience call this simply ς_k. Now we have

$$R_k \;=\; \varsigma_k \epsilon_k \;+\; \beta \epsilon_k \tag{6.22}$$

Instead of the geometric-mean based chromaticity $\boldsymbol{r}$ in eq. (6.11), let us make use of the standard L_1-norm based chromaticity [194]

$$\boldsymbol{\rho} \;=\; \{R, G, B\}/(R+G+B) \tag{6.23}$$

Thus here we have

$$\rho_k \;=\; \frac{\varsigma_k \epsilon_k \;+\; \beta \epsilon_k}{\sum_{j=1}^{3} (\varsigma_j \epsilon_j \;+\; \beta \epsilon_j)} \tag{6.24}$$

Let us define the Log-Relative-Chromaticity (LRC) as the above chromaticity divided by the chromaticity for the lighting itself, ρ_k^{ϵ}. The planar constraint [48] says that for near-specular pixels, LRC values are orthogonal to the light chromaticity, provided we have chosen the correct illuminant to divide by.

To see how this constraint arises, form the LRC, which we denote as ψ_k:

$$\psi_k \;=\; \log\left(\frac{\rho_k}{\rho_k^{\epsilon}}\right) \;=\; \log\left(\frac{\varsigma_k \epsilon_k \;+\; \beta \epsilon_k}{\sum_{j=1}^{3}(\varsigma_j \epsilon_j) \;+\; \beta \sum_{j=1}^{3} \epsilon_j} \cdot \frac{\sum_{j=1}^{3} \epsilon_j}{\epsilon_k}\right) \;=\; \log\left(\frac{\varsigma_k + \beta}{\frac{(\sum_j \varsigma_j \epsilon_j)}{E} + \beta}\right) \tag{6.25}$$

For convenience, now define $E \equiv \sum_{j=1}^{3} \epsilon_j = |\boldsymbol{\epsilon}|$ where $|\cdot|$ is the L_1 norm.

Near a specular point, we can take the limit as $(1/\beta) \to 0$. Let $\alpha = 1/\beta$. Then in the limit, $\boldsymbol{\psi}$ goes to

$$\psi_k \;\;=\;\; \lim_{\alpha \to 0} \log\left\{(\alpha\varsigma_k + 1) \,/\, \left(\alpha \textstyle\sum_j (\varsigma_j \epsilon_j)/E + 1\right)\right\} \;\simeq\; \alpha\left(\varsigma_k - \frac{\sum_j \varsigma_j \epsilon_j}{E}\right) \tag{6.26}$$

The above is the Maclaurin series, accurate up to $O\left(\alpha^2\right)$. By inspection, we have that the LRC vector, ψ_k, is *orthogonal to the illuminant vector:* $\sum_{k=1}^{3} \psi_k e_k = 0$, and hence also orthogonal to the illuminant chromaticity, $\sum_{k=1}^{3} \psi_k \rho_k^e = 0$.

The planar constraint therefore suggests finding which illuminant amongst several candidates is the correct choice, for a particular image, by minimizing the dot-product over illuminants, for pixels that are likely near-specular [48]. Define ζ as the dot-product between $\boldsymbol{\zeta}$ and the chromaticity for a candidate illuminant, with $\boldsymbol{\zeta}$ formed by dividing by this same illuminant chromaticity:

$$\zeta \;=\; -\boldsymbol{\psi} \;\cdot \boldsymbol{\rho}^{\,e} \;=\; -\log(\boldsymbol{\rho}\,/\boldsymbol{\rho}^{\,e}) \cdot \boldsymbol{\rho}^{\,e} \tag{6.27}$$

Then we seek to solve an optimization as express in algorithm 6.2

Algorithm 6.2 Illumination Estimation by Zeta using Optimization

$$\begin{array}{ll} Minimize & \min_{\boldsymbol{\rho}^{\,e}} \; \sum_{\boldsymbol{\psi} \,\in \Psi_0} |\zeta| \\ \text{subject to} & \sum_{k=1}^{3} \rho_k^e = 1 \,,\; 0 < \rho_k^e < 1 \,,\; k = 1..3 \end{array} \tag{6.28}$$

where Ψ_0 is a set of pixel dot-product values with the candidate illuminant chromaticity $\boldsymbol{\rho}^{\,e}$ that are likely to be near specular, e.g. those in the lowest 10-percentile.

To include the Daylight Locus constraint, for a camera calibrated as above in algorithm 6.2, we consider only natural illuminants lying on the curve (6.20).

6.5.2 Experimental Results

We apply our proposed method to two different real-image datasets [14, 40] and compare our results to other colour constancy algorithms. The motivation here is to investigate whether the derived daylight locus correctly helps identify illuminants that are indeed daylights. We show that this is the case.

Laboratory Images

Our first experiment uses the Barnard dataset [14], denoted here as the SFU Laboratory dataset (introduced above in §6.4.1). This contains 321 measured images under 11 different measured illuminants. The scenes are divided into two sets as follows: minimal specularities (22 scenes, 223 images – i.e., 19 missing images); and non-negligible dielectric specularities (9 scenes, 98 images – 1 illuminant is missing for 1 scene). In this dataset the illuminant for 86 of the images are fluorescents. To compare to other colour constancy methods, we consider the following algorithms: White-Patch, Grey-World, and Grey-Edge implemented by [180]. For Grey-Edge we use optimal

Table 6.1: Angular errors for several colour constancy algorithms, for SFU Laboratory dataset [14].

Method	all		non-fluorescent	
	Median Er	Mean Er	Median Er	Mean Er
White-Patch	6.5°	9.1°	6.9°	9.9°
Grey-World	7.0°	9.8°	6.4°	9.4°
Grey-Edge ($p = 1$,$\sigma = 6$)	3.2°	5.6°	2.9°	5.3°
Gamut Mapping pixel ($\sigma = 4$)	2.3°	3.7°	1.8°	3.5°
Planar Constraint Search	1.9°	4.3°	1.9°	4.6°
Daylight Locus using Planar Constraint	2.4°	5.1°	**1.6**°	4.4°

settings, which differ per dataset [98] ($p = 7$, $\sigma = 4$ for the SFU Laboratory dataset and $p = 1$, $\sigma = 1$ for the GreyBall dataset below). We also show the results provided by Gijsenij et al. [102] for pixel-based gamut mapping, using the best gamut mapping settings for each dataset.

How the daylight locus information is used is as an additional constraint to the optimization (6.28), whereby candidate illuminants are restricted to the daylight locus determined by our calibration, for the camera used in taking images.

Table 6.1 lists the accuracy of the proposed method for the SFU Laboratory dataset [14], in terms of the mean and median of angular errors, compared to other colour constancy algorithms applied to this dataset. Since the daylight locus is designed for natural lights (Planckian illuminants) and not fluorescents, we expect performance to be better for non-fluorescents, and this is indeed the case for the 86 scenes imaged under fluorescent lighting. As well, we break out results for all methods for non-fluorescent illuminants (235 images). The results show that in fact using the daylight locus outperforms all other methods in terms of median error, notwithstanding the fact that it is a much less complex method than the gamut-mapping algorithms and does not require any tuning parameters.

The main conclusion to be drawn from this experiment is that the daylight locus does aid a planar-constraint driven illuminant identifier when illuminants are indeed natural lights. This justifies the suitability of our daylight-locus formulation as a useful physics-based constraint of natural lighting.

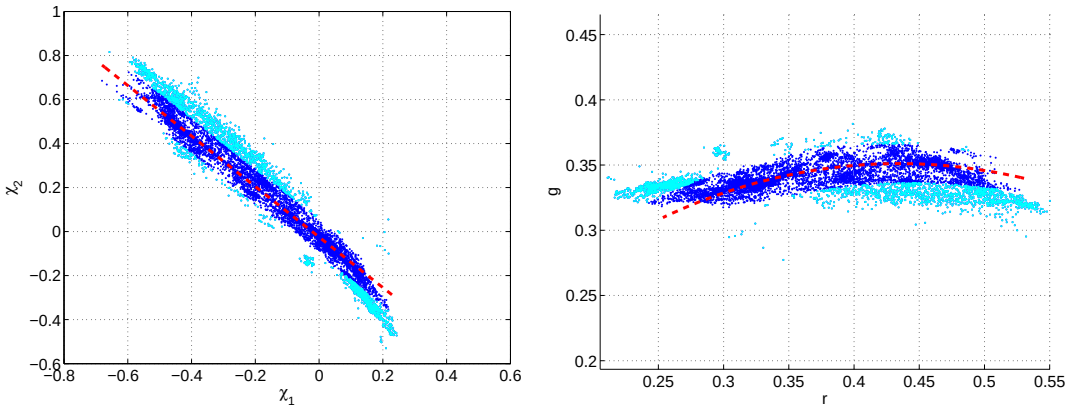
Real-World Images

For a more real-world (out of the laboratory) image experiment we used GreyBall dataset provided by Ciurea and Funt [40]: this dataset contains 11346 images extracted from video recorded under a

wide variety of imaging conditions. The images are divided into 15 different clips taken at different locations. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of the image. This grey sphere must be masked during experiments.

Fig. 6.3(a) shows the illuminants for this image set, mapped into 2-D χ colour space eq. (6.17). We see that these illuminants do approximately follow a straight-line path in 2-space; the LMS-based robust regression method finds a straight-line regression line shown red-dashed. Transformed back into standard L_1-norm based chromaticity space (6.23) the path is curved, as in Fig. 6.3(b).

Table 6.2 shows results for this dataset. We find that the Daylight Locus using the Planar Constraint does better than all the other methods save one: it is only bested by the far more complex Natural Image Statistics method [100]. This is a machine learning technique to select and combine a set of colour constancy methods based on natural image statistics and scene semantics. Again, we find that adding the Daylight Locus information improves the Planar Constraint approach since here lights used are natural daylights.



(a) (b)

Figure 6.3: (a): 11346 illuminants of GreyBall data set [40] in χ 2-space: they approximately follow a straight line locus. (b): The illuminants transformed back into a curve in L_1-norm based chromaticity space.

6.6 Re-Lighting Images

We have shown that by means of calibrating the camera we can recover the specular point for a new image not in the calibration set. That is, the method recovers an estimate of the temperature T

Table 6.2: Angular errors for several colour constancy algorithms for GreyBall dataset [40].

Method	Median Er	Mean Er
White-Patch	5.3°	6.8°
Grey-World	7.0°	7.9°
Grey-Edge ($p = 1$,$\sigma = 1$)	4.7°	5.9°
Gamut Mapping pixel ($\sigma = 4$)	5.8°	7.1°
Natural Image Statistics [100]	3.9°	5.2°
Planar Constraint Search	4.6°	5.9°
Daylight Locus using Planar Constraint	**4.1**$^\circ$	5.6°

for the actual natural illuminant in a test image. Moreover, we have a curve that illuminants must traverse as the lighting colour changes. Consequently it should be possible to re-light an image by changing the position of the specular point along the curve, thus generating new images with a different illuminant.

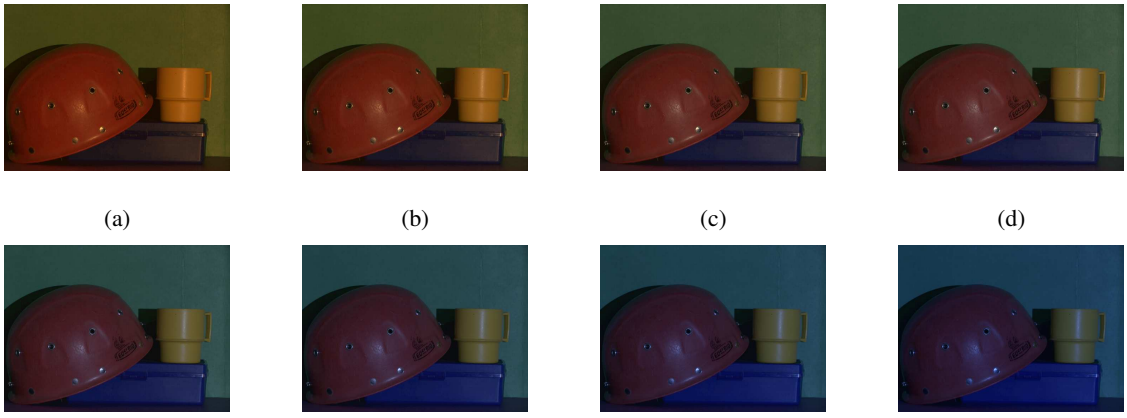
If we again adopt the assumption that camera sensors are narrow-band, we can use a diagonal colour space transform [38] to move the image into new light conditions, via the following equation:

$$\begin{aligned} \boldsymbol{M} &= diag(\boldsymbol{\rho}^{e'})\, diag(\boldsymbol{\rho}^{e})^{-1} \\ \boldsymbol{R}' &= \boldsymbol{R}\ \boldsymbol{M} \end{aligned} \tag{6.29}$$

where $diag(\boldsymbol{\rho}^{e})$ is a 3×3 diagonal matrix with values from vector $\boldsymbol{\rho}^{e}$, and with $\boldsymbol{\rho}^{e}$ the current specular point and $\boldsymbol{\rho}^{e'}$ the new specular point; $\boldsymbol{R}$ and $\boldsymbol{R}'$ are the original and transformed RGB vectors for each image pixel.

Fig. 6.4 shows the same image for different Planckian illuminants from 1500°K to 10000°K, using the proposed re-lighting method. The method arguably produces reasonable output images corresponding to the colour of the lights involved.

In another experiment, we compare the error of using daylight locus for re-lighting, via eq. (6.29) compared to using the actual value of illuminants. Fig. 6.5 shows the same image transferred to other measured images, using their estimated illuminants on the daylight locus. In all, we generated re-lit images for a fixed object under 8 different illuminants (56 re-lightings). In terms of PNSR error for generated images compared to measured ones, we found a median PSNR value of 33.8dB, with minimum and maximum values of 28.2 and 43.5dB. These values demonstrate acceptable faithfulness of rendition for images under new lighting. QAs another comparison, instead of using



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

(e) (f) (g) (h)

Figure 6.4: (a-h): Images generated by re-lighting with Planckians of differing tempertures T=1600°K, 1900°K, 2400°K, 2750°K, 3900°K, 4950°K, 6750°K, 10600°K.

illuminants on the locus we instead used actual measured illuminants in transforming the image via eq. (6.29). Now the min/median/max PSNR values are 28.2, 34.0 and 43.2, almost identical with those found using the illuminant approximation derived from the locus. This demonstrates that using the locus is nearly as good as using the actual illuminant, for this re-lighting task, with negligible difference in results.

6.7 Matte Image from Angle Image

We would like to generate a matte output image, which will then act as an invariant image free of shading and specularity (which could then be used as input to a segmentation scheme, for example). However, our specular-invariant quantity is the angle from the recovered specular point, to each image pixel in feature space. However, this angle encapsulates hue information. The main point is that the angle from the specular point to the feature point of a pixel is approximately independent of the presence or absence of specular content at that pixel. Hence, if there is any structure in the image feature space χ from specular content, then by going over to this 2-D chromaticity space radii from the specular point will be in the same direction for pixels of the same body colour with or without specular content.

Based on the chromaticity-space model [120], a pixel value is a linear combination of the light

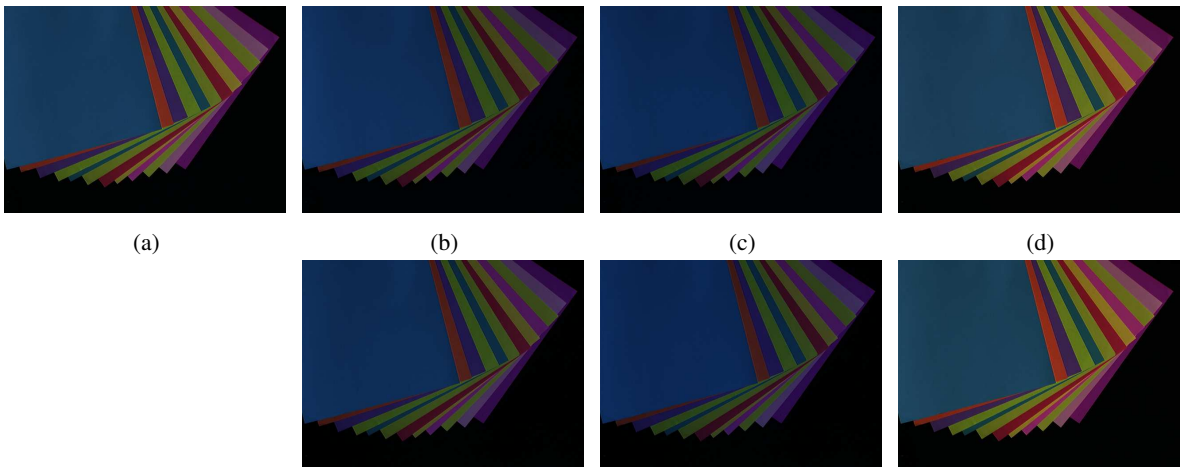


Figure 6.5: (a-d): Input image; (e-g): Images generated by re-lighting of images (a) using (b-d) estimated illuminants on daylight locus. The PSNR for (e-g) are respectively: 42.3, 37.1 and 33.9.

colour and the matte colour, as measured by the camera, resulting in a line in chromaticity space starting from the matte point for any particular colour and leading towards the illuminant colour. Since we already know the light, assumed to be the colour of the specular point, we have this line direction for each pixel, leading from from specular point to that pixel. Moreover, these lines correspond to the angular values that we already assigned to each pixel. We can therefore consider the pixels with the same angular value as belonging to the same matte object — although in real images it is possible that two matte values fall on the same line toward the specular point. Here we initially simply take any such cases as belonging to the same matte value; however, below, considering spatial information we can in fact separate these two matte values from each other.

To make the calculation simpler we transform the chromaticity of the specular point to the origin and use polar coordinate (r, θ). We discretize angle values by using $\lfloor \theta \rfloor$ to have 360 bins. Therefore for each chromaticity point $\boldsymbol{v}$, we consider $(r_v, \theta_v) = polar(v - S)$, where S is the specular point.

The final step to generate a specularity-free colour image is to find a matte value for each pixel. We take the farthest-most pixel from the specular point (i.e., maximum radius r) for each θ as the matte colour (after removing outliers). So the matte colour for each pixel at that angle is identified with the farthest pixel. We call this process "angular projection to matte colour". In other words we are projecting chromaticity points to the border of chromaticity values for each angle, considering

the specular point as the center of projection:

$$matte(\boldsymbol{v}\,) = maxIndex_{\theta_u=\theta_v}(r_u) \tag{6.30}$$

Fig. 6.6 illustrates the projection for chromaticity points for a real image by angular projection to matte colour.

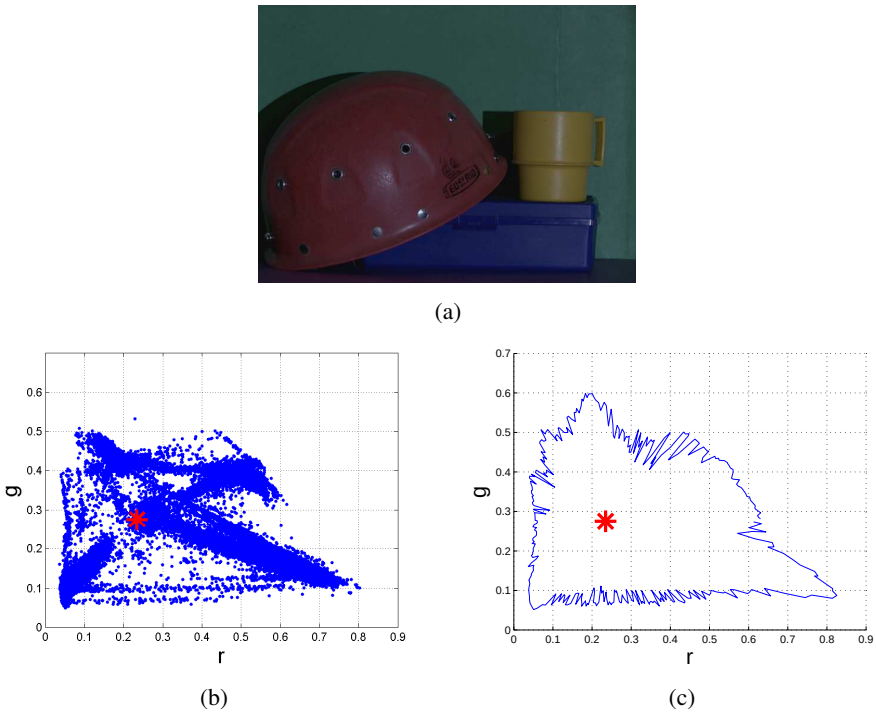
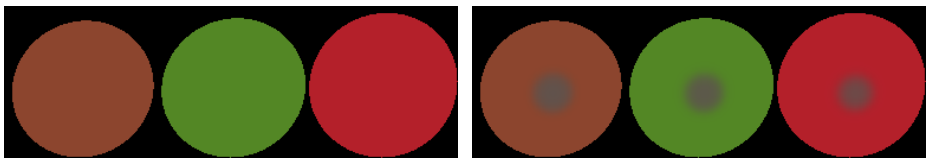


Figure 6.6: (a): A real image. (b): Chromaticity points for (a); the red star is the correct specular point. (d): Angular projection to matte colour for image points.

The angular projection is more sensitive to noise the closer are image feature points to the specular point. Generally, because of noise angular projection to matte colour may completely fail for highlights. Hence we deal with the 10% of pixels that are close to each candidate specular point differently — we iteratively inpaint these pixels using matte colour data from neighbouring pixels that correspond to the same angular value (1-D inpainting). That is, we use voting based on the matte colour of the pixel's neighbours: the new matte colour for that pixel will be the majority of its neighbours' matte colour if it garners at least half of the votes.

For our synthetic example, the resulting chromaticity image $\boldsymbol{\rho}$ is shown in Fig. 6.7(a). Comparing to the *input* chromaticity image in Fig. 6.7(b), we see that the algorithm performs very well for

generating the underlying matte image — specularities in the center of each sphere are essentially gone. In comparison, Fig. 5.2(a) shows the theoretical, correct, matte image, which is indeed very close to the algorithm output in Fig. 6.7(a).



(a) (b)

Figure 6.7: (a): The chromaticity image resulting from angular projection to matte colour. (b): The *input* chromaticity image.

Fig. 6.8 shows results, including finding the specular point and generating a matte colour image, for 4 of input images: whereas the original images' chromaticity clearly shows highlight effects and some shading, output for the proposed method effectively eliminates these effects.

6.8 Conclusion

In this chapter we present a new camera calibration method aimed at recovering parameters for the locus followed by illuminants in a special 2-D chromaticity space. The objective is to discover the colour-temperature of the illuminant in the scene, for a new image not in the training set but captured using the calibrated camera.

As a testing method to verify the validity of the proposed locus idea, we compare illuminant recovery making use of the suggested locus as opposed to not using it. We determined that adding the locus constraint does indeed help identify the scene illuminant. While the effect is not large, nonetheless the experiments do provide a justification of the locus approach — a new insight in physics-based vision.

As an additional capability, we can subsequently generate a new version of the input image, shown as it would appear re-lit under new lighting conditions by considering different illuminant values as the illuminant moves along the specular-point locus.

In future work we will investigate how to make the method more robust to illuminants that differ more substantially from Planckians.

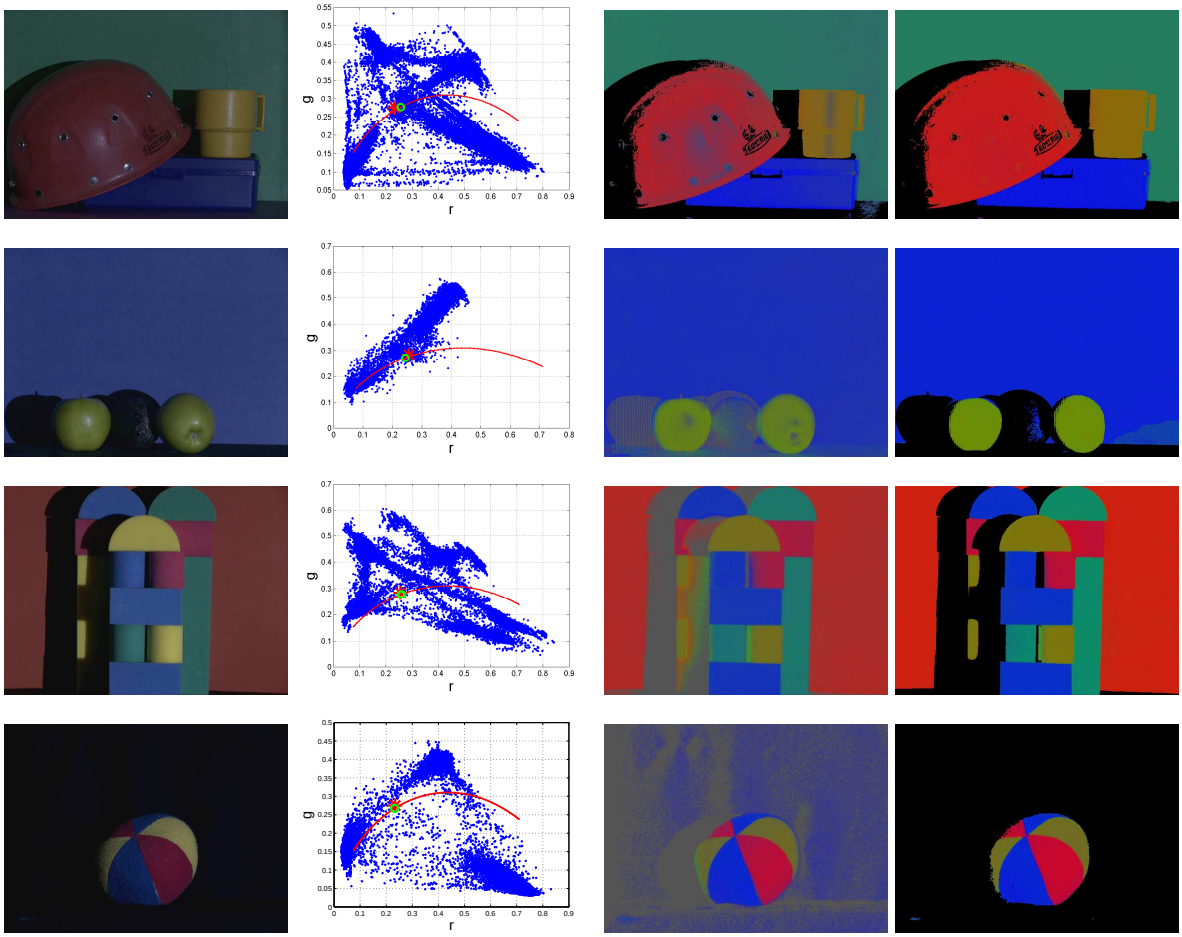


Figure 6.8: Left column: Input image; Second column: chromaticity and illuminant estimate; Third column: input image chromaticity shows highlights; Fourth column: proposed method removes shading and highlights.

Chapter 7

Exemplar-Based Colour Constancy

Exemplar-based learning or, equally, nearest neighbour methods have recently gained interest from researchers in a variety of computer science domains because of the prevalence of large amounts of accessible data and storage capacity. In computer vision, these types of technique have been successful in several problems such as scene recognition, shape matching, image parsing, character recognition and object detection. Applying the concept of exemplar-based learning to the problem of colour constancy seems odd at first glance since, in the first place, similar nearest neighbour images are not usually affected by precisely similar illuminants and, in the second place, gathering a dataset consisting of all possible real-world images, including indoor and outdoor scenes and for all possible illuminant colours and intensities, is indeed impossible. In this chapter we instead focus on *surfaces* in the image and address the colour constancy problem by unsupervised learning of an appropriate model for each training surface in training images. We find nearest neighbour models for each surface in a test image and estimate its illumination based on comparing the statistics of pixels belonging to nearest neighbour surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate. We show that it performs very well, for standard datasets, compared to current colour constancy algorithms, including when learning based on one image dataset is applied to tests from a different dataset. The proposed method has the advantage of overcoming multi-illuminant situations, which is not possible for most current methods since they assume the colour of the illuminant is constant all over the image. We show a technique to overcome the multiple illuminant situation using the proposed method and test our technique on images with two distinct sources of illumination using a multiple-illuminant colour constancy dataset. The concept proposed here is a completely new approach to the colour constancy problem and provides a simple learning-based framework.

7.1 Introduction

Many computer vision applications as well as image processing problems for both still images and video can make use of colour constancy processing as a prerequisite to ensure that the perceived colour of the surfaces in the scene does not change under varying illumination conditions. The observed colour of the surfaces in the scene is a combination of the actual colour of the surface, i.e., the surface reflection function, as well as the illumination. Estimation of illumination is the main goal of the colour constancy task.

Recently, notwithstanding large amounts of accessible data, many problems can be simply solved by a search through data instead of applying sophisticated algorithms. Sometimes these methods make use of nearest neighbour methods. Such use of these techniques occurs in a variety of computer vision problems such as shape matching [19], character recognition [18], human pose estimation [158], image parsing [126], scene recognition [195] and object detection [130]. As an example, Torralba et al. [171] gathered a large data set of some 80 million tiny 32×32 images, each labelled with a word. They solve different computer vision problems such as scene recognition, object classification, person detection, object categorization, picture orientation determination and even colorization by nearest neighbour methods using this large dataset.

Learning based on a previously seen examples is not a new concept. This concept appears in different domains such as exemplar theory in psychology as a model of perception and categorization, case-based reasoning in artificial intelligence and instance-based methods [5] in machine learning.

Many colour constancy algorithms have been proposed (see [110, 104] for an overview). The White-Patch, or Max-RGB, method estimates the light source colour from the maximum response of three different colour channels [117]. Another well-known colour constancy method is based on the Grey-World hypothesis [32], which assumes that the average reflectance in the scene is achromatic. Grey-Edge is a recent version which assumes that the average of the reflectance differences in a scene is achromatic [180]. Shades of Grey [75] is another grey-based method which uses the Minkowski p-norm instead of regular averaging. The Gamut Mapping algorithm [79], a more complex and more accurate algorithm, is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours. As mentioned in [36], these methods deal with an image as a bag of pixels and the spatial relation between pixels is not considered.

Applying the concept of exemplar-based or instance-based learning to the colour constancy problem seems to be an odd idea at first glance since similar or nearest neighbour images are not usually affected by precisely similar illuminants and moreover gathering a dataset consisting of all

possible real world images including indoor and outdoor scenes for all possible illuminant colours and intensities is indeed impossible. In contrast, what can we say about surfaces themselves? Every moderate sized dataset of real images includes thousands of surfaces under different viewing and lighting conditions. We can make these surfaces weakly invariant to illumination changes by simple colour constancy algorithms. Therefore, using the exemplar theory concept we can reduce our illumination estimation task down to the following steps: (1) finding surfaces in an image; (2) finding a similar surface or surfaces in the training dataset for each of our image surfaces; (3) estimating the illumination for each surface based on comparing the statistics of pixels belonging to similar surfaces with the target surface; (4) combining these estimated illuminants into a unique estimate.

In this chapter we present a wholly new line of approach to the colour constancy problem, which we call Exemplar-Based Colour Constancy. We use both texture features and weakly colour-constant three-channel RGB colour values in order to find the nearest neighbour surfaces from training data for each surface. Then we estimate the possible illuminant for each surface based on histogram matching of each surface to its nearest neighbour surfaces from training data. The final step is integrating these estimates into a unique illuminant estimation for the whole image. Since we have no labelled or clustered data for our training process as would be the case for a semantic segmentation task or texture detection task, we lack information for providing confidence for our mapping (such as k-nearest neighbour). Nevertheless, although we find some amount of mismatching for surfaces, the illumination estimation process simply considers these cases as outliers compared to the other estimates. Operating on three standard colour constancy datasets, we show that exemplar-based colour constancy produces excellent results that are better than for previous colour constancy algorithms.

Most colour constancy algorithms assume that the spectral distribution of light source is uniform across the image and therefore that the colour of illuminant is constant all over the image, Hence, estimation of this uniform illuminant is the main goal of such colour constancy method as discussed below. Although this assumption works well in most cases and is widely used in commercial cameras, nevertheless there exist common cases in which this assumption is violated in real images, including: skylight from windows plus indoor light; in-shadow plus non-shadow lights; or two different light sources in an indoor room. This situation, multiple illuminants or multiple light sources with different colours, is a common source of failure for current colour constancy methods. Exemplar-based colour constancy has the advantage of succeeding even in the multiple illuminant situation, which is not possible for most current methods. Hence as another contribution we show a technique to overcome the multiple illuminant situation using our proposed method, and test our technique on standard images having two distinct sources of illumination.

The implications for a useful discounting or regularizing for light in images are substantial in various tasks in computer vision.

The outline of the chapter is as follows: we discuss related work in §7.2 and then in §7.3 we introduce the proposed method by explaining our surface model, the process of learning surface models for training images, and the proposed illumination estimation procedure. In §7.4 we apply our proposed method to three standard colour constancy datasets, comparing performance to current colour constancy methods. In §7.5, we review previous multiple illuminant colour constancy methods, discuss our proposed method in the multiple illuminant situation, and carry out experiments. Finally, we conclude the chapter and discuss future work in §7.6.

7.2 Related Works

Illumination estimation methods can categorized into two groups: (1) static methods which try to estimate the illuminant for each image based on its statistical or physical properties and (2) learning-based methods which try to estimate the illuminant using a model that is learned on training images. Grey-based methods, which form a main part of static methods, have been formalized into a single framework [75, 181]:

$$\left(\int \left\| \frac{\partial^n I_k(x)}{\partial x^n} \right\|^p dx \right)^{\frac{1}{p}} = e_k \tag{7.1}$$

where e is estimated illuminant colour, k denotes R, G or B, p denote the Minkowski norm and n is grey-edge order. If $n = 0$, for $p = 1$ the equation is equal to the grey-world assumption, for $p = \infty$ it is equal to colour constancy by White-Patch and it is Shades of Grey and for $1 < p < \infty$. For higher n it is Grey-Edge.

Static colour constancy methods also include some physics-based methods such as methods that use specularity to estimate illuminant chromaticity [165, 48]. Drew et al. [48] present an effective physics-based colour constancy method, called the Zeta Image, which makes use of a log-relative-chromaticity planar constraint involving specular reflection. This method is fast and requires no training or tunable parameters.

One of the first colour constancy methods which estimates the illuminant by a model that is learned on training images is the Gamut Mapping algorithm [79]. It is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours; therefore, colours forming a "canonical" gamut which contains possible colours can be observed under a canonical illumination in a training phase, and an estimate of a test-image illuminant can

be derived by mapping current pixel colours to that canonical gamut. Several extensions have been proposed for gamut mapping algorithms. Colour-By-Correlation [69] is a discrete implementation of gamut mapping, where the canonical gamut is replaced by a correlation matrix.

Another learning-based approach to the illumination estimation problem is the Bayesian approach [150, 93], in which the variability of reflectance and illuminant is modeled as independent random variables. These methods estimate illuminant colour from the posterior distribution condition learned from training images. Here the illuminant prior could be uniform over a subset of illuminants [150] or could be an empirical distribution of illuminants in training images [93]. Other machine learning techniques includes using neural networks [35], in which binarized chromaticity histograms are used to estimate 2D illuminant chromaticity via a neural network system, or support vector regression (SVR) [196].

Besides static colour constancy methods such as Max-RGB, Grey-World, Grey-Edge and Shades-of-Grey, which as mentioned above are based on simple assumptions, recently efforts at fusing these algorithms have generated better performance than for the individual algorithms. One of the first attempts in this direction was carried out by Cardei and Funt [34], who applied a weighted committee mechanism over several of these methods. More complex methods try to learn to select the best algorithm or combination of algorithms for each image using pixel information as well as spatial information, and hence they do not deal with the image as simply a bag of pixels.

As mentioned in [104], these learning-based colour constancy methods that try to find the best or a combination of algorithms for each image using extracted features go through a similar procedure. They extract texture, shape or colour features from sets of training images, and estimate the colour of the illuminant for each image using several statistical illumination estimation algorithms. They then learn a model based on extracted features as well as the error of these estimates compared to known ground truth. This type of model could e.g. learn the set of weights associate with estimates of these illumination estimation algorithms [100, 193] or directly learn the colour of the illuminant [21, 124]. Figure 7.1 shows this procedure in both the training and test phases. It can be stated that the main differences amongst this kind of algorithm are in the feature extraction blocks, where the feature could be simple, such as a colour histogram [20, 21], or edge direction histogram [21], or more complex features such as Weibull features [100, 124, 193], Wiccest features [193], or Wavelet statistics [21].

As an example, Gijsenij and Gevers [100] clustered the images by a k-means algorithm using natural image statistics to characterize the images on the basis of Weibull distribution parameters. They then correspond each cluster with the best single algorithm for training images for that cluster.

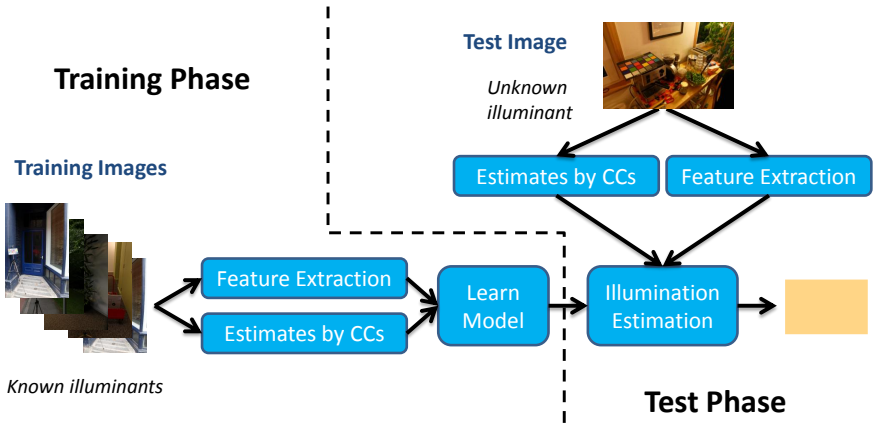


Figure 7.1: The common procedure of learning-based colour constancy methods that try to find the best algorithm or a combination of algorithms for each image using extracted features in both training and test phases.

To estimate the illuminant of a test image, they select the algorithm according to its cluster or combination of the individual algorithms according to the distances to neighbouring clusters.

In a different approach to selecting best algorithms, Wu et al. [193] introduce a multi-resolution texture descriptor based on an integrated Weibull distribution to extract texture information. They used an image similarity measure derived from the Wiccest feature and spatial pyramid matching to find the K most similar images for a test image from amongst training images, and with these neighbouring images they provide a combination for uniting the data-driven strategy and prior knowledge.

Van de Weijer et al. [182] extend the grey world hypotheses to say: the average reflectance of semantic classes in an image is equal to a constant colour, rather than being just grey. Therefore, for each of the semantic classes present in an image they compute that illuminant that transforms the pixels assigned to that class into the average reflectance colour of that semantic class in the training images. They call this a top-down approach as opposed to bottom-up approaches in many other colour constancy methods. They also make use of high-level visual information to select the best illuminant out of a set of possible illuminants generated by other methods. In a similar approach [146], the special visual object categories (called here memory-colour categories) which have a relatively constant colour such as sky or grass and foliage (which were used in their experiment) are detected using the Bag-of-Features machine learning method. Then the initial estimate

provided by a statistical colour constancy method can be adjusted to map the observed colour of the category to its actual colour which is determined in the training phase. The main difference between this work and [182] is that the visual object categories are known and hand labeling and tagging with the category label is required for training images.

In [175], we introduced the exemplar-based colour constancy approach, and in this chapter we go on to investigate the method much more substantially. Firstly, the method's details are set out considerably more comprehensively, and with illustrations. The method is first delineated in the context of previous learning-based approaches. Then entirely new tests are carried out, as well as challenging the method by applying colour constancy via exemplars obtained from one dataset to test images for a different image dataset – an inter-dataset test which is very demanding. Moreover here we go on to investigate how to structure the new algorithm in the face of the very difficult multiple-illuminant situation, and an entirely fresh set of tests on multiple-illuminant images with ground truth is carried out.

7.3 Proposed Method

The proposed method falls into the learning based colour constancy category, in which a model needs to be learned from training images. The main distinctions between this work and other learning based colour constancy methods that use spatial information by local feature descriptors, such as [100, 193, 182], is that they use this information to determined the best or combination of best possible illumination estimation algorithms (the procedure is shown in Fig. 2.7), while we use selected instances for illumination estimation. Compared to the top-down approach [182, 146] in which they assign a semantic classes (or memory-colour categories) to each patch of an image based on models learned in the training phase, in our proposed model we assign to segmented regions we call "surfaces", from training images to each surface of the test image. As well, [182] used the extended version of the grey world assumption to estimate the illuminant whilst we use the ground truth of corresponding surfaces for illumination estimation.

On the other hand, scenes with a single or a just few number of surfaces (such as images captured of grass or sky) are a common failure for grey-based methods which form the core of most recent learning based methods, since the grey assumption is not satisfied for these images. The Gamut Mapping method also fails for these images since only a limited number of colours are seen in the image and that is not enough to map the input gamut to the canonical gamut. We will see that our exemplar-based method can overcome this problem since the exemplar-based method estimates the

illuminant based on similar surfaces and there is no assumption that more than one surface is needed (although more surfaces do make the estimate more robust).

The exemplar-based methods classify or estimate test examples based on examples already seen; this is related to a similar concept in learning in humans. Hence, we can consider exemplar-based methods as belonging to learning-based, notwithstanding the fact that they do not appear to have a similar structure to other learning methods.

7.3.1 Surface Model

We find surfaces for both training and test images by mean-shift segmentation, implemented via [41]. Since the pixels in the margin of segmented areas affect texture information, we remove margin pixels of segments by dilating segment edges as well as small segments. In order to define a model for each surface we use both texture features and colour features. For the purpose of texture features, the MR8 filter bank [183] on three channels is selected for use because of its good performance in texture classification applications [184] and also its fast implementation.

The MR8 filter bank consists of 38 filters (6 orientations at 3 scales for 2 oriented filters, plus 2 isotropic) but only 8 filter responses. The filter bank contains filters at multiple orientations but it records only the maximum filter response across all orientations. We use the normalized histogram of frequency of appearance in that particular surface for each colour channel as our colour features. Since we deal with illumination variation, we divide each channel by its maximum value before computing each histogram. This makes our surface model weakly colour constant [53]. This means that we are not interested in specific colours for our surface matching, but instead on its relative distribution.

In the learning stage, training images are convolved with a filter bank to generate dense filter responses. Exemplar filter responses are chosen as textons via K-Means clustering (with $K = 1000$) and are collected into a dictionary. The histogram of frequency of textons belonging to this dictionary is a common description for texture detection [123] although other local descriptors such as scale-invariant feature transform (SIFT) [127] may used instead of the MR8 filter bank.

Given a surface in a training image, its corresponding model is generated by first convolving it with the filter bank and then labelling each filter response with the Euclidean nearest neighbour texton in the texton dictionary. The histogram of textons, i.e. the frequency with which each texton occurs in that surface, forms the first histogram in the corresponding model for that training surface. We then add a coarse three-channel histogram to that surface's model (we use only 10 bins for each

channel to be robust again noise). In order to make our model weakly invariant to variation in illuminant colour, we stretch the histogram for each channel to have maximum equal to 1 or, equally, divide the values of each colour channel by its maximum value among pixels within that surface (thus making it colour constant for Max-RGB via using a diagonal transformation). Therefore, each surface model includes four normalized histograms that are then stored in a single vector. We also need to store some meta-data for each model, consisting of ground truth illumination colour for that image as well as the maximum response for each channel used for stretching histograms. Fig. 7.2 shows a surface and its normalized histogram of textons and three weakly colour constant normalized histograms of colour channels. In summary, the training phase for exemplar-based colour constancy is expressed in algorithm 7.1. Here $MaxRGB$ function make the input pixels colour constant by the illuminant estimated by Max-RGB method as output.

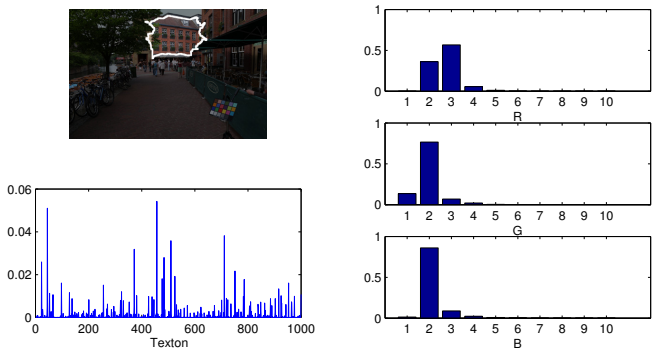


Figure 7.2: A surface and its normalized histogram of textons and three weakly colour constant normalized histograms of colour channels.

In the test stage, the same procedure is followed to build the model (one histogram of textons and three histograms of colour channels) corresponding to each surface in the test image. This model is then compared with the models corresponding to training surfaces by nearest neighbour classifier with the chi-squared statistic employed to measure distances. We select M nearest neighbours from training surfaces (with $M = 10$). Fig. 7.3 shows some test image surface examples and their eight nearest surface models from training data. We carry out our experiment on the re-processed version of the dataset [93, 160] (denoted "ColorChecker") (refer to §7.4 for details). Since we have no labelled or clustered data for our training process as would be the case for a semantic segmentation task or texture detection task, we lack information for providing confidence for our

Figure 7.3: Surfaces from test images (on the left) and their 8 nearest surface models from training images.

mapping. Nevertheless, although we find some amount of mismatching for surfaces as shown in Fig. 7.3, the illumination estimation process simply considers these cases as outliers compared to the other estimates.

7.3.2 Illumination Estimation

Given a test surface model and its nearest neighbour surface model from training models, we can transfer the test surface's colours to its corresponding training surface's colours linearly by a 3×3 matrix. We can approximate this matrix by a diagonal matrix as discussed in [61] and solve the transformation for each channel separately based on their channel histograms. Therefore, we can write this matrix which transforms test surface to training surface as follows:

$$D = \mathcal{M}_{test}^{-1} D_H \mathcal{M}_{train} \tag{7.2}$$

where $\mathcal{M}$ is the weakly colour constant diagonal transformation of surface colour from the Max-RGB method and D_H is the transformation of the test surface's histograms to training surfaces' histograms. Since we use this histogram to find similar models these histograms are usually approximately identical or not far from each other; this means that D_H is approximately the identity matrix. Finally, since the illumination colour of training surface e_{train} is known, the estimation for test surface illumination colour is:

$$e_{test} = D e_{train} = \mathcal{M}_{test}^{-1} D_H \mathcal{M}_{train} e_{train} \tag{7.3}$$

Algorithm 7.1 Training Exemplar Based Model

Generate Texton Dictionary

1: *features* ← *convolve all training images with MR8 filter*
2: *textons* ← *K-Means clustering of features (*$k = 1000$*)*

Finding Surfaces

1: *surfaces* ← *mean-shift seg. of all training images*

Generate Surface Models

1: ***for all*** S *in surfaces* ***do***
2: *features* ← *convolve* S *with MR8 filter*
3: *label* ← *NN(features,textons)*
4: *texture hist* ← *normalized histogram of labels*
5: $S_{cc} \leftarrow MaxRGB(S)$
6: *colour hist* ← *normalized histogram of each colour channel in* S_{cc} *(10 bins)*
7: $trainmodel_S \leftarrow$ *(texture hist, colour hist)*
8: ***end for***

Now we can consider a specific surface under two different illumination conditions (either by capturing images under different lights or by transforming to an image to different illumination conditions by a colour correction technique). We expect to have a similar vector model for corresponding surface models since the texture features are the same and the colour histograms are approximately the same, using our Max-RGB method. On the other hand, if both of these surface models are in out training sets (with similar surface vector), and if they are chosen as nearest neighbour surfaces, the estimation will also be approximately the same since the term $\mathcal{M}_{train}e_{train}$ in equation (7.3) will be approximately equal, since the change e_{train} discounts according to the changes in our Max-RGB estimation $\mathcal{M}_{train}$. As an example if if we synthesize a new surface by applying a 3×3 matrix transfer M to our original surface, we end up by having Me_{train} as the new illuminant ground truth and our Max-RGB illumination estimation multiplied by M, or, equally, our colour-constancy diagonal matrix will be $\mathcal{M}_{train}M^{-1}$; hence this ends up in $\mathcal{M}_{train}M^{-1}Me_{train} = \mathcal{M}_{train}e_{train}$, which is equal to our estimation using the original surface.

Given a test image, we will have n large enough surfaces and M nearest neighbour surfaces from training data, or equally M illumination estimates by eq. (7.3) corresponding to each. The final estimate can be the median or the mean after removing outliers of all of these estimates in rg chromaticity space ($\{r, g\} = \{R, G\}/(R+G+B)$). We can also use weighted averaging by defining weights for each estimated illuminant according to the confidence of estimation for each surface, which we compute based on the standard deviation of estimates for that single surface and also

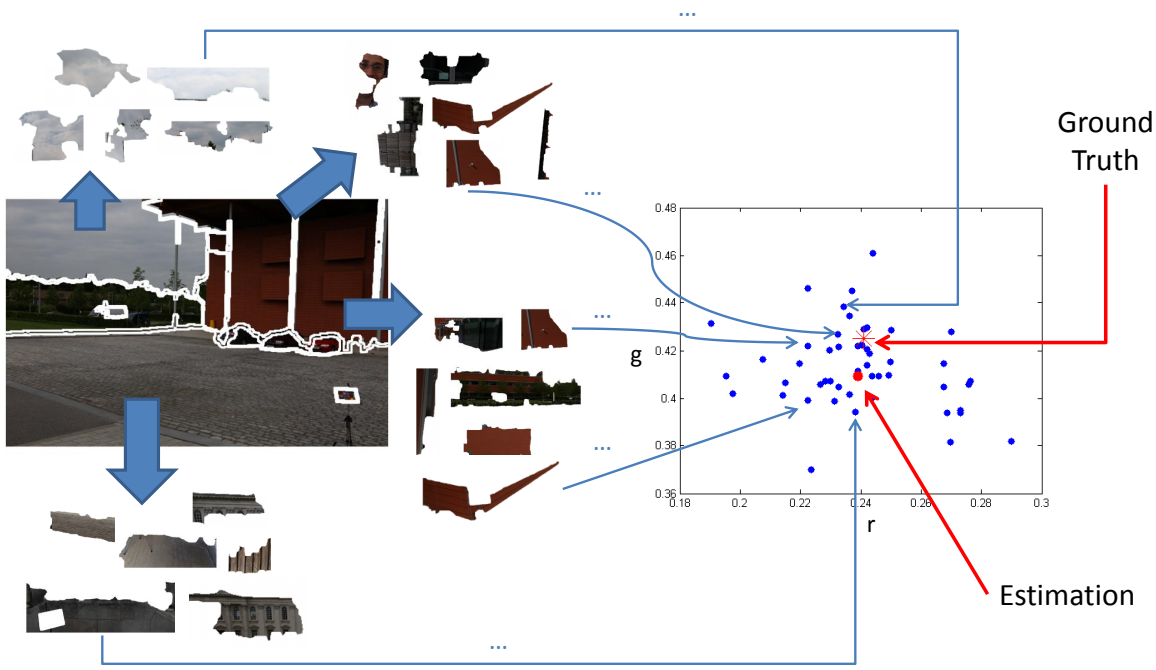


Figure 7.4: The procedure of estimating illuminant for a test image using exemplar-based color constancy. A test image and its nearest neighbour surface models from training images on left and estimated illuminants according to each model in rg chromaticity space on right.

similarity which we compute based on chi squared distance between their normalized histograms. Experiments show that none of these techniques outperforms the others, and therefore for simplicity we estimate the final illuminant by finding the median over all estimates for the three channels separately.

Figure 7.4 shows the procedure for illuminant estimation for a test image using exemplar-based color constancy. Figure 7.5 shows a test image and the angular error of estimated illuminant for its surfaces; as mentioned, we do not form any estimate for small segments. We see that more textured surfaces such as grass or textured road have more precise estimates compared to smooth road. In summary, the proposed illumination estimation for a test image is express in algorithm 7.2.

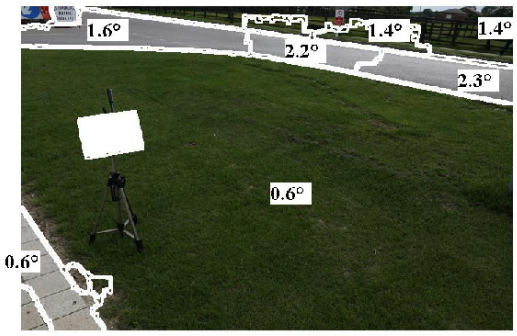


Figure 7.5: A test image and angular errors of estimated illuminant for its surfaces (computation is not carried out for small segments, and the ColorChecker is masked off).

7.3.3 Colour Correction

Once the colour of the light has been estimated, the input image, which was captured under an illuminant as estimated, should be transferred back to an image as it would appear under the canonical illuminant. This procedure is usually done using the diagonal model. Although exemplar based colour constancy estimates a unique illumination for an image, and this can be used for colour correction by a diagonal model, the method actually estimates a distinct illuminant for each surface, and this can be used for colour correction for each surface separately. The diagonal model for pixels that belong to surface edges or to small surfaces can be constructed as a weighted linear combination of those for its neighbour pixels.

We call the matrix in which every pixel is assigned to its own illuminant estimate "back-projection", where this usage derives from [105]. Obviously, in the case when the uniform illumination assumption is made, in our experiment in §7.4, the back-projection is a constant. The colour correction process using estimated back-projections, $\mathcal{D}$, which have equal dimension to the input image I, is as follows:

$$I^*_{ijk} = I_{ijk} * e^*_k / \mathcal{D}_{ijk} \quad , \; k = \{R, G, B\} \tag{7.4}$$

where I^* is colour-transferred images under canonical illuminant e^* and ij is pixel address. Figure 7.6 shows an input image, its estimated back-projection using uniform and surface illumination, and uniform illumination and colour transferred output images using these two back-projections.

Algorithm 7.2 Illumination Estimation of image I

surfaces $\leftarrow$ *mean-shift segment of* I
for all S *in surfaces* ***do***
  features $\leftarrow$ *convolve* S *with MR8 filter*
  label $\leftarrow$ *NN(features, textons)*
  texture hist $\leftarrow$ *normalized histogram of labels*
  $S_{cc} \leftarrow MaxRGB(S)$
  colour hist $\leftarrow$ *normalized histogram of each colour channel in* S_{cc} *(10 bins)*
  $model_S \leftarrow$ *(texture hist , colour hist)*
  for all i *in KNN(*$model_S$,$trainmodels$*)* ***do***
    $estimates_{Si} \leftarrow$ *eq. (7.3)*
  end for
end for
return $median(estimates)$

The estimated back-projection for each surface is computed based on similar known surfaces (similarity for both colour and texture); therefore even if they are not accurate they seem reasonable, as seen in the colour of the grass in Fig. 7.6(d). Even though this colour is not necessarily accurate compared to the known illumination, nonetheless it seems reasonable to us since it is similar to some other grass surfaces in the training images.

(a) (b) (c)

(d) (e)

Figure 7.6: (a) An input image, (b) constant back-projection (c) back-projection by surface illuminant estimates (d) colour corrected image via b and (e) colour corrected image via c.

Table 7.1: Angular errors for original ColorChecker dataset [93] in terms of mean and median for several colour constancy algorithms.

Method	Median Err	Mean Err
Do nothing	6.8°	9.5°
White-Patch	6.0°	8.1°
Grey-World	7.3°	9.8°
Grey-Edge ($p = 4$,$\sigma = 1$)	5.2°	7.0°
Bayesian [93]	4.7°	6.7°
Gamut Mapping pixel ($\sigma = 5$)	4.9°	6.9°
Gamut Mapping 1jet [102] ($\sigma = 5$)	4.9°	6.9°
Bottom-up+Top-down [182]	4.5°	6.4°
Natural Image Statistics [100]	4.5°	6.1°
Zeta [48]	5.2°	7.2°
Exemplar-Based	**3.7**$^\circ$	**5.2**$^\circ$

7.4 Experiments

We applied our proposed method to three standard colour constancy datasets of real images of indoor and outdoor scenes. The first dataset is the Gehler colour constancy dataset [93], denoted the ColorChecker dataset, This dataset consists of 568 images, both indoor and outdoor. The illuminant ground truth for these images is known because each image has a Macbeth ColorChecker placed in the scene — which must masked off in tests. The images are captured by auto white balance setting of the camera. For this dataset, we used three-fold cross-validation to learn our models using this original dataset, as used by other learning based methods we compared to. The second dataset is the re-processed version of the above ColorChecker dataset, provided by Shi and Funt [160]. This dataset which includes the same number of images, but contains raw image data of ColorChecker dataset in an attempt to recover linear sensor values, which is in principle critical for our 3×3 matrix transformation.

Tables 7.1 and 7.2 indicate the accuracy of the proposed methods for the ColorChecker dataset and its re-processed version, in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. For those methods which need tunable parameters, we utilize optimal parameters for this dataset. We see that our exemplar-based method makes a substantial improvement over previous algorithms.

Another dataset, which contains lower quality real images (image resolution 360×240), is

Table 7.2: Angular errors for re-processed version of ColorChecker dataset [160] in terms of mean and median for several colour constancy algorithms.

Method	Median Err	Mean Err
Do nothing	13.5°	13.4°
White-Patch	5.7°	7.4°
Grey-World	6.3°	6.4°
Grey-Edge ($p = 1$,$\sigma = 6$)	4.5°	5.3°
Bayesian [93]	3.5°	4.8°
Gamut Mapping pixel ($\sigma = 4$)	2.5°	4.1°
Gamut Mapping 1jet [102] ($\sigma = 9$)	2.5°	4.1°
Bottom-up+Top-down [182]	2.5°	3.5°
Natural Image Statistics [100]	3.1°	4.2°
Zeta [48]	2.8°	4.1°
Exemplar-Based	**2.3**$^\circ$	**3.1**$^\circ$

the GreyBall dataset of Ciurea and Funt [40]; this contains 11346 images extracted from video recorded under a wide variety of imaging conditions. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of the image (again masked for experiments). In order to learn our models for this dataset, we use 15 folds each of which represents a recorded video as provided by the dataset itself, as for the other learning based methods we compared to. Table 7.3 shows the performance of our proposed method for the GreyBall dataset in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. Again we utilize optimal parameters for this dataset for those methods which need tunable parameters. Here again, we see a substantive improvement over previous approaches, even those using complex methods.

In these experiments, we compare our proposed method to statistical illumination estimation methods (White-Patch, Grey-Word and Grey-Edge) as well as the Bayesian method [93], Gamut mapping and its edge-based extension [102] and the Zeta method which was proposed in chapter 5. Moreover we compare our method to the best learning-based colour constancy methods according to the recent colour constancy survey [104], which are Natural Image Statistics [100] and Bottom-up+Top-down [182]. To our knowledge, for these three standard datasets, widely used for evaluating colour constancy methods, Exemplar-Based Colour Constancy does best in terms of both mean and median angular error compared to any reported colour constancy methods, even those using a combination of algorithms such as Natural Image Statistics [100].

Table 7.3: Angular errors for GreyBall dataset [40] in term of mean and median for several colour constancy algorithms.

Method	Median Err	Mean Err
Do nothing	6.7°	8.3°
White-Patch	5.3°	6.8°
Grey-World	7.0°	7.9°
Grey-Edge ($p = 1$,$\sigma = 1$)	4.7°	5.9°
Gamut Mapping pixel ($\sigma = 4$)	5.8°	7.1°
Gamut Mapping 1jet [102] ($\sigma = 9$)	5.8°	6.9°
Natural Image Statistics [100]	3.9°	5.2°
Zeta [48]	4.6°	5.9°
Exemplar-Based	**3.3**$^\circ$	**4.4**$^\circ$

In Figure 7.7 we show colour-corrected images from GreyBall dataset based on Exemplar-based method compare to Grey-World, Grey-Edge, gamut mapping and Zeta-Image methods, along with their angular error compared to ground truth as obtained from the grey sphere mounted onto the video camera. The proposed exemplar-based method works well overall, and moreover also works significantly better compared to other methods for images of a single surface or a few surfaces, such as a grass scene as in Fig. 7.7: this constitutes a failure case for grey-based methods since the grey assumption is not satisfied, as well as for gamut mapping since only a limited number of colours are seen in the image.

7.4.1 Inter Dataset Cross Validation

Although we have shown excellent performance for exemplar-based colour constancy, outperforming existing methods for the standard colour constancy data sets studied, we would also like to investigate whether our proposed methods also work well for any arbitrary images using a *fixed* learned model, inter-datasets. A three-fold cross validation for the ColorChecker datasets and 15-fold cross validation for the GreyBall dataset were already designed for this purpose, intra-dataset, and we applied them in the tests for all learning based colour constancy methods. However images from the same dataset may be to some degree correlated to each other because of the limitation of gathering image data, and therefore doing cross validation between different datasets is a more challenging task and has not been considered in most papers on learning-based illumination estimation methods.

Figure 7.7: Examples of colour-corrected images from GreyBall dataset based on the Exemplar-based method, compared to Grey-World, Grey-Edge, Gamut Mapping and Zeta methods, along with their angular error compared to the ground truth as obtained from the grey sphere mounted onto the video camera. Here the first column, labelled "Original", is the error for the video frame as compared to that under the canonical illuminant, and is the same as the "Do Nothing" entries in the Tables.

For this purpose, we run our proposed method, exemplar-based colour constancy, for the Grey-Ball dataset, but using surface models learned by images from ColorChecker dataset. The median and mean of angular errors of our illumination estimation for these 11346 images are respectively 5.3° and 6.6°, which is acceptable.

As well, we also ran our proposed method on the ColorChecker dataset, but using surface models learned by images from the Greyball dataset. The median and mean of angular errors of our illumination estimation for these 568 images are respectively 5.1° and 6.5°, again quite acceptable.

Although the results for inter dataset cross validation are not as good as intra dataset cross validation, shown above, they are good enough to convince us that a general surface model is sufficient for estimating the colour of light for any arbitrary image using an arbitrary camera, considering the fact that in comparison either static methods such as Grey-Edge or learning-based methods such as gamut mapping or Bayesian have parameters which must be tuned for each dataset separately. Especially for the two datasets examined, the images differ significantly: the images in the GreyBall dataset were captured from a video recorder and contain approximately 0.1 megapixels (360×240), while the images in the ColorChecker dataset were captured by two high quality DSLR cameras (Canon 5D and 1D) and contain approximately 5 megapixels (813×541 or 874×583). This experiment is a worst-case cross validation scenario; obviously we can set up a different cross validation scenario in which the training and test sets are a *mixture* of these two datasets (ColorChecker and GreyBall sets) where we expect to see performance between that of intra dataset cross validation and inter dataset cross validation.

7.5 Multiple Illuminants

Most colour constancy algorithms assume that the spectral distribution of light is uniform in the image and therefore the colour of the illuminant is constant across the image, Estimation of this single illuminant is the main goal of most colour constancy methods. Although this assumption works well in most cases and is widely used in commercial cameras, there exist common cases in which this assumption is violated in real images. These include: skylight from windows plus indoor light; in-shadow pixels plus out-of-shadow pixels; ambient light and flash-light in photography; and two different light sources in an indoor room. This situation, which we call multiple illuminants or multiple light sources with different colours is a common failure case for current colour constancy methods.

Figure 7.8 shows a scene with two distinct illuminants: outdoor sky-light and indoor luminaire light. Assuming uniform illumination in the scene, the image on the left is colour-corrected using the outdoor sky-light and the image on the left is colour-corrected using indoor light. It is obvious that the colour constancy task, with a uniform-illumination assumption, fails for this scene even when we use the ground truth as our colour-correction illuminant.

Exemplar-based colour constancy has the advantage of working in the multiple illuminant situation, which is not possible for most current methods. White patch find the brighter illuminant, grey-based method as well as grey-edge methods may find combination of illuminant colours, and

Figure 7.8: A scene with two distinct illuminants. The image on the left is colour-corrected using outdoor sky-light and the image on the left is colour-corrected using indoor light.

gamut mapping approximately find the dominant illuminant.

7.5.1 Colour Constancy Methods for Multiple Illuminants

Most colour constancy methods assume a single source of light and uniform illumination, and very few methods explicitly focus on local illuminant estimation. In early research, it is shown in [65, 173] that a difference in illumination, once it has been identified correctly, provides additional constraints which can be applied to obtain better colour constancy; however these works do not provide algorithms to detect multiple illuminants automatically. Ebner [55] assumes a grey-world assumption works locally, and not just globally, and attempts a diffusion based methodology for pixel intensities. However a local grey-world can be noisy and inaccurate, especially for near uniform-colour scenes. In another approach, Bleier et al. [27] propose a method to overcome the multiple illumination problem in which the image is segmented into a set of superpixels based on colour; then a collection of colour constancy algorithms (Bayesian, plus different versions of grey-world, grey-edge and gamut mapping) is applied to each superpixel independently. The illumination estimate for each superpixel is computed using different fusion techniques such as the average, or a machine learning method such as Gradient tree boosting or Random forest regression. Riess et al. [148] apply a physic-based approach (inverse-intensity chromaticity space [165]. refer to 2.3 for more detail) to each superpixel to have a local illumination estimation, Then they group local estimates into regions with consistent and similar illuminant colour and finally obtain a new estimate per region. A local colour constancy algorithm is also presented by [116], which adjusts colours pixel-by-pixel based on its local area to solve the multi-illuminants for High-Dynamic-Range images. This estimates the illumination for each pixel from the colour information of its neighboring

pixels, weighted by the spatial distance, luminance intensity difference, and chromaticity.

A specific case of the multiple-illuminant scenario is encompassed in research on in-shadow and out-of-shadow regions. That is, the shadow removal problem [64, 70] can be considered as a colour constancy problem involving two light sources.

Recently, Gijsenij et al. [105] proposed a colour constancy method for multiple light sources. They obtain image patches by grid-based, keypoint-based, or segmentation-based sampling, and then estimate the illuminant for each image patch by some of the grey-based methods in eq. (7.1), assuming a uniform illuminant. Focusing on scenes with two distinct light sources and their combination, their final estimation of two distinct illuminant colours is either by clustering the illuminant estimates for patches or taking into account the spatial relations between these estimates by applying segmentation.

7.5.2 Exemplar-Based Colour Constancy for Multiple Illumination

As discussed above, given a test image with n surfaces we will have M nearest neighbour surfaces from the training set for each of them, and then finally we have nM different estimates for the colour of illuminant that will end up as our final estimation assuming uniform illumination for the test image. If we have more than one illuminant in the scene, however, we then need to carry out an estimation procedure for such illuminants using our nM estimates. As we already seen, since we are dealing with estimates for each of the test image surfaces, we can have separate estimates for each of these surfaces via the median of the estimates from M nearest neighbour surfaces from the training data. This is in the general case in which we have no knowledge about the number of distinct illuminants in the scene. However, knowing the number of illuminants can make our final estimate more robust and incorporate more resistance to incorrect surface matching, which in itself is inevitable as we showed above for the assumption of a single illuminant in the scene.

As in [105], this task can be done either by clustering the illuminant estimates or by taking into account the spatial relations between these estimates. Knowing the number of illuminants, K, we cluster our estimates into K clusters and make the procedure more accurate by finding the median of each cluster as our final illumination estimation. If two of the illuminants are close to each other it is likely that they become clustered into the same cluster, with extra clusters containing some other estimates. In this case we simply remove the extra clusters by removing small sized clusters; so therefore we may end up having fewer than K final illuminant estimates. In summary, the proposed illumination estimation for a test image with K illuminants is expressed in algorithm 7.3. The final

step is colour correction, in which the RGB value for back-projection of each surface will be the nearest one to the separate estimates for that surface. Figure 7.9 displays an image with two distinct illuminants, all illuminant estimates by exemplar-based method in rg chromaticity space, and two final estimates using clustering.

Algorithm 7.3 Estimation of K Illuminants by Exemplar-Based Method

1: $estimates \leftarrow$ *Algorithm 2*
2: *clusters* $\leftarrow$ *k-means(* $estimates$*,* K *)*
3: *remove the clusters with size less than a threshold:*
clusters $\leftarrow$ *k-means(* $estimates$*, updated* K *)*
4: *illums* $\leftarrow$ *median of each clusters*
5: ***return*** $illums$

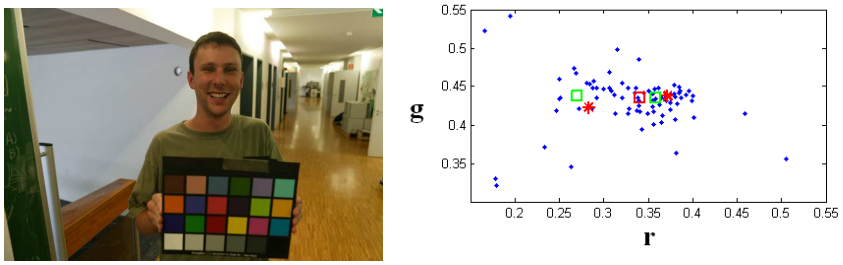


Figure 7.9: An input image with multiple illuminants. All illuminant estimates, using our exemplar-based method, are shown as blue dots in rg chromaticity space. Two red stars are the ground truth for two distinct illuminants in the scene. The red square shows the illuminant as estimated by the proposed method for the single illuminant assumption. And two green squares indicate the two illuminants estimated by the proposed method assuming two illuminants.

7.5.3 Experimental Results

We evaluate our proposed method with the multi-illuminant dataset provided in [105], which includes 9 outdoor low-quality images with two distinct illuminants for each scene. The ground truth for the light sources is provided by several grey balls placed in the scene and is manually annotated for each image. The images differ in size and are of quite low resolution, containing approximately

40 kilopixels. Since we are dealing with scenes with varying illuminations, we need to find the error across the scene. Therefore the angular error compared to ground truth for each pixel is computed and the average angular error throughout the image is considered as our measure.

First we run our proposed method as enunciated in section 7.3 to estimate a single illuminant for each image. Then we again run exemplar-based colour constancy but estimating the illuminant for each surface estimate separately via the median of estimates from their nearest neighbour surface models; thus each surface can have a distinct illuminant and there is no limitation on the number of illuminants in the scene. Finally, we assume there are exactly two illuminants in the scene and run exemplar-based colour constancy for multiple illumination as described in the last subsection, clustering the illuminants into two clusters with the estimates for each surface assigned to one of these two estimates. For all of these three experiments we used the texton dictionary and surface models learned from the GreyBall dataset, for an inter-dataset test.

Table 7.4 shows the median of per-pixel angular error for this dataset assuming a single illuminant and using the White-Patch, Grey-World and Grey-Edge methods; as well as assuming two illuminants with these algorithms with the method proposed in [105], compared to our three experiments as outlined above. For these images with multiple illuminants, almost all methods show significant improvement when including the knowledge that there are two distinct light sources in the image. Although the quality of the images is quite and indeed this may affect our texture features in our exemplar-based surface model, our proposed method work wells assuming a single illuminant is the goal. Moreover using the surface estimates (called Multi in Table 7.4) and our proposed method for multiple illuminants, with a 2-illuminant assumption, the performance of illumination estimation improves respectively by 14% and 25%. As already mentioned, knowing the number of illuminants makes our final estimate more robust and resistant to incorrect surface matching; therefore the exemplar-based method performs better for this dataset when we use the knowledge that the are exactly two illuminants in the scene. Figure 7.10 displays images from this dataset as well as calculated back-projectionimage using three mentioned experiment as well as ground truth.

7.6 Conclusion

In this chapter we present a completely new line of approach to the colour constancy problem which we call Exemplar-Based colour constancy. We use both texture features and weakly colour constant three channel colour values in order to find the nearest neighbour surfaces from training data for each surface, and then we estimate the illuminant for each surface based on histogram matching of each

Table 7.4: Median angular errors for a 9 outdoor image dataset [105] in terms of mean and median angular error, for colour constancy algorithms using a one- or two-illuminant assumption.

No. of Illuminants	Method	Median Err.
One	White-Patch	7.8°
	Grey-World	8.9°
	Grey-Edge (n=1)	6.4°
	Grey-Edge (n=2)	5.0°
Two (from [105])	White-Patch	6.7°
	Grey-World	6.4°
	Grey-Edge (n=1)	5.6°
	Grey-Edge (n=2)	5.1°
One	**Exemplar-Based**	**5.1**°
Two		**3.8**°
Multi		**4.3**°

surface to its candidate nearest neighbour surfaces from training data. The final step is integrating these estimates into a unique illuminant estimation for the whole image. The proposed method has the advantage of overcoming the difficulty of multi-illuminant situations, which is not possible for most current methods. We show that the proposed method performs very well for three standard datasets commonly used in colour constancy tests compared to current colour constancy algorithms.

We also extend our proposed method to overcome the problem of multiple illuminants in the scene by clustering all estimates correspond to nearest neighbour surfaces. The proposed method is shown to work well for a image set of outdoor images with two distinct light sources.

In future, we can apply more complex methods of integrating estimated surface illuminants into a unique illumination estimate. Also, we should construct a dataset for multi-illuminant colour constancy in order to evaluate Exemplar-Based Colour Constancy for images with more than one light source colour, in that this is a uniquely challenging scenario which could produce real benefits in image understanding.

Already, the method provides a real, substantive improvement over current methods, and further tests using inter-dataset calculations are called for to see whether we could generate a standardized much faster pipeline for colour constancy.

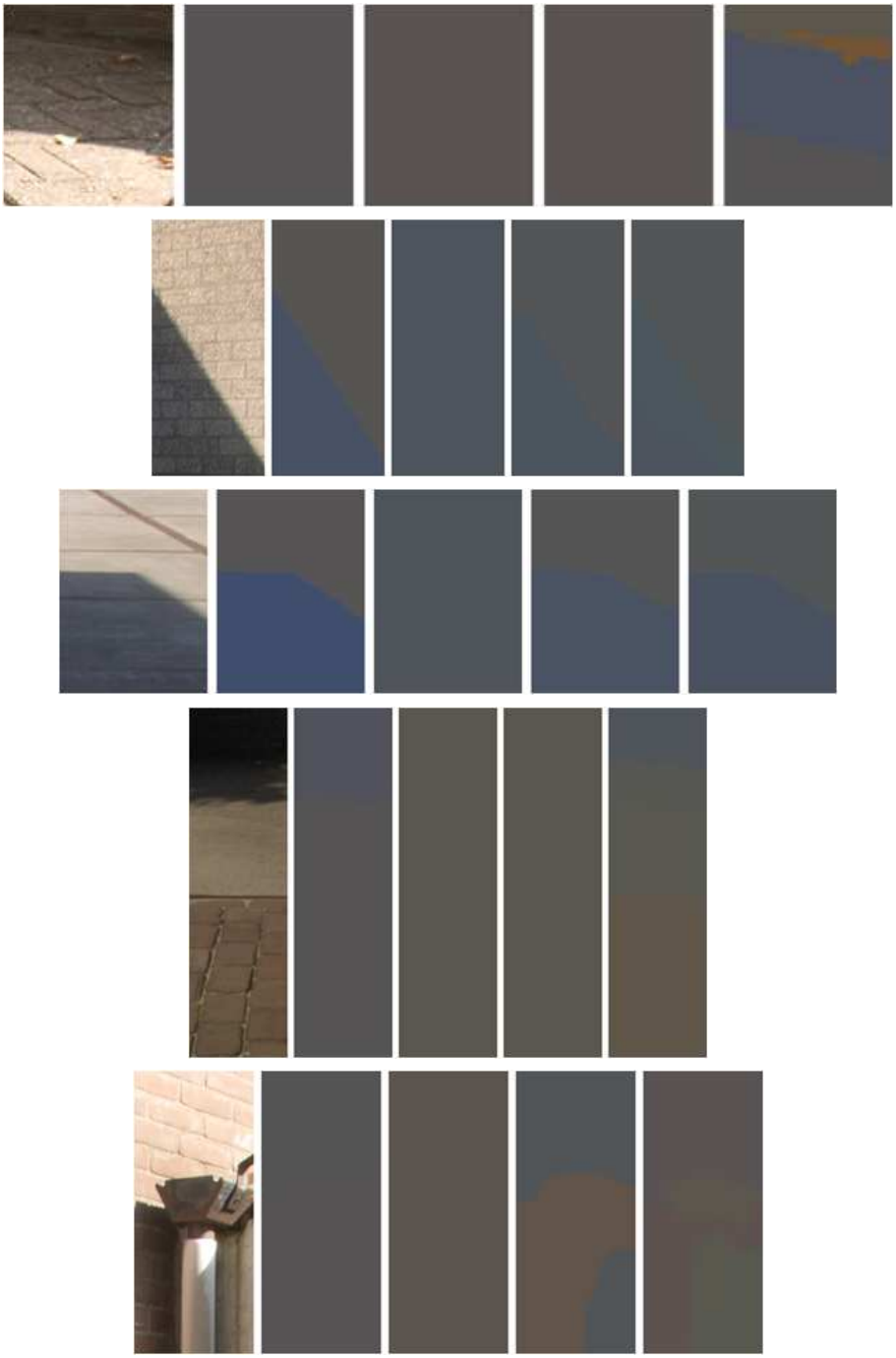


Figure 7.10: Five example images from an outdoor image dataset with two distinct illuminants [105]. For each example, from left to right: original image, the ground-truth back projection (chromaticity of pixelwise illumination) for that image and its estimated value using exemplar based colour constancy assuming single illuminant, two illuminants and illumination estimation for each surface separately.

Chapter 8

Conclusion and Future Work

Many illumination estimation methods have been proposed by researchers, and we categorize them into four general groups: statistical, physics-based, gamut based and learning-based methods. In this thesis, we propose five distinct methods to estimate the colour of illuminant; a statistical method, two physical-based methods, a gamut based method and a learning-based method which includes different approaches to this well known problem.

First, we investigate the effects of bright pixels on several current colour constancy algorithms such as White-Patch, Grey-Word, Grey-Edge and Shades-of-Gray methods. We describe a simple framework for an illumination estimation method based on bright pixels and compare its accuracy to well-known colour constancy algorithms. We also investigate failure and success cases, using bright pixels, and propose desiderata on input images with regard to the proposed method.

Then we use bright pixels in an extension of the gamut mapping colour constancy algorithm, one of the main illumination estimation methods. Specifically, we define the White Patch Gamut as a new extension to Gamut Mapping Colour Constancy, comprising the bright pixels of the image. Adding new constraints based on the possible White Patch Gamut to the standard gamut mapping constraints, a new combined method outperforms gamut mapping methods as well as any of its extensions.

Motivated by the effect of bright pixels in illumination estimation then we turn to using specular reflection, which does tend to be bright. In this direction, we present a new and effective physics-based colour constancy algorithm which makes use of a novel log-relative-chromaticity planar constraint. We call the new feature the Zeta-image. We show that this new feature is tied to a novel application of the Kullback-Leibler Divergence, here applied to chromaticity values instead of probabilities. The new method requires no training data or tunable parameters. Moreover it is simple to

implement and very fast. Our experimental results across datasets of real images show the proposed method significantly outperforms other unsupervised methods while its estimation accuracy is comparable with more complex, supervised methods and it is surprisingly fast. We also show that this new feature can be useful for the purpose of manipulating specular reflection and highlights.

Then we present a new camera calibration method aimed at finding a straight-line locus, in a special colour feature space, that is traversed by daylights and as well also approximately followed by specular points. The aim of the calibration is to enable recovering the colour of the illuminant in a scene, using the calibrated camera. We calculate the Zeta-image feature for points on a proposed straight-line locus to estimate the illuminant. Experimental result shows that using a calibrated locus and the Zeta-image feature we can find the illuminant with accuracy competitive with complex methods even though many test dataset images were not in fact captured under daylight conditions.

And finally, we focus on surfaces in the image and address the colour constancy problem by unsupervised learning of an appropriate model for each training surface in training images, which we call Exemplar-Based colour constancy. We find nearest neighbour models for each surface in a test image and estimate its illumination based on comparing the statistics of pixels belonging to nearest neighbour surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate. The proposed method has the advantage of overcoming multi-illuminant situations, which is not possible for most current methods. We hence also show a technique to overcome the multiple illuminant situation using the proposed method and test our technique on images with two distinct sources of illumination using a multiple-illuminant colour constancy dataset. The concept proposed here is a completely new approach to the colour constancy problem. We show that it performs very well, for standard datasets, compared to current colour constancy algorithms.

8.1 Comparison of our Proposed Methods

Gijsenij et al. [104] propose the following desiderata in their survey on colour constancy methods in assessing computational methods: (1) the requirement for training data; (2) the accuracy of the estimation; (3) the computational runtime of the method; (4) transparency of the approach; (5) complexity of the implementation; (6) number of tunable parameters.

We have already seen the accuracy of the estimation by our proposed methods and other well known colour constancy methods separately in each chapter, Table 8.1 shows angular errors for well known colour constancy algorithms for the SFU Laboratory dataset [14], the ColorChecker

Table 8.1: Angular errors for well known colour constancy algorithms for SFU Laboratory dataset [14], ColorChecker dataset [160] and GreyBall dataset [40] compared to proposed methods in this thesis.

Dataset	SFU Laboratory		Color Checker		Gray Ball	
Methods	Median	Mean	Median	Mean	Median	Mean
Do nothing	6.5°	9.1°	5.7°	7.4°	5.3°	6.8°
White Patch	6.5°	9.1°	5.7°	7.4°	5.3°	6.8°
Gray World	7.0°	9.8°	6.3°	6.4°	7.0°	7.9°
Gray Edge	3.2°	5.6°	4.5°	5.3°	4.7°	5.9°
Bayesian [93]	-	-	3.5°	4.8°	-	-
Gamut Mapping	2.3°	3.7°	2.5°	4.1°	5.8°	7.1°
Natural Image Statistics [100]	-	-	3.1°	4.2°	3.9°	5.2°
Bright Pixel Framework (Ch3)	1.9°	5.8°	2.6°	4.0°	4.7°	5.7°
White Patch Gamut Mapping (Ch4)	1.9°	3.4°	–	–	4.7°	6.0°
Zeta (Ch5)	2.1°	6.2°	2.7°	4.2°	4.7°	5.8°
Camera Calibration (Ch6)	2.4°	5.1°	–	–	4.1°	5.6°
Examplar Based (Ch7)	–	–	2.3°	3.1°	3.3°	4.4°

dataset [160] and the GreyBall dataset [40] compared to proposed methods in this thesis. In Figure 8.1 and Figure 8.2 we show colour-corrected images from the GreyBall dataset based on proposed methods in this thesis: the bright pixel framework, white patch gamut mapping, Zeta, camera calibration and exemplar-based methods, compared to white patch and Grey-World methods, along with their angular error compared to ground truth as obtained from the grey sphere mounted onto the video camera.

Table 8.2 compares the proposed methods in this thesis with other colour constancy algorithms in term of criteria mentioned in [104]. Run time values in this table which are the average run time for a single image are based on experiments set out for the GreyBall dataset, which include low quality images (240×360), for unoptimized Matlab running on a single PC. As we can see bright pixels framework, Zeta method and camera calibration methods are pretty fast while white patch gamut mapping and exemplar-based are computationally expensive. It is notable that in white patch gamut mapping extra constrain make it faster than standard gamut mapping. The main time-consuming task for Exemplar-based method is mean-shift segmentation and feature extraction which make it really slow. The advantages and disadvantages of these method is also mentioned in Table 8.2.

8.2 Contributions of the Thesis

The contribution of this thesis is that we propose five distinct methods to estimate the colour of the illuminant: a statistical method, two physical-based methods, a gamut based method and a learning-based method which includes different approaches to this well known problem. The most important contributions of this thesis are (1) introduction of the Zeta-image as a novel physics-based feature which is tied to the specular reflection for each pixel in the image, and (2) a novel Exemplar-Based colour constancy method which shows the best performance compared to reported results for current algorithms for estimating the colour of illuminant, over three standard colour constancy data sets. The individual contributions of this thesis, including the extent of my personal role in their devising, are discussed separately in the following subsections.

8.2.1 Bight Pixel Framework

First, we investigate the effects of bright pixels on several current colour constancy algorithms such as White-Patch, Grey-Word, Grey-Edge and Shades-of-Gray methods and show that using bright pixels only instead of all pixels makes the estimation for all of these algorithms better or at least equal and not worse in a few cases . Motivated by this, we describe a simple framework by adding a new parameter as threshold of bright pixels for a general grey-based colour constancy method for an illumination estimation method based on bright pixels [178]. We show this simple idea performs surprisingly well if we tune the parameters appropriately for current colour constancy datasets.

This is joint work of myself, Dr. Drew, Dr. Finlayson and his student Perla Troncoso Rey. The idea of investigating the role of bright pixels was developed by all the authors of that paper including myself, and the effect of bright pixels on well-known colour constancy methods and introduction of bright pixels framework was done my myself.

8.2.2 White Patch Gamut Mapping

We use bright pixels in an extension of the gamut mapping colour constancy algorithm, one of the main illumination estimation methods. Specifically, we define the White Patch Gamut [176] as a new extension to Gamut Mapping Colour Constancy, comprising the bright pixels of the image. Adding new constraints based on the possible White Patch Gamut to the standard gamut mapping constraints using convex programming implementation, a new combined method outperforms gamut mapping method as well as any of its extensions. We also show that when there is no white surfaces or specularities in the scene which is a failure for methods using bright pixels combined method has

no feasible solution so we detect failure and use standard gamut mapping for these cases. All this work was created and implemented by myself entirely.

8.2.3 Zeta-Image

We introduce a new and effective physics-based feature called the Zeta-image [48, 50] which is shown to be tied to a novel application of the Kullback-Leibler Divergence, here applied to chromaticity values instead of probabilities. We first prove a planar constraint which holds for near-specular pixels: we show that Log-Relative-Chromaticity values are orthogonal to the light chromaticity. Then we define the Zeta-image ζ (a greyscale image) as the dot-product of the log-relative-chromaticity for each pixel with a putative light direction. Based on the planar constraint the Zeta-image value is small (near zero) for near-specular pixels. The value of the Zeta-image for each pixel has an inverse relation with the amount of specular reflection component for each pixel. Using the Zeta-image we present a novel colour constancy algorithm. The new method requires no training data or tunable parameters. Moreover it is simple to implement and very fast. Our experimental results across datasets of real images show the proposed method significantly outperforms other unsupervised methods while its estimation accuracy is comparable with more complex, supervised methods and yet is surprisingly fast. We also show that this new feature can be useful for the purpose of manipulating specular reflection and highlights.

The general idea of this novel feature was first discussed by Dr. Drew and subsequently by Dr. Finlayson, but was a theoretical construct and not a working algorithm in practice. My idea of using an initial bright pixel set, followed by a filetering scheme for deciding on the solution illuminant estimate and its implementation brings the idea into a working algorithm. In sum, my main contribution was the post-processing method, the hierarchical grid search, and also the analytical algorithm. [1]

8.2.4 Camera Calibration for Daylight Specular-Point Locus

We present a new camera calibration method aimed at finding a straight-line locus, in a special colour feature space, that is traversed by daylights and as well also approximately followed by specular points [49]. The aim of the calibration is to enable recovering the colour of the illuminant in a scene, using the calibrated camera. We calculate the Zeta-image feature for points on a proposed straight-line locus to estimate the illuminant. Experimental results show that using a calibrated locus and

[1] Patent applied for this method.

the Zeta-image feature, we can find the illuminant with accuracy competitive with complex methods even though many test dataset images were not in fact captured under daylight conditions.

The general idea of a daylight specular-point locus is originated from the paper "4-Sensor Camera Calibration for Image Representation Invariant to Shading, Shadows, Lighting, and Specularities" [60] by Dr. Drew and Dr. Finlayson; my contribution was bringing the Zeta-image concept into the preceding illumination estimation method, with an aim as well of generating a re-lighting technique using a calibrated locus. [2]

8.2.5 Exemplar-Based Colour Constancy

Finally, we focus on surfaces in the image and address the colour constancy problem by unsupervised learning of an appropriate model for each training surface in training images, which we call Exemplar-Based colour constancy [175, 177]. We find nearest neighbour models for each surface in a test image and estimate its illumination based on comparing the statistics of pixels belonging to nearest neighbour surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate. The proposed method has the advantage of overcoming multi-illuminant situations, which is not possible for most current methods.

We hence also show a technique to overcome the multiple illuminant situation using the proposed method and test our technique on images with two distinct sources of illumination using a multiple-illuminant colour constancy dataset. The concept proposed here is a completely new approach to the colour constancy problem and different from previous learning-based colour constancy methods. We show that it performs the best, for three standard datasets, compared to results reported for current colour constancy algorithms. All this work was created and implemented by myself entirely.

8.3 Future Work

The main contributions of this thesis are introduction of the Zeta-image and of the Exemplar-Based colour constancy method. Both concepts have been shown to bring new results already as proposed, and moreover are likely fruitful for appropriate extension and further research in future.

[2] This method has reached the internal Invention Disclosure stage in a major company and patent will be applied for presently.

- In this thesis, we show the positive effect of using bright pixels on the process of estimating the colour of the illuminant for well-known statistical colour constancy methods such as Grey-World, Grey-Edge and Shades-of-Grey. We also extend the gamut mapping colour constancy method by defining the white patch gamut, which also improves the performance of this method. As future work, we will try using bright pixels for the more complex, semantic-information-based colour constancy methods.

- We have proposed the Zeta-image as a novel physics-based feature, which can be utilized for a very simple, fast and effective method for identifying the colour of the illuminant that requires no training data or tunable parameters and delivers a straightforward, basic, and novel specularity modification technique. Although this method performs very well, it may fail for scenes without any of white surface, light source or specularity. Future work for illumination estimation by the Zeta-image includes identifying failure cases, either by heuristics or by integrating with other methods. Further applications for this feature as well as its specularity modification technique likely form a rich seam to explore.

- Finally, we proposed a novel Exemplar-Based colour constancy method which shows excellent performance, which estimates the colour of the illuminant for each surface, using nearest neighbour surfaces from training images. We integrate estimated surface illuminants into a unique illumination estimate by a simple estimator such as the median. However, in future we can apply more complex methods of integration and further investigate this new approach to colour constancy.

Table 8.2: Comparison of the proposed methods in this thesis with other colour constancy algorithms in terms of their criteria (type, using training data, run time for a single image of the GreyBall dataset and parameters) and their cons and pros.

Method	Type	Train	Run Time	Params	Advantages	Disadvantages
White-Patch	statistical	-	$5ms$	-	• fast and simple	• not accurate
Grey-World	statistical	-	$5ms$	-	• fast and simple	• not accurate
Grey-Edge	statistical	-	$72ms$	p, n, σ	• fast and simple • somehow accurate	• many parameters
Gamut Mapping	gamut	Yes	$2.19s$	σ	• accurate	• slow • difficult implementation
Bright Pixel Framework	statistical	-	$15ms$	p, n, T	• fast and simple • somehow accurate	• many parameters
WP [a] Gamut Mapping	gamut	Yes	$1.27s$	σ	• better than GM [b] • detect WP failure	• slow
Zeta	physics	-	$85ms$	-	• fast and simple • no tunable parameter	• fail when no WP or specularity
Camera Calibration	physics	Yes	$84ms$	-	• fast • camera calibration	• not so accurate for non daylight
Examplar Based	learning	Yes	$4.91s$	-	• perform the best • multiple illumination	• very slow

[a] White Patch

[b] Gamut Mapping

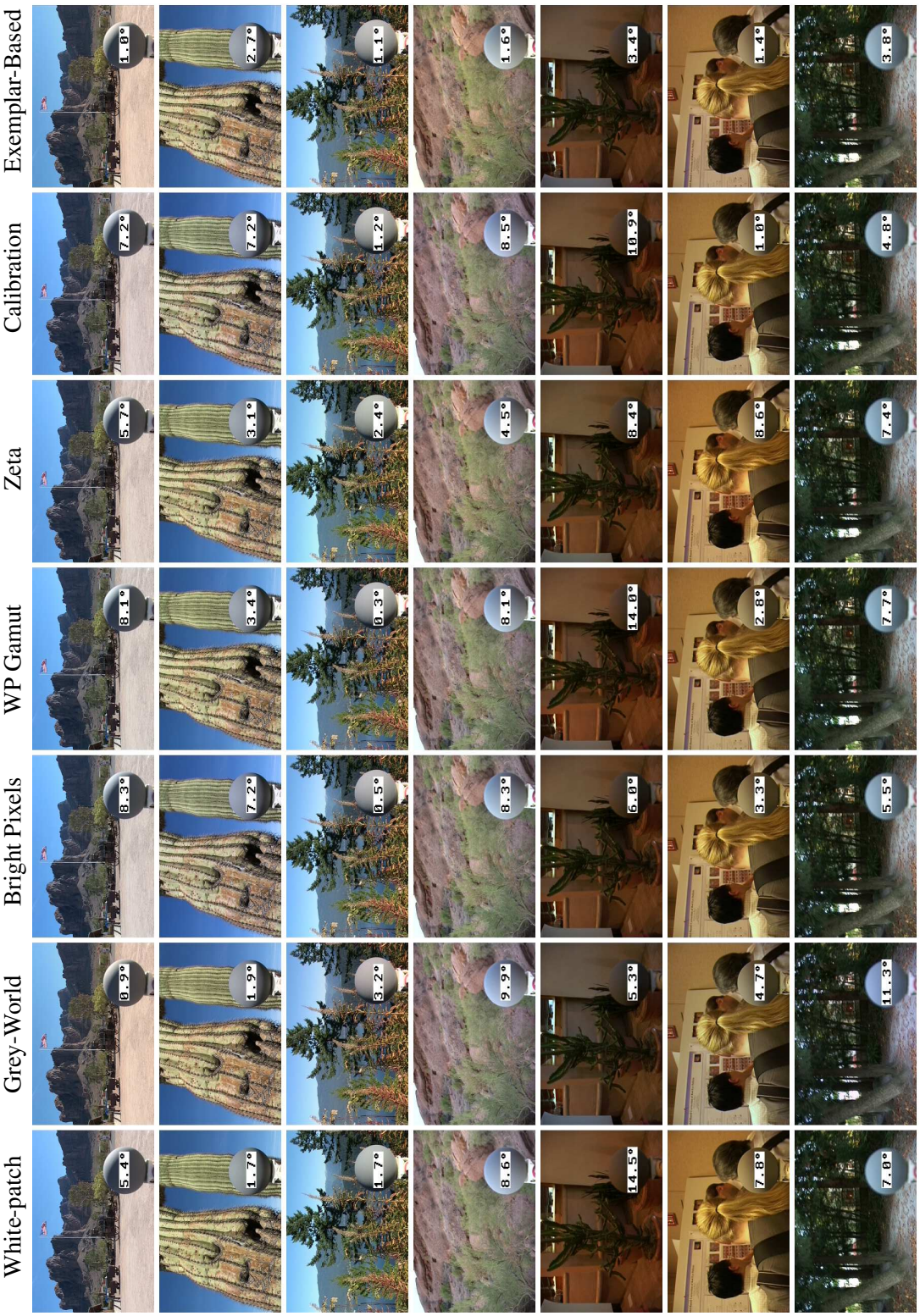


Figure 8.1: Examples of corrected images from GreyBall by White-Patch and Grey-World methods compared to our proposed methods: bright pixel framework, white patch gamut mapping, Zeta, camera calibration and exemplar-based methods and their angular error compare to ground truth.

Figure 8.2: Examples of corrected images from GreyBall by White-Patch and Grey-World methods compared to our proposed methods: bright pixel framework, white patch gamut mapping, Zeta, camera calibration and exemplar-based methods and their angular error compare to ground truth.

Appendix A

Specular Reflection

Specular reflection is the mirror-like reflection of light from a surface. In perfect specular reflection, light from a single incoming direction is reflected into a single outgoing direction. In that theoretical case, the law of reflection states that the direction of incoming light and the direction of reflected outgoing light make the same angle with respect to the surface normal.

(a)

Figure A.1: Example of specular reflection in an image. The figure is taken from [164].

When light illuminants a surface, some light is reflected from the interface between the surface and the air without entering the material. Some penetrates into the surface, is partially absorbed, and partially scattered within the surface material, and then is reflected back to the surface-air interface and into the air again. The first component is usually called the specular or *surface* component of

the reflected light, and the second component is called the diffuse or *body* component. The term *diffuse* is due to the fact that the radiance measured from that part of the reflected light is nearly independent of direction. Figure A.2 shows this reflection process.

For many surfaces of inhomogeneous materials (dielectrics such as plastic), the specular component has a spectral-energy distribution close to that of the incident light, so the colour of the specular component is close to the illuminant colour (in many works they are simply assumed to be the same). This is not true for surfaces made of homogeneous materials such as polished metal.

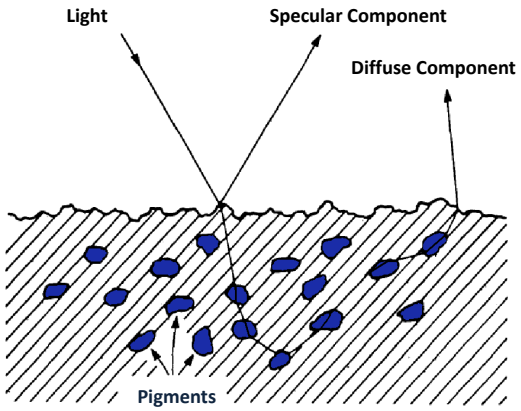


Figure A.2: Difference between specular reflection and diffuse reflection. Here, a dielectric material, such as plastic, is pictured.

Specular reflection is distinct from diffuse reflection, where incoming light is reflected in a broad range of directions. Fig. A.2 demonstrates the distinction between specular reflection and diffuse reflection. An example of the distinction between specular and diffuse reflection can be seen in glossy and matte surfaces. Matte surfaces such as dull painted surfaces, porous surfaces etc. have almost exclusively diffuse reflection, while glossy surfaces such as glass, ceramic, polished metal, and plastic, have both specular and diffuse reflection. However, metals have very different reflection properties from dielectric materials such as plastics.

A.1 Physical Theory

The Fresnel equations describe the transmission of light at an interface between two media of differing refractive properties. When unpolarized light impinges on a point of a smooth and flat surface,

the reflection coefficient is $R = 0.5(R_\perp + R_\parallel)$, where $R_\perp$ and $R_\parallel$ are the reflection of the electric field when it is perpendicular and parallel to the surface, respectively. This coefficient is dependent on the angle between the incoming light and the surface normal and on the refractive properties of the two media. The direction of specular reflection will follow the law of reflection, if the surface is perfectly smooth and flat and behaves as a perfect mirror. This implies that we can only see the specular reflected light if we are at the exact direction as the outcoming reflected light θ_r. However, in most objects we can still observe a certain degree of specularity even when we are somewhat away from the direction of θ_r. Therefore specular reflections do not only form a sharp line (spike) distribution of reflection, but instead form a lobe distribution. As a result, there are at least two components of specular reflections: (1) specular spike and (2) specular lobe. Fig. A.3 illustrated these two specular reflections.

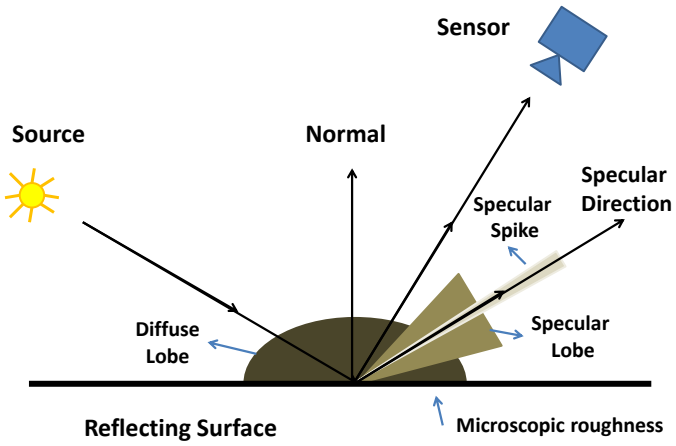


Figure A.3: The polar plot of the reflection components as a function of viewing angle for a fixed source direction.

The Torrance-Sparrow reflection model [172] provides an approximation of the specular lobe component in the form of the following formula:

$$\rho = \frac{FG}{cos\theta_r} exp(-\frac{\alpha^2}{2\sigma^2}) \tag{A.1}$$

where F is the Fresnel reflection coefficient , G is the geometrical attenuation factor, α is the angle between the surface normal and the bisector of the viewing direction and the light source direction, and σ represents the surface roughness.

The Torrance-Sparrow reflection model describes the mechanism of specular reflection by geometric optics, which is only valid when the wavelength of light is much smaller than the roughness of the surface. According to [139], the model uses a slope distribution model to represent the profile of a surface. The surface is assumed to be a collection of planar micro-facets, where their dimension is much larger than the wavelength of incident light. Each micro facet is assumed to be perfectly smooth. The orientation of each facet deviates from the mean orientation of the surface by an angle α. The model considers the masking and shadowing of micro-facets by adjacent facets. Adjacent facets may obstruct flux incident upon a given facet or a flux reflected by it. To include this effect, the geometrical attenuation factor, G, is introduced. The surface roughness, σ, represents the spatial distribution of the lobe. The larger the value of σ, the larger the lobe distribution (implying less-shiny surfaces), and vice versa. In this reflection model, the distribution of the specular reflections follow a Gaussian distribution, with mean α and standard deviation σ. Later, Cook and Torrance [42] replaced the Gaussian distribution with the Beckmann distribution function [17].

The Neutral Interface Reflection (NIR) model [121] formalizes the notion that the spectral power distribution (SPD) of specular reflections is the same as the illumination's SPD; this condition mostly obtains for the surfaces of optically inhomogeneous objects (such as ceramics, plastics, paints, etc.). However it does not usually hold for the surfaces of optically homogeneous objects (such as gold, bronze, copper, etc.) [107].

A.2 Specular Reflection Applications in Computer Vision

Computer vision algorithms such as detection, segmentation, classification and matching have treated specular reflection differently. They may ignore it completely, try to remove its effect, or alternatively use it as a source of information. They mainly use dichromatic reflection model which we discussed in . We classify computer vision algorithms into three categories based on their approach to specular reflection.

1. These methods disregard specularities or treat them as outliers; for Simplification such algorithms are completely based on a diffuse component analysis.

2. Since the presence of specular reflection is ubiquitous in the real world, this second group of methods, while also based only on a diffuse component analysis, do try to remove or reduce the effect of specular reflection. Most of these methods incorporate knowledge about specular reflection to make the above algorithms more robust.

3. Since specular reflection captures important scene information, such as photometric and geometric information, these methods try to analyze the scene considering the effect of specular reflection along with the diffuse component, thus making the model more complex. Generally, photometric information is important for material recognition, and geometric information is crucial for shape recognition.

In this section we review different computer vision algorithms related to specular reflection, including algorithms which separate diffuse and specular reflection components. We will also discuss colour constancy methods which consider specular reflection in the next section under physics-based colour constancy methods.

However, as mentioned above there is a third group of methods that use specular highlights as a source of information for recognition; this group is not included in this chapter because such methods are far from the core topic of this thesis. As an example, Osadchy et al. [142] proposed an algorithm for recognition of objects that have diffuse and specular components that uses specular highlights to determine the shape of object. Chen et al. [37] propose a progressive acquisition system that captures a dense specularity field as the only information for mesostructure reconstruction. Additionally, an image of a perfect mirror-like specular surface is a distortion of the surrounding environment. These images often convey useful shape information, such as the dent in a car, the imperfections in a building's window, or a word that is pressed into a specular sheet [1]. Some methods use this information in order to recover the surface shape. Since it is difficult or even impossible to distinguish between real images and virtual distortions on a specular surface, one needs more than one image to recover the shape. Some methods use stereo images and others use a sequence of images from motion [141, 1, 2]. The latter methods are usually referred to as *shape from specular flow* (SFSF). In this chapter we do not further discuss this kind of shape recovery.

In the remainder of this section we review algorithms separating diffuse and specular reflection components, including algorithms that use polarization and illuminant-dependent colour spaces as well as other algorithms.

A.3 Separation of Diffuse and Specular Reflection

Separation of diffuse and specular reflection components is an essential topic in computer vision. Many algorithms in this field assume perfect diffuse surfaces and ignore specular reflections. However, in the real world, the presence of specular reflection is ubiquitous since there are many dielectric inhomogeneous objects which have both diffuse and specular reflections. It is therefore

necessary to separate the two components robustly and accurately in order to properly acquire the diffuse-only reflection. Moreover, once this separation has been accomplished, the specular reflection component can be used to advantage since it includes useful information about surface properties.

Klinker et al. [113] showed that the colours of all pixels from an object patch form a colour cluster in its dichromatic plane. These colour clusters consist of:

1. Matte pixels, which are scene points that consist only of body reflection. The colours of the matte pixels form a matte line in the colour space, in the direction of the body reflection vector.

2. Highlight pixels, which are scene points that consist of both body reflection and also interface reflection, The RGB colour 3-vectors for these pixels thus form a straight highlight line in the colour space which is parallel to the interface reflection vector. The line starts from the matte cluster at the position representing the body reflection component of the highlight pixels.

3. Clipped colour pixels, which are highlight pixels at which the light reflection exceeds the dynamic range of the camera. In this case the highlight cluster bends near the wall of the colour cube that describes the limit of sensitivity of that colour band.

These colour pixels often cluster in the shape of a *skewed-T*, where the two limbs of the skewed-T correspond to diffuse and specular reflection (Fig. A.4). Gershon et al. [95] also reported the same observation by considering segmented neighbour colour pixels.

Klinker et al. [113] implemented an algorithm that uses the shape of the colour clusters to detect and remove highlights from colour images. Their method projects the pixels of selected image areas into the colour space and fits a dichromatic plane to the colour data from each image area. Then it searches within each dichromatic plane for the matte line, the highlight line, and lines of clipped colours based on their properties. Each colour pixel of the image is then broken up into its reflection components. In order to remove interface reflection, the algorithm then projects the colour of every pixel along the highlight vector onto the matte line. The result is an intrinsic matte image of the scene without highlight. They also use these clusters to estimate the scene illuminant, as we shall describe in §2.3.

Klinker et al. [115] applied the concept of these colour clusters to segment colour images with highlights. They compute an initial, rough description of colour variation in local image areas by investigation the shape of colour clusters for each local area. Principal component analysis (PCA) of pixel values of these local areas in colour space shows if they fall in a point, line, plane or general

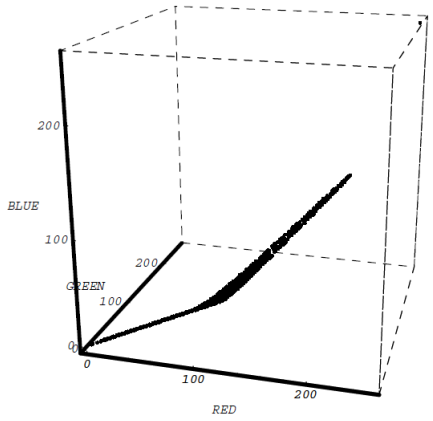


Figure A.4: The skewed-T pixels of a distinct colour surface with both diffuse and specular reflection in RGB colour space. The figure is taken from [155].

space. These authors merge clusters with their neighbours if possible. Moreover, segmentation is carried out over matte colour values after separating reflection components based on their work in [113].

While this basic methods works well for homogeneous, dichromatic surfaces in ideal cases, there are three main limitations that make it difficult to use in real world images. First, many surfaces are "textured", i.e., are not uniform in colour. Even when an image does contain homogeneous surfaces, a non-trivial segmentation process is required to identify them. Second, in order for the specular and diffuse limbs of the skewed-T to be distinct, the specular lobe must be sufficiently narrow, which is often not the case. Finally, when the diffuse and specular colours are the same, there is no way to distinguish between the two components, and no colour separation is possible.

Finlayson and Drew [60] used 4-channel images formed by a special 4-sensor camera to calibrate the camera to develop an image representation invariant to shading, shadows, lighting, and specularities. They first formed logs of colour ratios to reduce the dimensionality to 3 while eliminating light intensity and shading; then projecting into the plane orthogonal to the direction in the 3-D space corresponding to a lighting change direction, they arrived at generalized colour 2-vectors independent of lighting. They noted that in the 2-D space, specularities are approximately linear streaks pointing to a single specular point. Therefore they could remove specularities by the simple expedient of replacing each 2-D colour by the maximum 2-vector position at its particular direction

from the specular point.

Tan and Ikeuchi [164] propose a method which separates specular and diffuse reflection components using a single image. Unlike most other methods their method operates locally. This local operation is useful for handling texture, which is one of the limitations of older methods.

They use the standard diffuse (I^d) + specular (I^s) reflection model. They then define diffuse chromaticity Λ and illumination chromaticity Γ besides standard chromaticity ρ.

$$\rho_k = \frac{I_k}{\sum_{c\in\{1,2,3\}} I_c} \qquad \text{where } \; k = 1 \cdots 3 \tag{A.2}$$

$$\Lambda_k = \frac{I_k^d}{\sum_{c\in\{1,2,3\}} I_c^d} \qquad \Gamma_k = \frac{I_k^s}{\sum_{c\in\{1,2,3\}} I_c^s} \tag{A.3}$$

Based on these definitions, the reflection equation can be rewritten as :

$$I_k = \Lambda_k \sum_{c\in\{1,2,3\}} I_c^d + \Gamma_k \sum_{c\in\{1,2,3\}} I_c^s \tag{A.4}$$

They assumed that the illuminant chromaticity can measured or estimated by an unspecified colour constancy algorithm, so that the input image can normalized such that $\Gamma_1 = \Gamma_2 = \Gamma_3 = \frac{1}{3}$ which means $I_r^s = I_g^s = I_b^s$. They also defined a maximum chromaticity ρ_{max} and maximum diffuse chromaticity Λ_{max} which is the maximum value over the three channels. It is apparent that ρ_{max} and Λ_{max} are limited to the range from $\frac{1}{3}$ to 1.

They show that in a 2-D space of normalized pixel brightness and maximum chromaticity (both known values), all diffuse pixels approximately have equal maximum chromaticity while specular pixels form a curve which is the function of maximum diffuse chromaticity. In Fig. A.5, we can observe that specular points form a curved line in the space. A certain point in the curved line intersects with a vertical line representing the maximum chromaticity of the diffuse point. This curved line follows the following equation:

$$I_k^d = I_k - \frac{max_{\sum_{c\in\{r,g,b\}}} I_c - \Lambda_{max} \sum_{c\in\{1,2,3\}} I_c}{1 - 3\Lambda_{max}} \tag{A.5}$$

To generate a specular-free image, they simply set the diffuse maximum chromaticity equal to an arbitrary scale value for all pixels, regardless of their colour. In this case, a specular-free image may indeed be free of specularity but diffuse chromaticity may differ from the chromaticity of the real

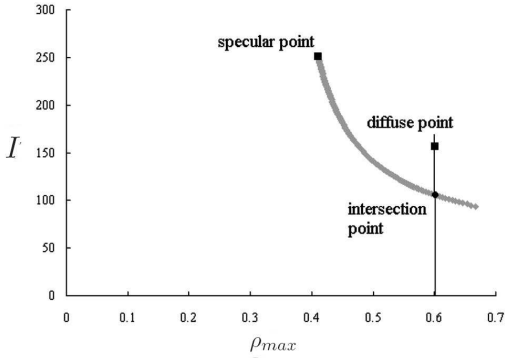


Figure A.5: Specular-to-diffuse mechanism. The intersection point is equal to the diffuse component of the specular pixel. By knowing diffuse chromaticity from diffuse pixels, the intersection point can be obtained. The figure is taken from [164].

image. The method uses the specular-free image to verify whether the original image has diffuse-only pixels, in which case the process terminates. Otherwise the the intensity of specular pixels is decreased from those pixels and they iteratively verify again if it is now diffuse-only pixels.

According to eq. (A.5), finding the diffuse component of reflection or the specularity removal problem can be reduced to searching for the maximum diffuse chromaticity Λ_{max} for each pixel. Moreover the maximum diffuse chromaticity Λ_{max} is the same as maximum chromaticity ρ_{max} except for specular pixels.

Yang et al. [200] assumed that the variance of Λ_{max} is very small in local patches that belong to the same surface colour. So local patches with consistent maximum chromaticity can be considered as non-specular pixels. They apply a low-pass filter to the maximum chromaticity ρ_{max}, to smooth out the variances due to specular highlights. Since the filter should preserve edges, they employ joint bilateral filtering [166] iteratively for smoothing and in order to diminish the effect of specular pixels in diffuse components during smoothing, they use λ as smoothing guidance:

$$\lambda_k = \frac{\rho_k - \rho_{min}}{1 - 3\rho_{min}} \tag{A.6}$$

For more detail about the relationship between λ and Λ_{max} please refer to [200]. They show that their method [200] normally converges in 2 to 3 bilateral filter iterations, which is 200 times faster than [164] while their results are competitive.

Besides algorithms which try to separate reflection components there are some methods which try to remove the effect of specular reflection or highlights using image inpainting. Inpainting is a technique for filling in an image region by propagating information into areas with missing data from the region boundaries. Unlike occluded image regions filled by inpainting, highlight pixels contain some useful information for guiding the inpainting process. As an example, Tan et al. [163] employed a total variation (TV) form of inpainting [152] that incorporates the illumination-based constraint of the dichromatic model. Then their inpainting solution is found by minimizing an energy function over the highlight region.

A.4 Polarization

Polarization is defined as the process or state in which rays of light exhibit different properties in different directions, especially the state in which the direction of electromagnetic oscillation takes place in one plane. Refer to [28] for detailed information.

If the source light is polarized by a linear polarizer filter, the diffuse component of reflection tends to be unpolarized [190]. In contrast, the specular component tends to be partially polarized; rotation of the polarization filter varies the specular component according to a cosine function [190]. Capturing more than one image of a scene with different polarizer orientations provides useful information about diffuse and specular components that may be use in different methods to separate these two components. Although polarization is a strong tool for separating reflection components, it has some restrictions such as the necessity of polarizer filters and multiple images.

Wolff and Boult [191] express the specular component as the sum of a specular constant I_{sc} and a specular varying term that is a cosine function with amplitude I_{sv}:

$$I = I_d + I_{sc} + I_{sv} cos2\,(\theta - \alpha) \tag{A.7}$$

where θ represents the angle of the polarization filter and α is the phase angle determined by the projection of the normal of the surface element onto the plane of the polarization filter. The exact values of I_{sc} and I_{sv} depend on the material properties and the angle of incidence. This dependence is determined by the Fresnel reflection coefficients $F_{\perp}$ and $F_{\parallel}$ which represent the polarization of the reflected light waves in the directions perpendicular and parallel to the plane of incidence, respectively.

They show that three discrete filter positions, θ_i and I_i, are enough to solve the linear system of equations from eq. (A.7) to obtain the parameters $I_d + I_{sc}$, I_{sv}, and α. For more than three discrete

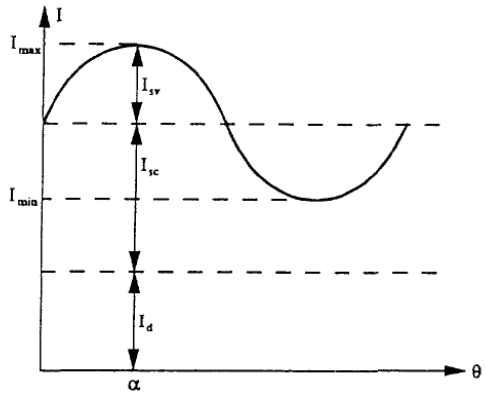


Figure A.6: Image brightness as a function of polarization filter position. The figure is taken from [138].

filter positions, we have an over-determined linear system that can be solved to obtain more robust estimation. They then use the degree of polarization [28], ρ, to classify points that are only diffuse ($\rho << 1$) and those that include a specular component.

$$I_{min} = I_d + I_{sc} - I_{sv} \qquad I_{max} = I_d + I_{sc} + I_{sv}$$

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{A.8}$$

Wolff and Boult's polarization-based method [191] assumes that the diffuse component is constant over the entire highlight region. They also assume that the material type and surface normal do not vary within the highlight region; these assumptions are often not practical in the context of real scenes.

Nayar et al. [138] presented an algorithm for the separation of specular and diffuse reflection components from images which uses colour and polarization simultaneously, to obtain new constraints on the reflection components. They first compute a degree of polarization, ρ by eq. (A.8) for each of the three colour bands. If the largest of three ρ estimates is less than a threshold value, the point is not sufficiently polarized and is assumed to be purely diffuse. For pixels which are not purely diffuse, the specular component varies along a straight line since the cosine functions in the

three colour bands are in phase while the diffuse component is unaffected (eq. (A.7)). This line obviously passes through I_{max} and I_{min}; therefore we have this line for each pixel which has specular component. The diffuse component of this pixel corresponds to a point on that straight line (I_d in eq. (A.7)).

Though they were unable to compute the diffuse component locally, the specular line gives useful constraints on the diffuse component. They use neighbouring image points where the diffuse component is already computed as an estimation of a pixel's diffuse part. Then, the line passing through neighbouring diffuse values from the origin and the specular line intersect to give a more accurate estimate of the diffuse component.

A.5 Illuminant-Dependent Colour Spaces

Recently, there has been some work involving defining colour space transformations that exploit knowledge of the illuminant colour to provide more direct access to the diffuse information in an image. When the illuminant colour is known, and the reflectance of surfaces can be represented by the dichromatic model, we can linearly transform the space of RGB vectors in a way that isolates specular reflection effects. Following the transformation, one, or two, channels are free of these effects, and such a resulting image thus constitutes a specular invariant. Moreover, if this operation is linear, the diffuse shading information is preserved by the transformation and the invariant can be exploited photometrically. Having a colour space with one illuminant-direction axis is ideal for detecting or removing specularities since we assume that specularity and highlights have illumination colour. However, knowledge of the illuminant colour for these colour space is a severe restriction. Although many illumination estimation methods have been proposed, error in such estimation is inevitable.

As discussed above, Tan and Ikeuchi [164] obtain a one-channel diffuse image through the transformation

$$I_d = \frac{3\max_k(I_k/e_k) - \Sigma_k I_k/e_k}{3\lambda^{-1}} \tag{A.9}$$

where $k \in \{1, 2, 3\}$, and the bounded quantity $1/3 < \lambda < 1$ is chosen arbitrarily. This transformation yields a positive monochromatic diffuse image, I_d, which is specular-free and depends directly on diffuse shading information. To show that I_d is independent of illuminant $\mathbf{e}$ let us assume I_k/e_k is maximized for the R channel, or equally $\max_k(I_k/e_k) = I_1/e_1$; then we can simplify I_d using the dichromatic equation, with the theoretical narrowband assumption in eq. (2.15):

$$\begin{aligned} I_d &= \frac{3I_1/e_1 - \Sigma_k I_k/e_k}{3\lambda^{-1}} = \frac{2I_1/e_1 - I_2/e_2 - I_2/e_2}{3\lambda^{-1}} \\ &= \frac{2m_b(\theta)s_1 + 2m_2(\theta) - m_b(\theta)s_2 - m_s(\theta) - m_b(\theta)s_3 - m_s(\theta)}{3\lambda^{-1}} \\ &= \frac{2m_b(\theta)s_1 - m_b(\theta)s_2 - m_b(\theta)s_3}{3\lambda^{-1}} \end{aligned} \tag{A.10}$$

which is independent of illuminant $\mathbf{e}$ and its geometric factor $m_s(\theta)$ and is specular-free.

An alternative transformation is proposed by Park [143], who isolates two predominantly diffuse channels while retaining a colour space similar to HSI. The transformation is composed of a linear transformation L_p and rotation R_p which is demonstrated in Fig. A.7:

- Transform the RGB cube into the Cartesian coordinate frame denoted "XYZ" — the transformed RGB cube within XYZ space will simply have its grey axis vertically aligned with the Z-axis.

- Rotate the RGB cube in XYZ space until the illumination axis vertically coincides with the Z -axis. Then

$$I_p = R_p L_p I \qquad with \qquad R_p L_p e = [0\ 0\ 1] \tag{A.11}$$

The matrices R_p and L_p are chosen such that the third colour axis is aligned with the illumination colour as mentioned before. Park called the new colour coordinates, $X_\alpha Y_\beta Z_\gamma$. As a result, the Z_γ channel contains the majority of the specular component, leaving the other two channels predominantly diffuse. Another major significance of the $X_\alpha Y_\beta Z_\gamma$ space is that all dichromatic planes are orthogonal to the $X_\alpha Y_\beta$ plane.

Park simply projects the pixel values in the new colour coordinates to the $X_\alpha Y_\beta$ plane in order to segment colour images having highlights. As mentioned earlier, in perfect dichromatic reflectance all dichromatic planes are orthogonal to the $X_\alpha Y_\beta$ plane, so each coloured object projected to the $X_\alpha Y_\beta$ plane forms a straight line passing through the center of the plane. Unfortunately, however, real images do not always comply with the dichromatic reflection model.

Another transformation, proposed by Mallick et al. [132], defines a colour space referred to as SUV colour space. The transformation is written

$$I_{SUV} = RI \qquad \text{with} \qquad Re = [1\ 0\ 0] \tag{A.12}$$

Here, they choose $R = R_G(-\theta_S)R_B(\phi_S)$ where $R_k(\theta)$ is a right-handed rotation about the k-axis by angle θ, and (θ_S, ϕ_S) are the elevation and azimuth angles of the source vector S in the

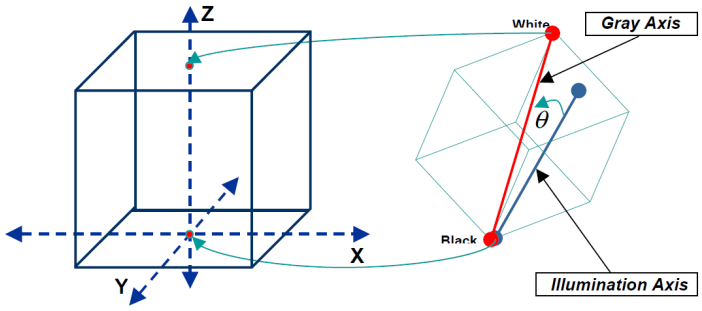


Figure A.7: Colour space transformation to $X_\alpha Y_\beta Z_\gamma$. The figure is taken from [143].

RGB coordinate system. Fig A.8 shows this transformation. Letting r_i^T denote the i^{th} row of R, considering the dichromatic model eq. (2.14) the diffuse UV channels are

$$I_U = r_2^T s f_d \hat{n}.\hat{l} \qquad I_V = r_3^T s f_d \hat{n}.\hat{l} \tag{A.13}$$

Similar to Park's transformation [143], one of the transformed axes in SUV space is aligned with the illuminant colour. Unlike Park's transformation, however, this channel includes the complete specular component, leaving the remaining two channels to be purely diffuse. Unfortunately, to develop SUV colour space we must already know the illuminant colour, by some unspecified means.

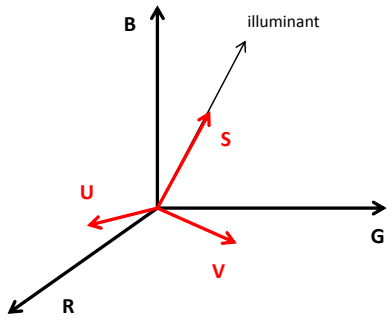


Figure A.8: Transformations of RGB colour space to SUV colour space. Three observations of the same material yield colour vectors I_1, I_2 and I_3 in the dichromatic plane spanned by the source and diffuse colours S and D.

Zickler et al. [203] attempt to generalize SUV colour space [132] for more complex light conditions. Unlike most other specular invariants which assume that illuminant colour is the same in all directions, they consider mixed-illumination environments. They show that, given an M-channel (possibly hyper-spectral) image and N incident illuminations, there exists a subspace of dimension $(M - N)$ that is independent of all these illuminant colours, and therefore invariant to specular reflections. For example, with $N = 2$, such as an office environment where the illumination is a mixture of daylight and fluorescent lights, there is a line invariant to specular reflections in RGB colour space for captured images.

Mallick et al. [131] presented a unified framework for separating specular and diffuse reflection components in images and videos of textured scenes by evolving a partial differential equation (PDE) that iteratively erodes the specular component at each pixel. They re-parameterized SUV colour space [132] in equation (A.13) using a combination of cylindrical and spherical coordinates:

$$\rho = \sqrt{I_U^2 + I_U^2} \qquad \theta = tan^{-1}\left(\frac{I_U}{I_V}\right) \qquad \phi = tan^{-1}\left(\frac{I_S}{\rho}\right) \tag{A.14}$$

Therefore ρ and θ are independent of specular reflection components, ρ represents diffuse shading and ϕ is the linear combination of specular and diffuse component $\phi = \phi_s + \phi_d$. Hence, the problem of computing a specular/diffuse separation is reduced to estimating $\phi_d(x, y)$, the diffuse contribution to ϕ at each image pixel. Once the scalar function $\phi_d(x, y)$ is known, the RGB diffuse component follows directly from inverting the transformations in eqs. (A.13) and (A.14). This will be accomplished by evolving a PDE that iteratively erodes the specular contribution to ϕ and converges to an estimate ϕ_d at each pixel. The erosion process is guided locally by the diffuse colour information provided by ρ and θ, and is formulated in the continuous domain using one of a family of non-linear PDEs that define multi-scale erosion [31].

They show that the PDE governing the evolution of ϕ for the three different cases of texture-less images, textured images, and video, can all be written as

$$\epsilon_t = -g(\rho, \nabla\rho)(\nabla\epsilon^T M \nabla\epsilon)^{1/2} \tag{A.15}$$

where M is a different matrix for each case and $g(\rho, \nabla\rho)$ is called the stopping function. There are two cases when erosion should stop, which are distinguished by stopping function:

- white surface or surface with colour equal that of the light: ρ is zero for this case, so there is no diffuse colour information.
- boundaries between regions of distinct colour: $\|\nabla\rho\|$ is usually large in this case.

A possible stopping function that is zero for both of these cases is

$$g(\rho, \nabla\rho) = \left(\frac{1 - e^{-\rho}}{1 + e^{-\rho}}\right) \frac{e^{-(\|\nabla\rho\| - \tau)}}{1 + e^{-(\|\nabla\rho\| - \tau)}} \tag{A.16}$$

A.6 Summary

In this appensix we describe the physical aspect of specular reflection compared to diffuse reflection. We reviewed algorithms which separate diffuse and specular reflection components; some of these use single images and some use multiple images. We introduced a group of algorithms that separate diffuse and specular reflection components by making use of polarization, and another group which develops illuminant-dependent colour spaces. Nonetheless, the problem of separating diffuse and specular reflection components accurately and robustly is still an open problem, although a great deal of research effort is presently focussed on this problem.

Bibliography

[1] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Toward a theory of shape from specular flow. In *the IEEE International Conference on Computer Vision*, 2007.

[2] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Shape from specular flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2054–2070, 2011.

[3] V. Agarwal, A. Gribok, A. Koschan, B. Abidi, and M. Abidi. Illumination chromaticity estimation using linear learning methods. *Journal of Pattern Recognition Research*, 4(1):92–109, 2009.

[4] V. Agarwal, A. Gribok, A. Koschan, and M. Abidi. Estimating illumination chromaticity via kernel regression. In *IEEE International Conference on Image Processing*, pages 981–984, 2006.

[5] D.W. Aha, D. Kibler, and M.K. Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.

[6] N. Akkiraju, H. Edelsbrunner, M. Facello, P. Fu, E. Mucke, and C. Varela. Alpha shapes: Definition and software. In *1st International Computational Geometry Software Workshop*, pages 63–66, 1995.

[7] L. Arend and A. Reeves. Simultaneous color constancy. *The Journal of the Optical Society of America A*, 3(10):1743–1751, 1986.

[8] K. Barnard. Improvements to gamut mapping colour constancy algorithms. In *European conference on computer vision*, pages 390–403, 2000.

[9] K. Barnard, F. Ciurea, and B. Funt. Sensor sharpening for computational color constancy. *Journal of the Optical Society of America A*, 18(11):2728–2743, 2001.

[10] K. Barnard and B. Funt. Experiments in sensor sharpening for color constancy. In *Proceedings of the 6th IS&T/SID Color and Imaging Conference: Color Science, Systems and Applications*, 1998.

[11] K. Barnard and B. Funt. Color constancy with specular and non-specular surfaces. In *Proceedings of the 7th IS&T/SID Color and Imaging Conference*, 1999.

[12] K. Barnard, B. Funt, and V. Cardei. A comparison of computational color constancy algorithms, part one; theory and experiments with synthetic data. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002.

[13] K. Barnard, A. Coath L. Martin, and B. Funt. A comparison of computational color constancy algorithms, part 2; experiments with images. *IEEE Transactions on Image Processing*, 11(9):985–996, 2002.

[14] K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for colour research. *Color Research and Applications*, 27:147–151, 2002.

[15] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. In *ICCV'01: International Conference on Computer Vision*, pages II: 283–290, 2001.

[16] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

[17] P. Beckmann and A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. Pergamon Press, 1963.

[18] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *Neural Information Processing Systems*, pages 831–837, 2000.

[19] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

[20] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Improving color constancy using indoor-outdoor image classification. *IEEE Transactions on Image Processing*, 17(12):2381–2392, 2008.

[21] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Automatic color constancy algorithm selection and combination. *Pattern Recognition*, 43(3):695–705, 2010.

[22] S. Bianco, F. Gasparini, and R. Schettini. Combining strategies for white balance. In *Digital photography III. IS&T.*, 2007.

[23] S. Bianco, F. Gasparini, and R. Schettini. Region-based illuminant estimation for effective color correction. In *Proceedings of the International Conference on Image Analysis and Processing*, pages 43–52, 2009.

[24] S. Bianco and R. Schettini. Color constancy using faces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[25] M.J. Black, G. Sapiro, and D. H. Marimont. Robust anisotropic diffusion. *IEEE Trans. Image Processing*, 7:421–432, 1998.

[26] A. Blake and G. Brelstaff. Geometry from specularities. In *ICCV'88: International Conference on Computer Vision*, pages 394–403, 1988.

[27] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup. Color constancy and non-uniform illumination: Can existing algorithms work? In *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on*, pages 774–781, 2011.

[28] M. Bom and E. Wolf. *Principles of Optics*. London:Pergamon, 1965.

[29] C.F. Borges. Trichromatic approximation method for surface illumination. *The Journal of the Optical Society of America A*, 8:1319–1323, 1991.

[30] M. Brill. Minimal von kries illuminant invariance. *Color Research and Application*, 33(4):320–323, 1986.

[31] R. Brockett and P. Maragos. Evolution equations for continuous scale morphology. *IEEE Trans. on Sig. Proc.*, 42(1):3377–3386, 1994.

[32] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.

[33] G. Burghouts and J.M. Geusebroek. Performance evaluation of local colour invariants. In *Comput. Vision Image Understanding*, volume 113, pages 48–62, 2005.

[34] V. Cardei and B. Funt. Committee-based color constancy. In *IS&T/SID Color Imaging Conference*, pages 311–313, 1999.

[35] V. Cardei, B. Funt, and K. Barnard. Estimating the scene illumination chromaticity using a neural network. *The Journal of the Optical Society of America A*, 19(12):2374–2386, 2002.

[36] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy beyond bags of pixels. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008.

[37] T. Chen, M. Goesele, and H. P Seidel. Mesostructure from specularity. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1825–1832, 2006.

[38] H. Chong, S. Gortler, and T. Zickler. The von kries hypothesis and a basis for color constancy. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[39] A. Choudhury and G. Medioni. Perceptually motivated automatic color contrast enhancement. In *CRICV09: Color and Reflectance in Imaging and Computer Vision Workshop*, pages 1893–1900, 2009.

[40] F. Ciurea and B. Funt. A large image database for color constancy research. In *IS&T/SID Color Imaging Conference*, pages 160–164, 2003.

[41] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[42] R. Cook and K. Torrance. A reflectance model for computer graphics. *ACM Computer Graphics*, 15(4):307–314, 1981.

[43] J.L. Dannemiller. A test of color constancy in 9- and 20-week-old human infants following simulated illuminant changes. *Developmental Psychology*, 25(2):171–184, 1989.

[44] P.B. Delahunt and D.H. Brainard. Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision*, 4(3):57–81, 2004.

[45] J.M. DiCarlo, F. Xiao, and B.A. Wandell. Illuminating illumination. In *Color Imaging Conf.*, pages 27–34, 2001.

[46] S. Dorr and C. Neumeyer. The goldfish–a colour-constant animal. *Perception*, 25(2):243–250, 1996.

[47] M.S. Drew and G.D. Finlayson. Multispectral processing without spectra. *The Journal of the Optical Society of America A*, 20(7):1181–1193, July 2003.

[48] M.S. Drew, H.R. Vaezi Joze, and G.D. Finlayson. Specularity, the zeta-image, and information-theoretic illuminant estimation. In *CPCV2012: European Conference on Computer Vision Workshop on Color and Photometry in Computer Vision*, 2012.

[49] M.S. Drew, H.R. Vaezi Joze, and G.D. Finlayson. Camera calibration for specularity removal and illuminant determination. *The Journal of the Optical Society of America A*, 2013. Submitted for publication.

[50] M.S. Drew, H.R. Vaezi Joze, and G.D. Finlayson. The zeta-image: illuminant estimation and specularity removal. *IEEE Transactions on Image Processing*, 2013. Submitted for publication.

[51] M.S. Drew and H.R. Vaezi Joze. Sharpening from shadows: Sensor transforms for removing shadows using a single image. In *Proceedings of the 17th IS&T/SID Color and Imaging Conference*, 2009.

[52] M.S. Drew and H.R. Vaezi Joze. Planckian regression temperature for least spectral error and least cielab error. *The Journal of the Optical Society of America A*, 28:1954–1961, 2011.

[53] M.S. Drew, J. Wei, and Z.N. Li. Illumination–invariant color object recognition via compressed chromaticity histograms of color–channel–normalized images. In *the Int. Conf. on Computer Vision*, pages 533–540, 1998.

[54] M.T. Dzhafarli, V.V. Maksimov, A.R. Kezeli, and N.B. Antelidze. Color constancy in monkeys. *Sensory Systems*, 5(3):200–204, 1991.

[55] M. Ebner. Color constancy using local color shifts. In *European Conference in Computer Vision*, pages 276–287, 2004.

[56] M. Ebner. *Color Constancy*. John Wiley & Sons, 2007.

[57] M. Ebner. Color constancy based on local space average color. *Machine Vision and Applications*, 20(5):283–301, 2009.

[58] M. Fairchild. *Color Appearance Models*. Wiley-IS&T Series in Imaging Science and Technology. John Wiley & sons, Chichester, UK, 2nd edition, 2005.

[59] G.D. Finlayson. Color in perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:1034–1038, 1996.

[60] G.D. Finlayson and M.S. Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *ICCV'01: International Conference on Computer Vision*, pages II: 473–480. IEEE, 2001.

[61] G.D. Finlayson, M.S. Drew, and B. Funt. Color constancy: diagonal transforms suffice. In *IEEE International Conference on Computer Vision*, pages 164–171, 1993.

[62] G.D. Finlayson, M.S. Drew, and B. Funt. Color constancy: generalized diagonal transforms suffice. *Journal of the Optical Society of America A*, 11(11):3011–3019, 1994.

[63] G.D. Finlayson, M.S. Drew, and B. Funt. Spectral sharpening: sensor transformations for improved color constancy. *The Journal of the Optical Society of America A*, 11(5):1553–1563, May 1994.

[64] G.D. Finlayson, M.S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *ECCV 2004: European Conference on Computer Vision*, pages 582–595, 2004. Lecture Notes in Computer Science Vol. 3023.

[65] G.D. Finlayson, B. Funt, and K. Barnard. Color constancy under varying illumination. In *Proceedings of the Int. Conf. on Computer Vision*, pages 720–725. IEEE, 1995.

[66] G.D. Finlayson and S. Hordley. Selection for gamut mapping colour constancy. *Image and Vision Computing*, 17(8):597–604, 1999.

[67] G.D. Finlayson and S. Hordley. Improving gamut mapping color constancy. *IEEE Transactions on Image Processing*, 9(10):1774–1783, 2000.

[68] G.D. Finlayson and S. Hordley. Colour constancy at a pixel. *The Journal of the Optical Society of America A*, 18(2):253–264, 2001.

[69] G.D. Finlayson, S. Hordley, and P. Hubel. Colour by correlation: A simple, unifying approach to colour constancy. In *Proceedings of the Int. Conf. on Computer Vision*, pages 835–842, 1999.

[70] G.D. Finlayson, S. Hordley, C. Lu, and M.S. Drew. On the removal of shadows from images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 28:59–68, 2006.

[71] G.D. Finlayson, S. Hordley, and R. Xu. Convex programming colour constancy with a diagonal-offset model. In *IEEE International Conference on Image Processing*, 2005.

[72] G.D. Finlayson, P. Hubel, and S. Hordley. Colour by correlation. In *Fifth Color Imaging Conf.*, pages 6–11, 1997.

[73] G.D. Finlayson and G. Schaefer. Convex and non-convex illumination constraints for dichromatic color constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 598–605, 2001.

[74] G.D. Finlayson and G. Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *Int. J. Comput. Vision*, 42(3):127–144, 2001.

[75] G.D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Twelfth Color Imaging Conference: Color, Science, Systems and Applications.*, pages 37–41, 2004.

[76] G.D. Finlayson and R. Xu. Convex programming color constancy. In *IEEE workshop on color and photometric methods in computer vision, Int. Conf. on Comp. Vision*, 2003.

[77] G. Fischer and M. Sajjaa. Whitebalpr a new method for automatic white balance. In *IS&Ts European Conference on Color in Graphics, Imaging and Vision*, pages 202–207, 2008.

[78] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, 2nd edition, 1990.

[79] D.A. Forsyth. A novel approach to color constancy. In *Proceedings of the Int. Conf. on Computer Vision*, pages 9–18, 1988.

[80] D. Foster, S. Nascimento, and K. Amano. Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience*, 21:331–336, 2004. `http://personalpages.manchester.ac.uk`.

[81] D.H. Foster, K. Amano, and S.M. Nascimento. Color constancy in natural scenes explained by global image statistics. *Visual Neuroscience*, 23:341–349, 2006.

[82] C. Fredembach and G.D. Finlayson. Bright chromagenic algorithm for illuminant estimation. *The Journal of Imaging Science and Technology*, 52(4):1–11, 2008.

[83] C. Fredembach and S. Susstrunk. Illuminant estimation and detection using near infrared. In *IS&T/SPIE Electronic Imaging, Digital Photography V*, 2009.

[84] B. Funt, V. Cardei, and K. Barnard. Learning color constancy. In *Proceedings of the IS&T/SID Color and Imaging Conference*, 1996.

[85] B. Funt and et al. Hdr dataset, 2010. http://www.cs.sfu.ca/ colour/data/funt_hdr.

[86] B. Funt and G.D. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:522–529, 1995.

[87] B. Funt and H. Jiang. Non-diagonal colour correction. In *IEEE International Conference on Image Processing*, 2003.

[88] B. Funt and B. Lewis. Diagonal versus affine transformations for color correction. *Journal of the Optical Society of America A*, 17(11):2108–2112, 2010.

[89] B. Funt and M. Mosny. Removing outliers in illumination estimation. In *Proceedings of the 20th IS&T/SID Color and Imaging Conference*, 2012.

[90] B. Funt and L. Shi. The effect of exposure on maxrgb color constancy. In *SPIE Volume 7527 Human Vision and Electronic Imaging XV*, 2010.

[91] B. Funt and L. Shi. The rehabilitation of maxrgb. In *18th Color Imaging Conference*, pages 256–259, 2010.

[92] B. Funt and Xiong W. Estimating illumination chromaticity via support vector regression. In *Proceedings of IS&T Color Imaging Conference*, pages 47–52, 2004.

[93] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *CVPR'08: Computer Vision and Pattern Recognition*, 2008.

[94] R. Gershon, A. Jepson, and J. Tsotsos. From [r, g, b] to surface reflectance: computing color constant descriptors in images. In *International Joint Conference on Artificial Intelligence*, pages 755–758, 1987.

[95] R. Gershon, A. Jepson, and J. K Tsotsos. The use of color in highlight identification. In *Proceedings of the 10th international joint conference on Artificial intelligence*, volume 2, pages 752–754, 1987.

[96] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and T. Gevers. Color constancy from physical principles. *Pattern Recognition Letters*, 24(11):1653–1662, 2003.

[97] T. Gevers and A.W.M. Smeulders. Color-based object recognition. *Pattern Recognition Letters*, 32:453–464, 1999.

[98] A. Gijsenij. Color constancy : Research website on illuminant estimation. `http://staff.science.uva.nl/~gijsenij/colorconstancy/index.html`.

[99] A. Gijsenij and T. Gevers. Color constancy using natural image statistics. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[100] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):687–698, 2011.

[101] A. Gijsenij, T. Gevers, and M. Lucassen. A perceptual analysis of distance measures for color constancy algorithms. *Journal of the Optical Society of America A*, 26(10):2243–2256, 2009.

[102] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, October 2008.

[103] A. Gijsenij, T. Gevers, and J. van de Weijer. Physics-based edge evaluation for improved color constancy. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009.

[104] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20:2475–2489, 2011.

[105] A. Gijsenij, R. Lu, and T. Gevers. Color constancy for multiple light sources. *IEEE Transactions on Image Processing*, 21(2):697–707, 2012.

[106] R. Gottumukkal and V. Asari. Skin color constancy for illumination invariant skin segmentation. In *Proc. of SPIE-IS&T Electronic Imaging*, volume 5685, pages 969–976, 2005.

[107] G. Healey. Using color for geometry-insensitive segmentation. *Journal of Optics Soceity of America A*, 6(6):920–937, 1989.

[108] G. Healey. Estimating spectral reflectance using highlights. *Journal of Image and Vision Computing*, 9(5):333–337, 1991.

[109] G. Healey and T.O. Binford. Local shape from specularity. *CVGIP*, 42(1):62–86, 1988.

[110] S. Hordley. Scene illuminant estimation: past, present, and future. *Color Research and Application*, 31(4):303–314, 2006.

[111] S. Hordley and G.D. Finlayson. Reevaluation of color constancy algorithm performance. *Journal of the Optical Society of America A*, 23(5):1008–1020, 2006.

[112] Y. Imai, Y. Kato, H. Kadoi, T. Horiuchi, and S. Tominaga. Estimation of multiple illuminants based on specular highlight detection. In *Computational Color Imaging*, volume 6626 of *Lecture Notes in Computer Science*, pages 85–98. Springer, 2011.

[113] G.J. Klinker, S.A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In *Proceedings of the Int. Conf. on Computer Vision*, volume 1, 1987.

[114] G.J. Klinker, S.A. Shafer, and T. Kanade. The measurement of highlights in color images. *International Journal of Computer Vision*, 2:7–32, 1988.

[115] G.J. Klinker, S.A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1):7–38, 1990.

[116] J. Kuang and W. Xiong. Color constancy for multi-illuminants high-dynamic-range scenes. In *IS&T/SID Color Imaging Conference*, 2008.

[117] E.H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, December 1977.

[118] E.H. Land and J.J. McCann. Lightness and retinex theory. *The Journal of the Optical Society of America*, 61:1–11, 1971.

[119] M. Lecca and S. Messelodi. Computing von kries illuminant changes by piecewise inversion of cumulative color histograms. *Electronic Letters on Computer Vision and Image Analysis*, 8(2):1–17, 2009.

[120] H.-C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *The Journal of the Optical Society of America A*, 3:1694–1699, 1986.

[121] H.-C. Lee, E. Breneman, and C. Schulte. Modeling light reflection for computer color vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(4):402–409, 1990.

[122] T.M. Lehmann and C. Palm. Color line search for illuminant estimation in real-world scenes. *The Journal of the Optical Society of America A*, 18(11):2679–2691, 2001.

[123] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[124] B. Li, D. Xu, and C. Lang. Colour constancy based on texture similarity for natural images. *Coloration Technology*, 125(6):328–333, 2009.

[125] B. Li, D. Xu, W. Xiong, and S. Feng. Color constancy using achromatic surface. *Color Research and Application*, 35(4):304–312, 2010.

[126] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing: Label transfer via dense scene alignment. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1972–1979, 2009.

[127] D.G. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 60, pages 91–110, 2004.

[128] C. Lu and M.S. Drew. Practical scene illuminant estimation via flash/No-flash pairs. In *Color Imaging Conference*, 2006.

[129] R. Malik and P. Bajcsy. Achieving color constancy across multiple cameras. In *in Proc. ACM International Conference on Multimedia*, pages 893–896, 2009.

[130] T. Malisiewicz, A. Gupta, and A.A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 89–96, 2011.

[131] S. Mallick, T. Zickler, P. Belhumeur, and D. Kriegman. Specularity removal in images and videos: A PDE approach. In *Computer Vision ECCV 2006*, pages 550–563, 2006.

[132] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *Computer Vision and Pattern Recognition*, pages 619– 626, 2005.

[133] B. Martinkauppi. *Face Colour Under Varying Illumination: Analysis and Applications*. PhD thesis, Oulu University, Finland, 2002.

[134] C.S. McCamy, H. Marcus, and J.G. Davidson. A color-rendition chart. *J. App. Photog. Eng.*, 2:95–99, 1976.

[135] N. Moroney, M. Fairchild, R. Hun, C. Li, M. Luo, and T. Newman. The ciecam02 color appearance model. In *Proceedings of the 10th IS&T/SID Color and Imaging Conference*, pages 23–27, 2002.

[136] M. Mosny and B. Funt. Reducing worst-case illumination estimates for better automatic white balance. In *Proceedings of the 20th IS&T/SID Color and Imaging Conference*, 2012.

[137] S. Nascimento, F. Ferreira, and D. Foster. Statistics of spatial coneexcitation ratios in natural scenes. *Journal of the Optical Society of America A*, 19(8):1484–1490, 2002. `http://personalpages.manchester.ac.uk/staff/david.foster`.

[138] S. Nayar, X.S. Fang, and T. Boult. Removal of specularities using color and polarization. In *Computer Vision and Pattern Recognition*, pages 583–590, 1993.

[139] S. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):611–634, 1991.

[140] J. Nieves, C. Plata, E. Valero, and J. Romero. Unsupervised illuminant estimation from natural scenes: an rgb digital camera suffices. *Applied Optics*, 47(20):3574–3584, 2008.

[141] M. Oren and S. Nayar. A theory of specular surface geometry. *International Journal of Computer Vision*, 24(2):105–124, 1997.

[142] M. Osadchy, D. Jacobs, and R. Ramamoorthi. Using specularities for recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1512–1519, 2003.

[143] J.B. Park. Efficient color representation for image segmentation under nonwhite illumination. *SPIE*, 5267:163–174, 2003.

[144] C.A. Parraga, G. Brelstaff, T. Troscianko, and I. Moorehead. Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, 15(3):563–569, 2002. `http://psy223.psy.bris.ac.uk/hyper`.

[145] C.A. Parraga, J. Vazquez-Corral, and M. Vanrell. A new cone activation-based natural images dataset. *Perception*, 36(Suppl):180, 2009. `www.cvc.uab.es/color_calibration`.

[146] E. Rahtu, J. Nikkanen, J. Kannala, L. Lepisto, and J. Heikkila. Applying visual object categorization and memory colors for automatic color constancy. In *Proceedings of the International Conference on Image Analysis and Processing*, pages 873–882, 2009.

[147] J. Renno, D. Makris, T. Ellis, and G. Jones. Application and evaluation of colour constancy in visual surveillance. In *Proceedings of the 14th IEEE International Conference on Computer Communications and Networks*, pages 301–308, 2005.

[148] C. Riess, E. Eibenberger, and E. Angelopoulou. Illuminant color estimation for mixed-illuminant real-world scenes. In *IEEE Color and Photometry in Computer Vision Workshop*, 2011.

[149] C. Rosenberg, M. Hebert, and S. Thrun. Color constancy using kl-divergence. In *Proc. IEEE International Conference on Computer Vision*, pages 239–246, 2001.

[150] C. Rosenberg, T. Minka, and A. Ladsariya. Bayesian color constancy with non-gaussian models. In *Neural Information Processing Systems*, 2003.

[151] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

[152] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[153] M. Sajjaa and G. Fisher. Automatic white balance: whitebalpr using the dichromatic reflection model. In *in Proc. SPIER, vol. 7250 Digital Photography V*, 2009.

[154] I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):290–300, 2003.

[155] Y. Sato. *Object Shape and Reflectance Modeling from Color Image Sequence*. PhD thesis, Carnegie Mellon University, 1997.

[156] G. Schaefer, S. Hordley, and G.D. Finlayson. A combined physical and statistical approach to colour constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 148–153, 2005.

[157] S.A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10:210–218, 1985.

[158] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter sensitive hashing. In *Proceedings of the Int. Conf. on Computer Vision*, pages 750–757, 2003.

[159] L. Shi and B. Funt. Dichromatic illumination estimation via hough transforms in 3d. In *in IS&Ts European Conference on Color in Graphics, Imaging and Vision*, 2008.

[160] L. Shi and B. Funt. Re-processed version of the gehler color constancy dataset of 568 images, 2010. http://www.cs.sfu.ca/ colour/data.

[161] R. Shrestha and J.Y. Hardeberg. Computational color constancy using a stereo camera. In *6th European conference on color in graphics,imaging and vision: final program and proceedings*, pages 69–74, 2012.

[162] M. Storring, H. Andersen, and E. Granum. Skin colour detection under changing lighting conditions. In *7th Symposium on Intelligent Robotics Systems*, 1999.

[163] P. Tan, S. Lin, L. Quan, and H.Y. Shum. Highlight removal by illumination-constrained inpainting. In *International Conference on Computer Vision*, pages 164–169, 2003.

[164] R.T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 178–193, 2005.

[165] R.T. Tan, K. Nishino, and K. Ikeuchi. Color constancy through inverse-intensity chromaticity space. *The Journal of the Optical Society of America A*, 21(3):321–334, 2004.

[166] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV'98: International Conference on Computer Vision*, 1998.

[167] S. Tominaga, S. Ebisui, and B.A. Wandell. Scene illuminant classification: brighter is better. *The Journal of the Optical Society of America A*, 18:55–64, 2001.

[168] S. Tominaga and B.A. Wandell. Standard surface-reflectance model and illuminant estimation. *The Journal of the Optical Society of America A*, 6(4):576–584, 1989.

[169] J. Toro. Dichromatic illumination estimation without pre-segmentation. *Pattern Recognition Letters*, 29(7):871–877, 2008.

[170] J. Toro and B. Fun. A multilinear constraint on dichromatic planes for illumination estimation. *IEEE Transactions on Image Processing*, 16(1):92–97, 2007.

[171] A. Torralba, R. Fergus, and W.T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1958–1970, 2008.

[172] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of Optics Soceity of America*, 57:1105–1114, 1966.

[173] M. Tsukada and Y. Ohta. An approach to color constancy using multiple images. In *IEEE International Conference on Computer Vision*, 1990.

[174] H.R. Vaezi Joze and M.S. Drew. Improved machine learning for image category recognition by local color constancy. In *Proceedings of IEEE Int. Conf. on Image Proc.* IEEE, 2010.

[175] H.R. Vaezi Joze and M.S. Drew. Exemplar-based colour constancy. In *Proceedings of the British Machine Vision Conference*, pages 26.1–26.12, 2012.

[176] H.R. Vaezi Joze and M.S. Drew. White patch gamut mapping colour constancy. In *Proceedings of IEEE Int. Conf. on Image Proc.* IEEE, 2012.

[177] H.R. Vaezi Joze and M.S. Drew. Exemplar-based colour constancy and multiple illuminant scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013. Submitted for publication.

[178] H.R. Vaezi Joze, M.S. Drew, G.D. Finlayson, and P.A. Troncoso Rey. The role of bright pixels in illumination estimation. In *Proceedings of the 20th IS&T/SID Color and Imaging Conference*, 2012.

[179] K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

[180] J. van de Weijer and T. Gevers. Color constancy based on the grey-edge hypothesis. In *Int. Conf. on Image Proc.*, pages II:722–725, 2005.

[181] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007.

[182] J. van de Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. In *IEEE Conference on Computer Vision (ICCV)*, 2007.

[183] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the 7th European Conference on Computer Vision-Part III*, pages 255–271, 2002.

[184] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1-2):61–81, April 2005.

[185] J. Vazquez-Corral, C.A. Parraga, M. Vanrell, and R. Baldrich. Color constancy algorithms: Psychophysical evaluation on a new dataset. *Journal of Imaging Science and Technology*, 53(3):563–569, 2009. `www.cvc.uab.es/color_calibration`.

[186] J. von Kries. Influence of adaptation on the effects produced by luminous stimuli. In D. MacAdam, editor, *Sources of Color Vision*, pages 109–119. MIT Press, 1970.

[187] M. Vorobyev, J. Marshall, D. Osorio, N.H. de Ibarra, and R. Menzel. Colourful objects through animal eyes. *Color Research and Application*, 26:S214–S217, 2001.

[188] N. Wang, D. Xu, and B. Li. Edge-based color constancy via support vector regression. *IEICE Transactions on Information and Systems*, E92-D(11):2279–2282, 2009.

[189] G. West and M. Brill. Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy. *Journal of Mathematical Biology*, 15(2):249–258, 1982.

[190] L.B. Wolff. Using polarization to separate reflection components. In *Computer Vision and Pattern Recognition*, pages 363–369, 1989.

[191] L.B. Wolff and T. Boult. Constraining object features using a polarization reflectance model. *Trans. on Pattern Analysis and Machine Intelligence*, 13(7):611–634, 1991.

[192] J. Worthey and M. Brill. Heuristic analysis of von kries color constancy. *Journal of the Optical Society of America A*, 3(10):1708–1712, 1986.

[193] M. Wu, J. Sun, J. Zhou, and G. Xue. Color constancy based on texture pyramid matching and regularized local regression. *The Journal of the Optical Society of America A*, 27(10):2097–2105, 2010.

[194] W. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulas*. Wiley, New York, 2nd edition, 1982.

[195] J. Xiao, J. Hays, K. A Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pages 3485–3492, 2010.

[196] W. Xiong and B. Funt. Estimating illumination chromaticity via support vector regression. *Journal of Imaging Science and Technology*, 50(4):341–348, 2006.

[197] W. Xiong and B. Funt. Stereo retinex. *Image and Vision Computing*, 27(1-2):178–188, 2009.

[198] W. Xiong, B. Funt, L. Shi, S. Kim, B. Kang, S. Lee, and C. Kim. Automatic white balancing via gray surface identification. In *Proceedings of the 15th IS&T/SID Color and Imaging Conference*, 2007.

[199] J. Yang, S. Hanazawa, M.K. Yamaguchi, and I. Kuriki. Color constancy in 4- to 5- month old infants. *Journal of Vision*, 9(8), 2009.

[200] Q. Yang, S. Wang, and N. Ahuja. Real-time specular highlight removal using bilateral filtering. In *Proceedings of the European Conference on Computer Vision*, pages 87–100, 2010.

[201] F. Zaraga and G. Langfelder. White balance by tunable spectral responsivities. *Journal of the Optical Society of America A*, 27(1):31–39, 2010.

[202] H. Zhang, H. Liu, and S. Quan. Multi-illuminant color constancy for HDR images through exposure segmentation. In *SPIE*, pages 829204–829204, 2012.

[203] T. Zickler, S. P Mallick, D. J Kriegman, and P. N Belhumeur. Color subspaces as photometric invariants. *International Journal of Computer Vision*, 79(1):13–30, October 2007.

Index