Understanding Colour Image: Colour Constancy

Bozhi Liu

School of Computer Science

The University of Nottingham

This dissertation is submitted for the degree of

Doctor of Philosophy

March 2018

# Acknowledgements

It is really not easy right now for me to review my PhD study, there are so many people I would like to thank during the past 7 years.

First and foremost, I would especially like to thank my PhD supervisor, Professor Guoping Qiu, who gave me the chance to learn how to do academic research, and lead me into this interesting research area since my master study. Qiu is always being optimistic, always encourage me to do the things I like, to keep the passion for research also the life. Qiu never gave me up during the whole study even the worst time, he tried his best to help me back on track, and finally I accomplished my PhD study. Thank you very much.

I would also like to thank my viva examiners, Prof Tony Pridmore and Prof Lik-kwan Shark, for giving me insightful comments, feedbacks and advices, to improve the quality of my thesis, and finally qualified to submit it as PhD level.

I would like to thank Hao Fu, as my best model researcher, who also gave me a lot of good ideas in research. I also thank Jie Shu for being my great labmate and gym partner. And many thanks to Yujie Mei, Qian Zhang, Wenwen Bao, Min Zhang, Orod Razeghi, Salvador Garcia Bernal, and Mercedes Torres Torres, who are all my best labmates worked together in our VIP Lab.

I would like to extend thanks to my friends who spent amazing time with me at Nottingham for years, thanks to Wenchao Jiang, Jia Fu, Chaoyu Ye, Yike Fang, Jianhua Shao, Huijin Yu, Yixing Xu, Yin He, Qiong Yin and Wenchong Du. The Life is short, but the memory will last long.

I would like to thank Xi Gao, Minghua Liu, Yi Nie, and Penghe Zhang, for sharing the pleasant time with me during my visiting year in Ningbo Campus, and also, thanks to Xia Li, Xiaosong Zhou, Pengfei Cao, Xiang Luo, Luming Chen, Chao Zhang, Zhan Xu, Zhuo Lei, Qile He, who were around during my daily life in Ningbo and enriched the experience more graceful.

I am thankful for all the administrative support from my school and university officers, including international office and student service, specially thanks to Christine Fletcher, Nadine Holmes and Hong Zhang.

I would like to thank Bolei Xu, Jingxin Liu, Ke Sun, Xianxu Hou, Wei Pan, Yuexiang Li, Qing Li and Rui Cao for thoughtful discussions and ideas sharing, also continuing working with me in Shenzhen University right now in our new Vista Lab. I would also like to thank Bozhi Liu for his help during my visiting at Shenzhen University.

I would like to give a big thank you to my parents, Dongsheng Liu and Yuehua Cao, for supporting me as a self-funding student to pursue my dreams. I am deeply indebted to the love from them. This thesis is dedicated to them with my sincere gratitude.

Last but not least, thanks to the colourful world, for giving me passion and interest to aspire after truth.

Abstract

Human visual system has a mechanism which ensures that the perceived colour of an object remains almost constant under varying illumination conditions, and this mechanism is called color constancy. Electronic imaging systems such as digital cameras do not naturally have this ability. The color appearance of images of an object under different lighting conditions changes with the colour of the light sources and this can cause problems in many computer vision applications such as object recognition. To deal with this problem, many algorithms have been developed to estimate the input image's illuminant, and then recover the intrinsic colour of the scene correctly. In this thesis, we focus on this topic, try to produce new colour constancy algorithms in both images and videos, to improve the performance of the state of the art.

This thesis makes four technical contributions. First, we have developed a new image representation scheme suitable for developing learning based colour constancy algorithms; second, we introduce a new method that formulates the colour constancy problem as one that infers the illuminant class of the input image; third, we introduce a novel clustering classification colour constancy framework (the 4C method); and finally, we extend our method from still image into video processing, create a new framework to deal with the colour constancy problem in videos.

As in many computer vision problems, one of the crucial issues is how to effectively represent the input events. Colour constancy is no exception and we need to first represent the input image. As we are only interested in the colours of the image, colour histogram is a natural choice. However, traditional colour histogram is content dependent. As our task is estimating the colours of the illuminant rather than the colours of the image, we need a representation that is relatively independent of the image content. Based on this reasoning, we introduce the novel concept of a binary colour histogram where it records if a colour has appeared in the image or not and disregards the frequency of the colours appear in the image. We will present experimental results to demonstrate that our new binary histogram representation is particularly suitable for learning based colour constancy and that it provides better performances than other traditional representation schemes.

The colour of a digital image is directly affected by the colour of the illuminant. We reason if we can recognize or classify the illuminant source of the image, we can then correct the colour of the image. Based on this rationale, we formulate the colour constancy problem as an illuminant classification problem. We assume that each image has an associated class of illuminant and the task of colour constancy is that of recognizing the illuminant class of the image. To accomplish this, we make use of our newly introduced binary colour histogram representation scheme and employ a powerful machine learning method called the Random Forest to construct the illuminant recognition system. We will present experimental results to show the effectiveness of our new method.

Encouraged by the success of our illuminant recognition framework, we have developed a novel clustering classification colour constancy (the 4*C*) framework. We reason that similar illuminants will result in similar white point colours in an image. Based on this assumption, we first use a clustering algorithm to group similar white point colours of the training samples into the same cluster. We then treat the images in the same cluster as belonging to the same illumination source and each cluster as one class of illuminants. The colour constancy problem, i.e., that of estimating the unknown illuminant of an image, becomes that of identifying which illuminant class (cluster) the image's illuminant falling into. We again make use of our novel binary colour histogram representation and our random forest based illuminant classification methods to implement our new 4C colour constancy framework. We present experimental results on publicly available testing datasets and show that our new method is competitive to state of the art.

As a practical application, we have successfully extended our novel colour constancy methods from still image into video processing. The video tonal stabilization problem is still an unsolved problem, and current algorithms are only focusing on keeping the tonal stable during video playing, not really trying to recover the incorrect illuminant. We tackle these two problems together by keeping the tonal stable and recovering the frame colour to a canonical illuminant. Our approach first divides video frames into shots containing similar illuminant characteristics. We then correct the frames in the same scene by using the Random Forest illuminant estimation framework. A smooth function is applied to prevent flick and flash from occurring at the boundary of the neighboring scenes. Experimental results show that our new methods can improve video quality effectively.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Projection Background

Alongside other features such as shape, texture and size, colour helps humans to make sense of the world around them. The perception of object colour by the human visual system in varying conditions of illumination demonstrates a certain level of consistency. For instance, a white piece of paper seen in an outdoor environment when it is sunny will still be perceived as white when seen at different times of the day with daylight of varying wavelength composition. From a formal perspective, the colour of the light is ignored by the human visual system via certain algorithmic mechanisms, meaning that colour descriptors are visualised without dependence on the illuminant. Colour constancy refers to this capability of relying on an object's surface characteristics to discern its colour [L$^+$77]. The conditions under which visualisation occurs determine to some extent the level of this colour constancy. Despite this, the colour constancy that the human visual system is capable of is usually acceptable.

Colour constancy could have numerous valuable applications in the context of computer vision, provided that the colour descriptor is addressed (i.e. surface properties not dependent on the colour of the light). Among the potential applications that could benefit from this are digital photography, object recognition, aerial image analysis, image retrieval, and automatic quality inspection systems [FBL96]. However, people's perception of colour, which depends on how an object is reflected, and image development physics (i.e. image capture by the camera) are incongruous because the colour of the dominant illuminant confuses camera images. This is illustrated in Figure (1.1), where an identical scene is imaged under various illuminants. Alterations in illumination give a different appearance to the scene in the three images, even though it is actually the same scene.

Fig. 1.1 The same scene was photographed using the same camera but under three distinct illuminants, namely, daylight, white fluorescent light, and tungsten light. It is obvious that the illuminants determine the colour of the objects.

Colour constancy in computer vision has not be addressed as an issue in itself, but as a pre-processing step to enable the use of colour in activities like object recognition. Automatic location and identification of objects in digital images is the purpose of object recognition methods. To this end, a series of features are used to characterise the target objects and thus set them apart from additional image elements. These features ought to be independent of effects like translation, rotation, scale, and, most importantly for the present study, illumination alteration [HS94, JZQ11]. Colour constancy in computer vision can also benefit digital photography, particularly in procedures like colour balancing, which involves "correction" of the image colours based on the illumination type using estimations of ambient illumination from the digital camera data. The need to address colour constancy in computer vision led to the development of several algorithms, which has been frequently done in conjunction with human visual processing. Illuminant colour estimation or inference from an image is the basic mechanism underpinning all algorithms because this enables elimination of illumination-related colour bias [HHF^{+}97].

Many studies have sought to solve the issue of computational colour constancy, but so far, no computationally affordable and high-performance approach has been developed. The premises underpinning the majority of uncomplicated colour constancy algorithms may be false in some cases, making the algorithms unsuccessful [BCF02, BMCF02]. Another disadvantage of this category of algorithms is that their performance is not as good as that of the more advanced algorithms of greater complexity [GLG12, JD14, OK16]. The latter typically rely on training data-sets for model learning, but they are not viable to be applied in real-time in devices like digital cameras, video recorders and robots, because they are not computationally cheap. Therefore, the purpose of the current study is to develop straightforward and rapid algorithms by addressing the issue of illumination inference based on the Random Forest method as the primary data source.

1.2 Outline of the Thesis

The outline of this thesis is provided below.

In Chapter 2, we discuss background material and previous works related to the topic of this thesis. We start by learning the cause of colour, find out what is colour constancy, describe the computational colour constancy problem, then review algorithms and categorize illumination estimation methods into four general groups: statistical, gamut based, learning-based and physics-based algorithms methods, then review several colour constancy methods in each of these categories.

In Chapter 3, under the assumption that images taken under the same illuminant should have similar colour appearance, we present a novel learning-based technique for colour constancy. Based on a powerful machine learning model, Random Forest, we treat colour constancy as a problem of inferring the illuminant class of the input image. Random Forest is a tree-based data structure; our technique is therefore intrinsically fast and efficient. This work makes two main contributions we made in this work: Firstly we have developed a new image representation feature suitable for developing classification or learning based colour constancy, and secondly we have developed a novel colour constancy framework that classifies or recognises illuminants based on Random Forest. We will present experimental results on a number of publicly available testing databases and show that our new technique outperforms state of the art techniques.

However, there are obviously disadvantages to the classification framework for illuminant estimation: if the data were only roughly classified or did no have enough classes provided, the potential number of the estimation results would be based on the number of illuminant classes provided by the data sets. In a ideal condition, the natural illuminance distribution should be a continuous function. To deal with this problem, There are two potential ways to improve the algorithm: the first way is to use more classified data set for training, and the other is to classify the data more specifically by using the clustering method.

In Chapter 4, in trying to adopt a more classified data set with synthetic images, we describe our approach to the implementation and compare it with a number of leading algorithms, and reporting the results using synthesised data. Experiments using synthesised data are important because we can generate an arbitrarily large number of images, such that we obtain stable results when testing the algorithms, and because the ground truth is known, which means that possible confounds due to camera characterisation and pre-processing are absent, and various factors affecting colour constancy can be efficiently investigated because they can be manipulated individually and precisely. Furthermore, the data could be used to generate more illuminant spectral information, thus potentially providing more classes than the previous data for our classification framework training. We have evaluated our algorithm performance with the synthetic data and have also conducted another test with real-world image data-sets to determine whether the synthetic data are useful and practical.

In Chapter 5, to classify the data more specifically, we introduce a novel clustering classification colour constancy framework (referred to as the 4*C* method). In the new 4*C* method, we assume that the training data contain the *RGB* values of the white point of the scene, referred to as $R_wG_wB_w$, which is obviously the colour of the illuminant under the imaging system. It is always possible to obtain these measurements by placing a neutral white patch in the scene when collecting training images, and some of the data-sets we use in our work [CF03, GRB$^+$08] employ this approach to obtain the $R_wG_wB_w$ values of the white point (patch) of the scene. Based on the assumption that similar illuminants will have similar illumination colours (colours of the white patch of the scenes), we first use a clustering algorithm to group similar $R_wG_wB_w$'s of the training samples into the same cluster. We then treat the $R_wG_wB_w$'s in the same cluster as belonging to the same illumination source and each cluster as one class of illuminant. The colour constancy problem, i.e., that of estimating the unknown illuminant of an image, becomes that of identifying which illuminant class (cluster) the image's illuminant belongs to. To achieve this, we use a classification algorithm to classify the image into the illuminant class (cluster). We present experimental results on publicly available testing datasets and show that our new method is competitive to state-of-the-art methods.

In Chapter 6, we address the colour constancy problem from still image to video. The video tonal stabilisation problem is still an unsolved problem, and current algorithms are only focusing on keeping the tonal stable during the video playing, not really trying to recover the incorrect illuminant. From this point, we are trying to combine these two problems together, as firstly keeping the tonal stable and secondly recovering the frame colour to canonical illuminant, and propose an algorithm for improving the performance of the video. We firstly treat the problem as a video illuminant recognition problem, use scene detection to cut the video into different shots based on illuminants, and for all the frames in each illuminant, we adopt our still image colour constancy algorithm to recover these frames with one single correction parameter. Subsequently, to avoid the tonal gap between different neighbouring scenes, we generate a simple smoothing method. We test our video colour correction framework on a published data set, to present some experiments to show the effect.

Finally, in Chapter 7, we present our general conclusions and future work.

Chapter 2

Literature Review

This chapter is intended as an introduction to research on computational techniques for colour constancy. Several aspects are addressed, including the determinant of colour, the colour reflection model framework, as well as the existing methods that have been proposed to deal with the issue of computational colour constancy.

2.1 Colour Determinant

Colour plays a key role in the processing of an image, facilitating object recognition and characterisation [Bra04]. For example, individuals can find their way quicker if the colour of a landmark is included in the directions they receive, or the colour of a fruit helps determine how ripe it is. Attribution of a specific colour to an object is an exceptional capability. To shed light on this capability, the representation of information regarding the spectral features of an object in the retinal image must be addressed.

On what physical conditions is production of varying colour stimuli dependent? Why are the colours of different objects distinct? A close inspection of the colours in the everyday environment reveals that the majority of objects with colours that stand out obtained their colour through either tinting or dying, they do not have any colour of their own because they do not possess the pigments that produce colour [BW86]. For instance, textiles are naturally pale or white in colour and they are coloured through treatment of their fibres with a substance or dye. Just like snow or clouds, textiles that are white owe their colour to a certain transparent substance without colour that is present in the single fibres. The white colour is the result of the reflection and refraction of light that cause the homogeneous dispersal of most of the light in every direction [Fin00].

Things stand differently if pigments are present in the fibre tissue; then, a portion of the light is absorbed in certain wavelength areas by every fibre serving as a coloured filter

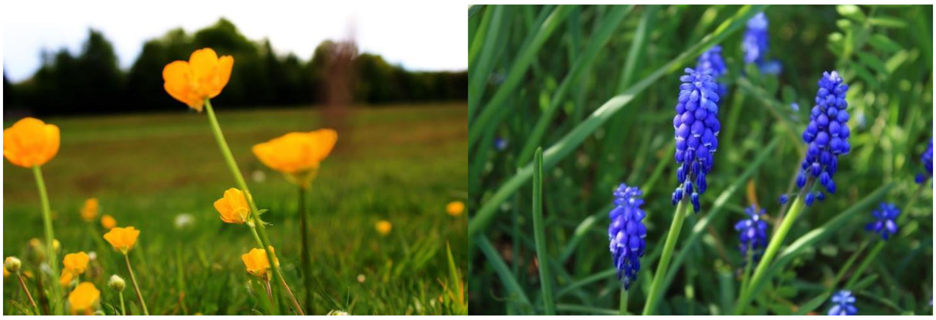


Fig. 2.1 Flowers appear differently coloured due to the various pigments they contain.

[Bra04]. The illuminate light emitted by a source enters the human eye after going through one or multiple fibres and it conveys the colour of the fibre(s) permissible by that portion of the light wavelength range through which light filtering occurred.

The same principle applies to flowers: flowers that are "snow-white" lack pigments and their colour is given by the numerous tissue gaps that are filled with air and determine light to be reflected and refracted in different ways. On the other hand, pigments are present in coloured flowers and the specific colour that is perceived by the human eye is the outcome of special colour filters. The two effects responsible for the colouring of various items are selective absorption of light and light diffraction/reflection by the existing dye molecules in different wavelength areas, the item colour being given by the reflected and diffracted light that enters the human eye. An example is shown in Figure 2.1.

The colour reflection framework integrates illuminants, objects and observers, which is shown in Figure 2.2 [L+77]. A series of illuminated objects make up a scene. Since the manner in which the illumination is spatially distributed is usually complex, differences occur between the illuminants falling on various objects. If it is assumed that the illumination demonstrates homogeneity throughout a scene, then its spectral power distribution $E(\lambda)$ (λ = light source wavelength) can be used for its characterisation [FCB96], indicating the amount of power possessed by the illuminant at every wavelength. The human eye perceives the illuminant as reflection off objects and concentrates it to create the retinal image. This image, in turn, helps to make sense of how the scene is constituted.

Light is not absorbed and reflected in the same way by the surfaces of objects. The wavelength, incident light angle in relation to the surface normal, and the reflected light angle are the factors that usually determine reflection [For90]. To enable characterisation of every object surface based on a spectral reflectance function $S(\lambda)$, geometric aspects should be simplified or overlooked [FCB96]. The portion of incident illumination that an object reflects at different wavelengths is indicated by this function.

Human eyes and digital cameras can both be classified as observers. The amount of light energy assimilated at different wavelengths, which is known as the system sensor sensitivity, is associated with the observers as well. In the colour reflection framework, this is denoted by the $C(\lambda)$ function [L^{+}77]. The colour signal represents the light that every visible scene location reflects to the eye. Since it has already been assumed that the illuminant is homogeneous in a scene, the three outlined functions of $E(\lambda)$, $S(\lambda)$, and $C(\lambda)$ can be used to determine the spectral power distribution of the colour signal:

$$\rho_k = \int E(\lambda)S(\lambda)C_k(\lambda)d\lambda \tag{2.1}$$

In general, three different sensor types (k = 3) are incorporated in the majority of imaging devices. Therefore, at a certain pixel, the light response is given by a triplet of responses $\rho = (\rho_1, \rho_2, \rho_3)$. Figure 2.2 outlines the framework.

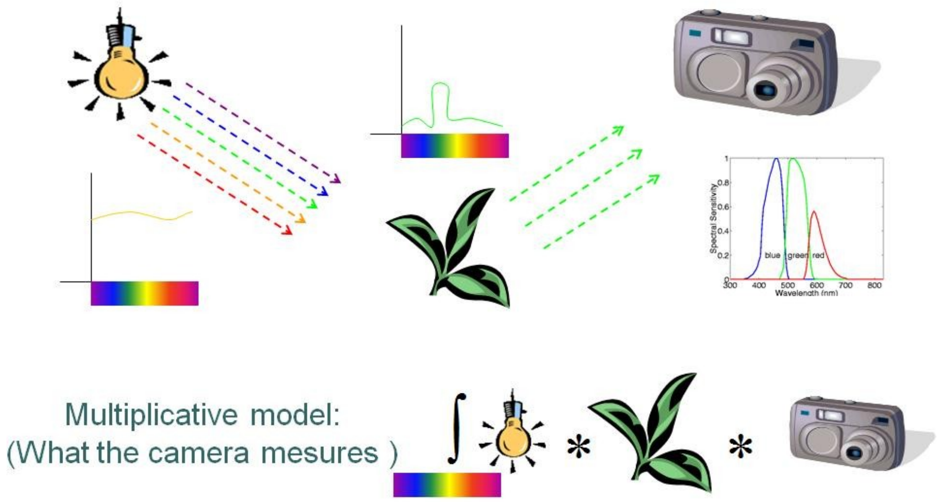


Fig. 2.2 The colour reflection framework: According to the attributes of the pigments, a portion of the illuminant light wavelength is reflected by an object and the signal is converted by the camera into RGB values subsequently captured in the image. The camera functioning mechanism is closely replicated by the human eye.[L^{+}77]

Equation (2.1) defines the imaging model known as the **Mondrian World** [L^{+}77], which is based on the premise of spatial homogeneity of the light source, object flatness and coplanarity, and Lambertian surface reflectance [Bra04, For90, FCB96]. Furthermore, to minimise the complexity of the colour reflection framework, the characterisation of how the illuminant is spatially distributed, how the scene is structured, and how the direction of the incident and reflected light affects the spectral reflectance of every object are not considered by the **Mondrian World** [BMFC02]. However, for real scenes, the Mondrian

World premises do not stand. For reasons of practicality and realism, the spatial distribution of the illuminant, the scene structure, and the influence of the direction of incident and reflected light on the spectral reflectance of every object need to be characterised [Fol90]. Nevertheless, for preliminary analysis the **Mondrian World** is a sufficiently suitable framework [Bra04, For90, FCB96].

2.2 Definition and Features of Colour Constancy

The simulation technique is used to determine the function $C(\lambda)$, which is constant for a specific observer [Bra04]. The involvement of the illuminant and surface reflectance in the production of the colour signal is indicated by Equation (2.1). These two physical factors are respectively inherent and external to the object, with surface reflectance being informative about the identity and attributes of the object, while the illuminant does not supply any information regarding the object.

What enables attribution of a clear colour to an object if the illuminant and surface properties are a source of confusion for the colour signal at an image location? As indicated by the expression of Equation (2.1), modifications in the surface reflectance of an object can be closely replicated by modifications in the illuminant; therefore, it can be anticipated that a stationary object will undergo significant alterations in appearance when the illuminant fluctuates. Figure 2.3 shows this physical process through the example of an identical flower presented under four distinct sources of light. It is apparent that the colour of the light source greatly influences the colour of the flower [Fos03].

The differences in colour appearance are not so pronounced when the four patches are visualised within the setting of the images they belong to [Bra04]. In Figure 2.3, such appearance stabilisation is only partial, as the viewer sees only printed images of small size derived from an illuminated medium of larger proportions. The difference in the perceived colour would usually be negligible if the viewer stood in front the depicted flower, meaning that co-analysis of multiple image areas is undertaken by the visual system to address the ambiguous nature of the colour signal. Thus, the object surface colour is perceptually represented in a stable manner based on the complete scene setting. This is what colour constancy is.

Colour is an outcome of cognitive and retinal processing, rather than a property that can be assigned to objects in the surrounding environment. The illuminant has no connection to the ability of the human visual system to discern object colours, which is known as colour constancy [Zek93]. This ability is not exclusive to humans, but occurs in other species as well, including goldfish and honeybees [Tov96], highlighting the importance of colour as a

Fig. 2.3 As Equation (2.1) shown, changes in illuminants can make large changes in the appearance of a fixed object. In this example, the same flower is depicted four times, each rendered under a different light source. As can be seen, the colour of the flower is strongly dependent on the colour of the light source.[GRB$^+$08]

biological signalling mechanism. Indeed, reliable recognition of objects according to their colour cannot be achieved without colour constancy. Remarkably, fluctuations in illuminant do not interfere with the development of fixed descriptors by the visual system.

2.3 Computational Colour Constancy in Digital Image Processing

As defined above, perception of object colour without dependence on illuminant colour is known as the ability of colour constancy, which is displayed by the human visual system. However, there are several question marks about this ability and its underlying mechanism, especially why it is demonstrated by the human visual system, while the camera lacks it. Computational colour constancy is geared towards addressing the issue of modifications in the colour of the same object as a result of modifications in the light source, as indicated by Equation (2.1).

The mechanism can be better grasped through the examination of the propagation of a ray of light. After it is emitted by a source of light, the light ray reaches an object at a given moment, when the portion of reflected incident light is dictated by the reflectance according to the physical features of the object. Subsequently, the reflected light penetrates the eye and is subjected to measurement by three different kinds of receptors present within the retina, which are capable of absorption of light with long, middle, and short wavelengths, respectively. The reflectance of the whole visible spectrum needs to be measured to precisely calculate the amount of light that the object reflects. This measurement can be undertaken by using white light or another light source of known power distribution to illuminate the object. However, reflectance measurement cannot be undertaken by a device such as a camera because it lacks knowledge of the power distribution of the light source.

What the camera does measure is the outcome of object reflectance and how much light falls upon the object. However, there are two aspects that are not known in this regard, namely, the reflectance and the type of light source. The illuminant will cause variation in the output if the measuring device is used without any extra processing. This effect ought to be familiar to amateur photographers. The kind of illuminant employed determines the appearance of an image. For example, the colours generated by sunlight or candle light are warmer than those generated by flash, irrespective of the type of camera used (i.e. analogue or digital). In the case of the objects of a scene illuminated by a yellowish source of light, a white surface will be coloured yellow because it will reflect the incident light the same for every wavelength. This is not a problem if the image comes in digital form because it can be edited to improve the colour reproduction. Image colour correction can be achieved with a range of well-known programs, including Adobe Photoshop, Google's Picasa, and GIMP.

The procedure is referred to as automatic white balance in photography. There are a number of alternatives for white balance management in the case of digital cameras. The white balance can be established to the suitable setting if knowledge exists regarding the kind of light source illuminating a scene, which means that the light source is a light bulb, neon light, sunlight or cloudy sky. An image of the same scene as under a white illuminant can be produced by the camera due to having knowledge of the colour of those light sources. Certain cameras can capture an image of a white patch as well, subsequently using the image contents to determine the illuminant colour. Another technique employed by some cameras to identify the illuminant colour is ambient light measurement based on the sensors they are equipped with. To ensure that the appearance of an image is the same as under a white light source, the automatic white balance option is available for automatic selection of the optimal alternative. Naturally, production of a representation of a scene that is identical to the one observed by the photographer based on the image data is the overall aim.

Object reflectance is also an important aspect from the perspective of machine vision. If it is known, the reflectance can help to recognise an object on the basis of its colour, as well as to divide scenes into sections or for colour-based optical flow determination. Every facet of colour-based computer vision depends on reflectance information. For example, the functioning of an autonomous service robot should be independent of whether the light source is artificial light or sunlight. Since the light source has no influence on reflectance, the type of light source will be irrelevant if reflectance is employed by the computer vision algorithm. However, by the same token, illuminant fluctuation can cause problems if colour is employed as a cue to address basic vision activities (e.g. scene segmentation and object recognition and tracking). Therefore, extensive attention is being paid in the field of computer vision research to address the problematic issue of colour constancy.

Various computation strategies have been developed to determine the true colour of object surfaces and solve the colour constancy problem. In this process, the light source colour must be eliminated, as it contributes to the colour of the object surface alongside the real surface colour. Elimination of the light source colour is simplified if the illumination is identified, and hence, this is typically the focus of colour constancy. As previously mentioned, the colour vision system of both humans and animals depends greatly on colour constancy, which facilitates object recognition, regardless of discrepancies in illumination. It is thus not far-fetched to assume that any smart visual system would hinge on colour constancy to a significant degree. Numerous applications in computer vision rely on colour constancy, including image recovery, colour reproduction, and object detection [JD10]. This prompted a proliferation of different colour constancy algorithms [Hor06, GGVDW11].

One study exemplified the colour constancy application based on the use of straightforward colour constancy techniques in a conventional issue of object categorisation [JD10]. Thus, a number of 20 objects (e.g. bird, bottle, car, dining table, motorbike, people) were categorised through application of colour constancy algorithms as pre-processing of the bag-of-words learning techniques followed by performance comparison between these techniques and constant descriptors like C-SIFT [VDSGS10] and Opponent-SIFT [BG09]. Although the performance of colour constancy as pre-processing did not exceed that of colour invariant descriptors, a notable enhancement of over 10% in the performance of object categorisation was achieved by integrating local colour constancy techniques and colour invariant descriptors [VDSGS10]. Such findings suggest that computer vision studies stand to benefit considerably from good colour constancy.

2.4 Colour Space with Computational Colour Constancy

Colour has different representation ways in different colour spaces, defined by different use [Poy97]. For example, standard *RGB* colour space is the most widely used colour space for LCD display, and *CMYK* is another colour space used to define colours of printer. There are a large number of other different colour spaces as well, and most of the colour spaces could transfer to another colour space by a linear transformation. In this section, we will discuss some generally used colour spaces in colour constancy area.

2.4.1 RGB Colour Space

RGB colour space is an additive colour space model, with means any colour in this space could be mixed by using Green, Red and Blue three primary colours with a certain percentage(degree in 3D space) [SG31]. In the *RGB* colour space, the colour can be presented in a unit cube. The cube have three axes, which is red, green and blue colours in each direction. Any point in this cube is a unique colour in *RGB* colour space. Generally, as our image format are 8 bits, the range of R, G, and B in digital images is from 0 to 255, when all the three channels are 0, the colour is defined as black, and in converse, in the opposite point of the cube, white colour is the maximum value with (255,255,255). When $R = G = B$, the colour is greyscale, and with the value increasing, the intensity of the illuminant is enhancing.

Since the *RGB* colour space is the most widely used standard to capture or generate pictures, a lot of colour constancy algorithms just work on this. For example, the greyworld algorithm [Buc80], the white patch algorithm [FCB96], and the first version of gamut mapping[For90], the details of these algorithms will be discussed in the later section.

2.4.2 sRGB Colour Space

The *sRGB* is developed from *RGB* colour space, to solve the problem which is the same *RGB* colour will appear differently on LCD displayers, TVs, or digital cameras. With the development of internet, to handle the device independent colour representation problem, Hewlett-Packard and Microsoft proposed this standard (Since at that time, the two companies nearly monopoly the market of personal computers.) Currently, as the default standard, the images transfering in the internet are attached with a profile, which include the parameter setting to make sure the image display on different device with the same colour measured by human eyes. The approach to achieve this is by applying a gamma correction on the images with various factors, and the pixel value could be transferred to linear *RGB* colour space. Why we need to pay attention on images under *sRGB* space is for some image processing or computer vision problem, the image need to be preprocessed to the linear space, also required when obtaining the ground truth of the illuminant in colour constancy datasets.

2.4.3 rg Chromaticity Space

rg Chromaticity space is an illuminant intensity independent colur space, the two colour signals r and g are normalized from R,G, and B signals by the following Equation:

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B}, r+g+b = 1 \tag{2.2}$$

An R,G, and B combination is defined a absolute, unique colour, but in the normalized rg space, a color is represented by the proportion of red, green, and blue in the colour, instead of the intensity of each. In this colour space, the colour features could be only involved just red and green, since sum up all the signals must always a total of 1.

The inadequate of intensity does not mean the rg chromaticity is less useful than RGB space, since it is not relative with the intensity of the illuminant, it could avoid the spotlight effect when extracting the colour feature. And furthermore, in computer vision area, not all the algorithms require the colour as the essential parameter, and actually most of them are trying to keep away from an illuminant changing environment (That is why we need to tackle with the colour constancy problem.) In another way, some of the colour constancy works [Fin95, FH00] assume only the chromaticity is the useful information and based on this assumption, to develop algorithms with good performance. Also it reduces the dimension of the feature, to simplify the problem in computation cost.

There are several works under other colour spaces(Lab, HSV)[HF14, Fai13], since these colour spaces are not mainly applied in colour constancy topic, the algorithms introduced lately will only focus on the colour spaces we described above.

2.5 Framework for Computational Colour Constancy Algorithm

Computational colour constancy is concerned with formulating a relevant characterisation of a scene from an image captured under unknown conditions of lighting without dependence on the illuminant [BCF01, BCF02]. This usually involves a two-stage process whereby the illuminant parameters are determined (see the following section for more details) and afterwards illumination is determined with the obtained parameters without dependence on surface descriptors [BVdW11, BW86, GRB$^+$08]. To deal with the considerable level of abstraction exhibited by the surface descriptors, the working assumption in this study is that the characterisation of an image from a scene independent of illumination captured under known, conventional lighting conditions [BF97]. The standard illuminant is selected somewhat at random. In this case, an illuminant for which the camera is balanced is employed because it is most appropriate for tasks of image reproduction [CF03, BMFC02]. Considering the camera colour as denoting function, a framework is created on the premise of a diagonal model of illumination modification with autonomous scaling of different channels to map the image captured under a light source to the image captured under a different light source (e.g. the standard illuminant) [BW86, CF03, FH00].

According to Equation (2.1), after the accurate determination of $E(\lambda)$, it can be substituted with a standard canonical illuminant to accomplish colour constancy [For90]. Hence, finding out the image illuminant is the main obstacle to colour constancy [For90, BF97, FH00, WGG07, GRB$^+$08]. It is possible to express the issue in terms of retrieval of the scene illuminant *SPD* $E(\lambda)$ through inversion of Equation (2.1), provided that the issue of colour constancy is specified as the issue of the determination of the illuminant in a particular scene. Just three measurements (i.e. λ components) are available for one surface and neither $E(\lambda)$ nor $S(\lambda)$ are known. Hence, the issue is under-constrained. Extra constraints on the illuminant can be achieved through the addition of more surfaces, based on the assumption that $E(\lambda)$ does not change in an imaged scene. On the other hand, the added surfaces give rise to additional unknowns because their reflectance functions are unspecified. This unsuccessful approach has prompted the development of numerous other colour constancy techniques in the literature that involved the introduction of extra assumptions regarding the world to deal with the issue of insufficient constraints [For90, BF97, BMF00, JD14].

The present thesis is concerned with techniques that rely on just one image taken with a standard camera to determine the colour of the light source. Subsequently, the techniques can be used not only with a random image taken already or with Internet images, but also as incorporated white balancing methods within the camera. Certain techniques for determining the colour of the light source employ extra data from special hardware systems, such as WhitebalPR [FS08, SF09], a novel series of camera sensors [NPVR08, ZL10], near-infrared information [FS09], stereo images [XF09], extra images [LM09, SH12], or image sequence information (e.g. video) [MB09, RMEJ05]. However, these techniques cannot be applied to images that have already been taken.

Illumination estimation can be understood as determination of the geometry or direction of light [SSI03, BJK07], of the complete spectral power distribution of light [DXW01], or of the colour of light. In the present case, it is understood to refer strictly to determination of the colour of light, since this is the primary focus of the study.

Although additional techniques related to colour constancy exist, they are excluded from the review of colour constancy methods as they lack the conventional colour constancy stages of illumination estimation and colour correction, because they do not retrieve the true colour of object surfaces. The greyscale [FH01] or multi-channel images [FF95, GvdBSG03] produced by these techniques are independent of light source [FH01], specularity [FD01] or shadowing [FDL04, DJ09]. Since they are unaffected by lighting conditions, their outputs may not always be indicative of the colour or intensity of the initial image, which is why their main application is as input to advanced processes of computer vision (e.g. recognition) [FF95, GS99].

2.5.1 Image Colour Correction

As previously indicated, once the colour of the light source for an input image is determined, its effect must be eliminated. This can be achieved through a conversion that produces an output image seemingly captured under a reference illuminant (i.e. canonical light). In other words, every colour of the input image has to be converted into a new colour as appearing under the canonical light.

The Lambertian model that disregards specular reflection underpins the majority of colour constancy techniques, apart from physics-based techniques. This produces similarities between the image formation model and the dichromatic model with no specular reflection [Tom96]. This conversion can be construed as an instance of chromatic adaptation, provided that the colour of the light source is homogeneous across the whole image [Fai13]. The modelling of the conversion is typically based on a 3×3 linear transformation, which involves multiplying the 3×3 transformation matrix $D_{3\times3}$ to convert the colour of each pixel determined under the canonical light $(R_c, G_c, B_c)^T$ in the input image $(R_i, G_i, B_i)^T$:

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = D_{3\times3} \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \tag{2.3}$$

Simplification of the linear transformation to a diagonal transformation called the von Kries Model can be achieved based on the premise that the camera sensors or light represent delta functions (narrowband) [BW92]. In reality, actual camera sensor curves are not narrowband, as can be observed in Fig. (2.4(a)), which illustrates the sensor sensitivity curves determined for a Sony DXC-930 3-CCD camera. Several studies have addressed the conditions allowing for the substitution of the linear transformation with diagonal transformation [WB82, FDF94, FL00].

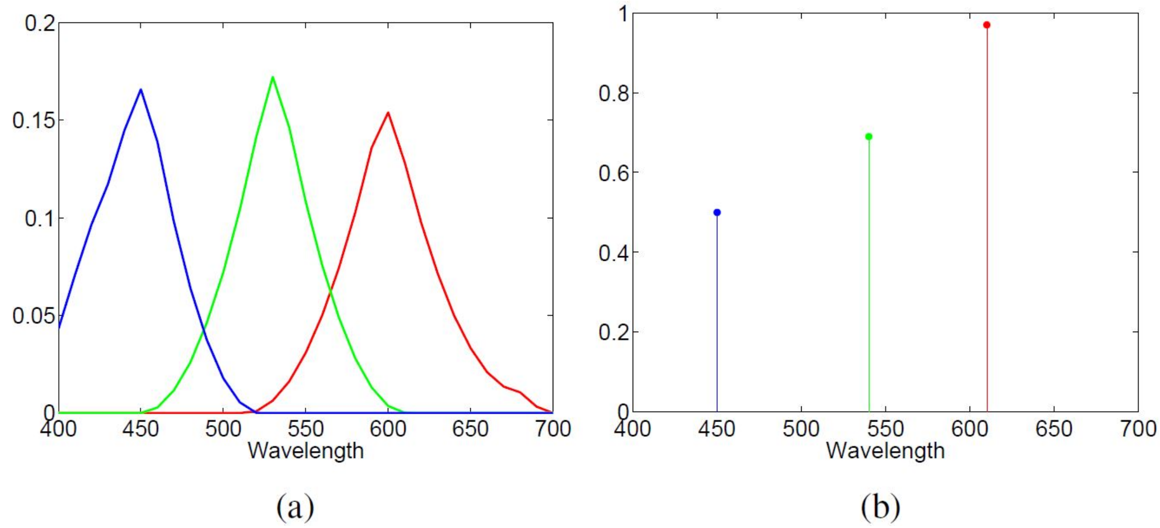


Fig. 2.4 (a): RGB sensors determined for a Sony DXC-930 camera; (b) theoretical narrow-band RGB camera sensors.[Ebn07]

To elucidate the fundamental mechanism underpinning the ability of human colour constancy, an uncomplicated chromatic adaptation model was developed by von Kries in 1902. The hypothesis formulated for this specified that there was no connection between the separate elements included in the organ of vision and only their particular function determined their adaptation or fatigue [Fai13]. Equation (2.3) illustrates this hypothesis, which has come to be known as the von Kries model.

In the case of a scene with a white patch, where (R_e,G_e,B_e) and (R_c,G_c,B_c) respectively denote the camera response to the white patch under the unknown light source and under the known canonical illuminant, the mapping of the response to the white patch from the unknown to the canonical illuminant can be achieved based on scaling the three channels:

$$\begin{bmatrix} R_O \\ G_O \\ B_O \end{bmatrix} = \begin{bmatrix} \frac{R_c}{R_e} & 0 & 0 \\ 0 & \frac{G_c}{G_e} & 0 \\ 0 & 0 & \frac{B_c}{B_e} \end{bmatrix} \begin{bmatrix} R_I \\ G_I \\ B_I \end{bmatrix} \tag{2.4}$$

Colour correction through a diagonal transform is extensively employed in the domain of colour constancy because it is uncomplicated and provides an acceptable degree of estimation, despite being solely an approximate and depending on the premise of the homogeneity of the colour of light [WB86]. The majority of the examined colour constancy techniques are based on the diagonal transformation colour correction model, typically without the use of spectral sharpening. Creation of the transfer matrix is possible insofar as the same frame is compatible with non-white patches, provided that the same pixel colour values in the same position in the scene but under different light sources can be determined. The main application of the diagonal matrix model is in retrieval of the input image colour under the canonical illuminant, following determination of the illuminant with the image. The next parts will focus on the definition of the transfer parameters and different algorithms based on different premises and techniques for an image with neither light source nor reference canonical light source being known.

2.6 Colour Constancy Algorithms

Estimation of the unknown light source of the input image followed by the transfer of the colour value under the canonical light source is the main difficulty of colour constancy. Three major classes of techniques for overcoming this difficulty have been discerned in the literature and several key techniques in each class are presented in the next parts.

Techniques for determination of the illuminant colour can be categorised according to a range of classificatory schemes, including supervised and unsupervised techniques, static and learning-based techniques, and physics- and non-physics-based techniques. In the present thesis, four general classes of techniques for illumination estimation are distinguished, namely, statistical techniques, which use the statistical features of an image to determine its illuminant, gamut-based techniques, which achieve illumination estimation through comparison of canonical gamut with image gamut, learning-based techniques, which use training images for model learning to determine the illuminant and Physics-based algorithms. These classes are not wholly distinct but may merge into each other; for instance, statistical and physics-based techniques may be employed as static techniques, while gamut-based and learning-based techniques may be used as supervised techniques.

2.6.1 Statistical Methods

Low-level statistics or the physics-based dichromatic reflection model underpins the static techniques used on images with a constant parameter setting, such as Grey-World [Buc80], White-Patch [L$^+$77], Shades of Grey [FGJ98, FT04], and Grey-Edge [VDWG05, VdWSV07]. The premise on which the Grey-World algorithm is based is that the average colour in a scene is achromatic under a white illuminant [Buc80], while the assumption of the White-Patch algorithm is that perfect reflectance is the determinant of the maximum response in an image [L$^+$77]. The colour constancy algorithm produced by the Shades of Grey technique displays greater generality and is rooted in the Minkowski norm [FGJ98, FT04]. Last but not least, the Grey-Edge algorithm assumes that the average colour of the margins of objects in a scene is achromatic under a white illuminant [VDWG05, VdWSV07].

Grey World Methods

Put forth by [Buc80], the Grey World assumption relies on the average pixel colour for illuminant estimation; in other words, computation of an average of the light reaching the observer enables illuminant estimation. This assumption is a highly popular colour constancy algorithm and is at the basis of numerous other algorithms [Ebn04, Ebn03, FGJ98], all of which presume that the world is grey, on average. The following equation expresses how an illuminant can be estimated with the Grey World algorithm:

$$e = [mean(R) \quad mean(G) \quad mean(B)]^t \tag{2.5}$$

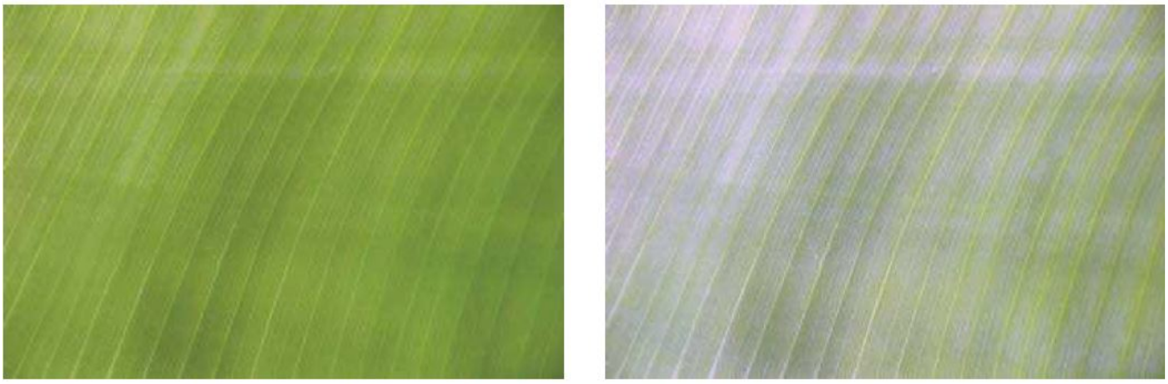
In the above, the 3D vector for determining the light source is denoted by e, while the three colour channels in the image are denoted by R, G and B.

The premise of the Grey World technique is that a pre-established value, called grey, represents the average of the surface reflectance of a common scene [Buc80]. It is necessary to clarify what is meant by grey. It may mean true grey, and particularly 50% homogeneous reflectance, or it may mean the average of the reflectance database. This is anticipated to exceed the performance of Grey World with synthesised data and the use of many surfaces can ensure an optimal outcome because a greater average achromatic value can be obtained the more colour is used [BCF02]. By contrast, the performance of Grey World with image data is anticipated to be lower because the actual average surface reflectance is unknown. Obviously, errors are more likely to occur when there is an insufficient diversity of colours in images. For instance, since the average value employed for estimation of reference white is incorrect, an image with numerous green leaves will be corrected as grey leaves.

Despite this limitation, the Grey World algorithm remains the colour constancy algorithm with the highest popularity, due to its facile application and rapidity. This technique is adopted by the white balance function incorporated in numerous commercially available camera models for correction of camera sensor responses.

Use of Channel Maximum to Estimate the Illuminant

A less complex version of the retinex algorithm, the Max-RGB or White-Patch algorithm relies on the maximum response in every channel to determine the light source [FCB96, FBL96, CF99]. The retinex algorithm depends on the existence of a bright patch in an image



(a) (b)

Fig. 2.5 If the number of colours in the scene is not high enough, accurate colours will not be yielded by the Grey World assumption; (a) banana plant leaf; (b) output image.[Ebn07]

location, which will reflect the greatest amount of light possible for every band, thus giving the illuminant colour. The following equation expresses how the illuminant is estimated with the White-Patch algorithm:

$$e = [max(R) \quad max(G) \quad max(B)]^t \tag{2.6}$$

In the above, the 3D vector for light source estimation is denoted by e, while the three colour channels in the image are R, G and B.

It is obvious that this technique displays sensitivity to the dynamic range associated with the visual system. Provided that a white surface exists in the image, Max-RGB can produce satisfactory outcomes. However, in the context of an environment of matte reflectance, the maximum reflectance in the scene will never exceed that of a pure white due to the bias inherent in the estimate of the illuminant magnitude. On the other hand, the maximum reflectance can exceed that of pure white if specularities occur. Noise sensitivity is an additional limitation of Max-RGB [MKNM95], as is the fact that it computes the illuminant estimate based on just a single value from each channel. On a more positive note, one study demonstrated that Max-RGB yields an effective estimate of illuminant chromaticity if considerable specularities exist and if the dynamic range of the visual system is good enough to avert the clipping of those specularities [RGM02]. In this way, the algorithm can exceed the performance of algorithms relying just on chromaticity input because it benefits from data regarding pixel brightness.

However, Max-RGB is limited by the fact that an erroneous estimate of the light source can be generated due to just one bright pixel. The estimate will be incompatible with the true colour of the illuminant if the image contains a highlight due to inhomogeneous reflectance of the illuminant colour by an object. Another issue is the occurrence of noise in the image. Furthermore, high sensitivity to clipped pixels is exhibited by the White-Patch retinex algorithm as well [FBL96]. It is not possible to produce a dependable estimate of the illuminant colour from the pixel of greatest brightness if one or multiple colour channels are clipped.

Shades of Grey

Finlayson and Trezzi's [FGJ98] demonstration that only slight changes were needed to generate results comparable to the results of complex colour constancy algorithms prompted renewed attention to be paid to low-level techniques. The two authors argued that the Grey World and Max-RGB constituted distinct versions of a more general, Minkowski norm-based colour constancy algorithm, namely, the Shades of Grey algorithm. A compromise between Grey World and Max-RGB has to be reached, given that they can produce good outcomes if the average colour in the scene is approximated to grey or the maximum is white. Finlayson and Trezzi applied the premise that the scene average was a shade of grey and used the following function to mathematically compute the technique:

The application scope of the algorithm is somewhat enlarged by the Minkowski-norm, thus allowing integration of the Grey World and White-Patch algorithms. Representing the normalised outcome forming the approximated illumination vector, the Minkowski-norm determines a weighted average of the pixel values and allocated higher weights to the pixels that have higher intensity. The Minkowski-norm p underpins the suggested weight function:

$$\left(\frac{\int (f(X))^p dX}{\int dX}\right)^{1/p} = Ke \tag{2.7}$$

For p values of 1 and ∞, the above equation respectively corresponds to the premise of Grey World and colour constancy by Max-RGB coupled with computation of the maximum value max f(x). Recurring experiments showed that the Shades of Grey algorithm achieved the highest estimation performance at a p value of 6.

Grey Edge Hypothesis

The use of just one homogeneous illuminant and illumination estimation based on pixel values are the defining features of the majority of techniques proposed by earlier studies. However, these features do not always hold true. For instance, more than one illuminant can affect an image. Thus, to achieve colour constancy under several different illuminants, the Grey Edge technique was developed [VDWG05, WGG07]. As the image margins display the largest number of details, this technique is geared towards enabling edge-based colour constancy. It assumes that the average edge discrepancy in a scene is achromatic, in keeping with the premise of the Grey World algorithm that the average surface reflectance is achromatic. The premise of the Grey Edge technique allows computation of the illuminant colour from the average colour derivative in the image. The technique produces results similar to those of more advanced algorithms, but it is more computationally affordable, easier to apply, and more time-effective in terms of computation performance. In addition to playing a key role in image analysis, edge detection constitutes an essential pre-processing step in the context of image processing domain as well. Feature extraction and texture profile analysis also benefit from it. However, efforts are still ongoing to develop an effective technique of edge detection that has minimal sensitivity to noise and can locate edges accurately. There are two types of edges derived from a 2D image of a 3D scene, namely, edges that depend on viewpoint, which may undergo alterations in keeping with viewpoint modifications and are generally indicative of scene geometry (e.g. mutual occlusion of objects) and edges that do not depend on viewpoint, which are usually indicative of 3D object intrinsic attributes (e.g. surface markings and surface shape).

The premise underpinning the use of low-level image features in the Grey Edge hypothesis [GS99] is that the average discrepancy of edges in the scene is achromatic. The principles of the Shades of Grey algorithm are adopted by the Grey Edge algorithm, with the exception that image derivatives instead of pixels are used. The basis of the Grey Edge algorithm is that the distribution of colour derivatives helps to obtain the maximal fluctuations in the direction of the light source. The direction of the light source is estimated with the help of the Minkowski norm of the derivatives. The technique is expanded through the introduction of derivatives of higher order or various types of derivation (e.g. convolution with Gaussian filters). The colour distribution presumed to occur in the image facilitates extraction of gradient information from the image.

However, if a gradient operator is unable to discern the gradient information, then the Grey Edge hypothesis does not ensure colour constancy. One study applied this hypothesis to a wavelet domain, with multi-resolution analysis being conducted to achieve colour constancy based solely on details (or edges) [WGG07]. However, no automated technique for determining the Minkowski norm estimate was put forth, although an adaptive technique was suggested for selection of the number of decomposition levels employed in discrete wavelet transform.

2.6.2 Gamut Mapping

Gamut mapping represents the second class of colour constancy algorithms [For90, FH00, GGVDW10]. It assumes that only a few colours can be visualised in real-world images with a certain light source [For90], implying that discrepancy in the light source colour is the reason for any unforeseen fluctuation in the colours of an image.

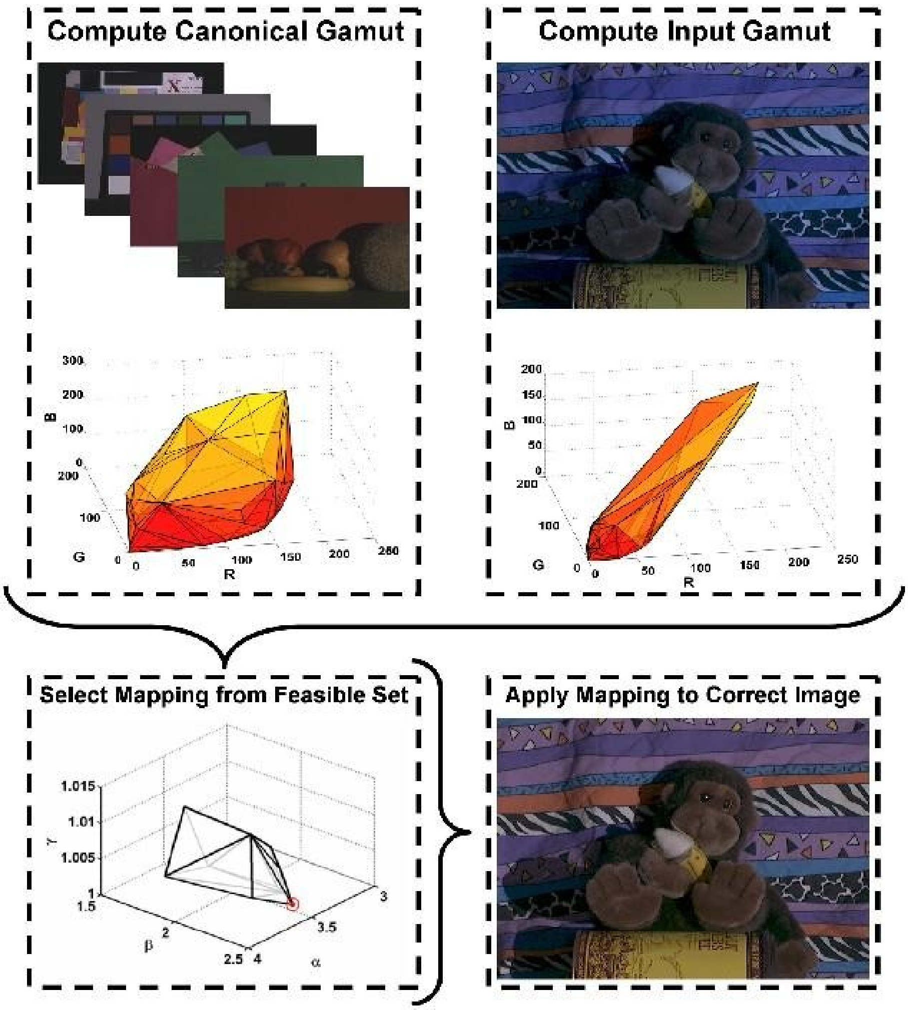


Fig. 2.6 The framework of the colour constancy algorithm of Gamut mapping.[For90]

Figure (2.6) presents the Gamut mapping framework. The technique starts with the formation of a set of the entirety of potential pixel (R,G,B) values related to real-world surfaces under a known, canonical source of light. Known as canonical gamut [For90] and taking the form of a convex hull, the finite colour set under canonical light source is learned off-line from a training set consisting of a series of actual images with ground truth illuminant or a series of hyperspectral surfaces (ground truth indicates that precisely represented colours under its own light sources are documented in the test environment, such as different colours of the same object under more than one light source). The experiment section will address several training and testing datasets.

The sets of every potential (R,G,B) under the unknown light source by its unknown convex hull are outlined. Based on the diagonal premise of illumination modification, a transfer can be created between the two hulls with a singular diagonal mapping. The calculation of the diagonal mapping parameters is particularly difficult from a technical perspective, but once they are obtained, the singular diagonal mapping between the canonical and the unknown light can be established. The estimation of mapping is schematically illustrated in the above figure with just one pixel value with ground truth; in reality, a greater amount of data with ground truth are needed to precisely estimate the mapping. A range of techniques for mapping estimation have been proposed [FH00].

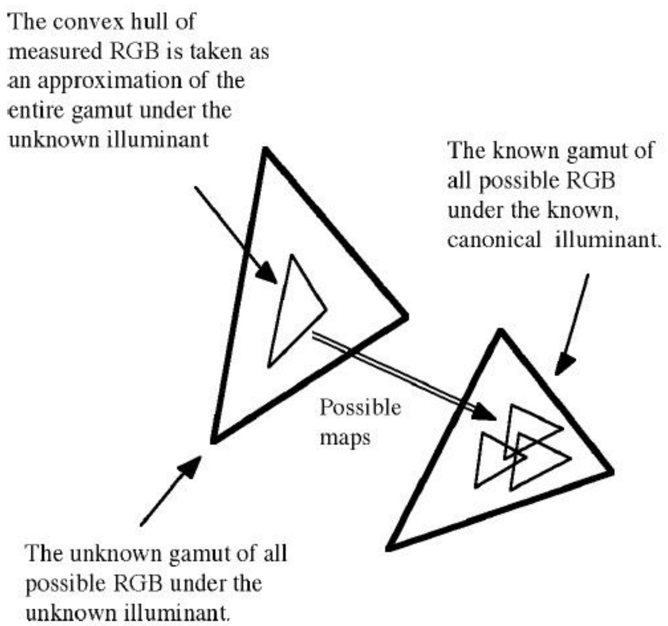


Fig. 2.7 Representation with two-dimensional triangles as to the operating principle of Gamut mapping colour constancy.[For90]

Gamut mapping is exemplified in Figure (2.7) with triangles denoting the gamuts. These represent three-dimensional polytopes in the complete algorithm version [For90]. The upper and lower thicker gamuts respectively denote the unknown gamut of the potential sensor responses under the unknown light source and the known gamut of sensor responses under the canonical light source. Although mapping between the sets is the overall aim, the first step is to determine the unknown set based on the observed sensor responses. The subset created by these responses has a convex hull denoted by the thinner triangle. The mapping to the canonical is not singular due to the fact that the observed set is usually an actual subset. To improve the efficiency of computation of a set of potential diagonal maps representing a convex set in the space of mapping coefficients, a particular technique was proposed in one study [For90].

The condition $(R,G,B) \in C$ (C = canonical gamut) must be met by a potential diagonal map (R,G,B), as it is a requirement for every observed (R,G,B) to be mapped into the canonical gamut. Therefore, $(R,G,B) \in C$, with C being the outcome of every (hull) point divided by (R,G,B) element-wise. The potential diagonal maps are subjected to a (convex) constraint by every one of these sets, and the intersection of the collection of convex sets derived from consideration of each actual observation gives the final solution set (it is enough to intersect just the sets equivalent to the vertex points of the convex hull of the observed RGB). The intersection of convex sets can be achieved by dividing up space into cubes and extracting the one present in every hull or by taking the convex hull of the suitable amounts in dual space [AGA07].

Incorrect premises or different errors (e.g. noise) may result in the lack of any shared points between the sets intended for intersection. In one study, the observed data were enhanced with the corner points of error boxes surrounding the data and the input of every corner point to $C_{R,G,B}$ was determined [FH00]. Again representing the convex hull of the contributing maps, the altered $C_{R,G,B}$ was subjected to expansion due to error modelling. This was different from mapping an expanded observed hull, since determination of $C_{R,G,B}$ required inversion of the observed points. Additionally, to prevent issues occurring during inversion, it was ensured in [FH00] that all R, G, or B used exceeded a small positive value based on threshold of the error box points. The increase in the level of error was done incrementally until the intersection was non-empty.

The Gamut mapping technique was enriched with two concepts by the Finlayson Colour in Perspective algorithm [Fin96], allowing compatibility of the technique with the chromaticity space $(R/B, G/B)$ and additional, non-convex constraint of the diagonal maps through their restriction to maps equivalent to the conjectured light sources. One study deemed that the combined solution set was the point where the convex constraint set associated with the initial surface constraints and the non-convex illuminant constraint set intersected [Fin96]. In a different study, the convex hull of the illuminant constraint set was used to estimate it, after which it was applied in the full case [GLG12].

The second key step of the algorithm after computation of the set of potential maps is selection of a solution from that set. This can be achieved in a number of different ways

[Fin00, FH00, For90]. The solution selected by the initial method represented the diagonal transform with maximal determinant, which enhanced the volume of the mapped set as much as possible [BF97]. An identical heuristic in chromaticity space is employed by the Colour in Perspective algorithm. Nevertheless, there is an inherent bias in this solution method and one study used both the chromaticity based algorithm and the (R,G,B) algorithm to determine the average of the constraint set [GLG12]. A different study chose to investigate the average three-dimensionally, due to the continued bias of this solution selection method in the chromaticity case [Fin00]. There is equivalence between the constraints on the mappings in perspective space and the cones in the space of mapping between the (R,G,B) gamuts. Monte Carlo integration is commonly applied for averaging across the non-convex illumination constraint. This average was estimated in [For90] based on a more straightforward version of numerical integration, not only in the chromaticity case, but also in the (R,G,B) case. Subsequently, the space was separated into cubes (or squares) and the cubes with the implied light source included in the cone of potential light sources were averaged. To determine the implied light source, the canonical illuminant $[(R,G,B)$ or $(R/B,G/B)]$ was divided by the mapping equivalent to the selected cube (or square) element-wise.

2.6.3 Learning-based Methods

Illumination estimation is achieved in learning-based techniques through the use of training data to learn a model. To this end, some algorithms involve application of learning or regression methods to a certain feature vector derived from the input image [XSF$^+$07, AGA07] or the use of low-level features rooted in the Bayesian formulation [BF97, RLM03, FHH99, GRB$^+$08].

Neural Network

The Neural Network solution [CFB02] involves training a network on an extensive set of images with known scene light source to determine how the observed image data and the scene light source are correlated. It is possible to construe the trained network as reflecting information regarding how images are statistically structured. In one study, network training was achieved not with image *RGB*s but with chromaticity information [CFB02]. Representing an image *RGB* without dependence on intensity, chromaticity coordinates can be determined in the following way:

$$\underline{c} = [c_1, c_2] = [\frac{\rho_1}{\rho_1+\rho_2+\rho_3}, \frac{\rho_2}{\rho_1+\rho_2+\rho_3}] \tag{2.8}$$

For a trained network, the calculation of an approximate of scene illuminant chromaticity is performed as follows:

$$\underline{\hat{c}}_w = network(\{\underline{c}_i \mid i = 1, \cdots m\}) \tag{2.9}$$

In the above, the trained network with m image chromaticities as input is denoted by network(.). The problem can be effectively solved with a neural network, if the statistical structure of the images occurring in practice is represented correctly by the network-encoded information. Similar approaches employ support vector regression [FX04, XF06, NDB09], linear regression methods [AGKA06, AGA07, AGK$^+$09], and thin-plate spline interpolation [SXF11]. Nevertheless, the efficiency of neural networks in estimating the light source is limited because training of a neural network to achieve generalisation (i.e. ability to correctly anticipate the scene light source in images not visualised before) is challenging.

Colour by Correlation

The Colour in Perspective technique [Fin96] was refined by Finlayson and colleagues, who proposed the new approach of Colour by Correlation [FHH97, FHH99, HF00]. This approach involves pre-computation of a correlation matrix indicating the level of congruence between the suggested illuminants and the occurrence of image chromaticities. Every matrix row is associated with a distinct training illuminant, while the columns of the matrix are equivalent to conveniently arranged potential chromaticity ranges produced by (r, g) space discretisation. One study proposed two versions of this approach [FHH97]. In one version, the computation of the correlation matrix components associated with a certain light source is undertaken in the following way: the camera sensors are used to determine the (r, g) chromaticities of the reflectances in the training set under the light source in question; the next step involves determining the convex hull of these chromaticities and the compatibility of all chromaticity bins found within the hull with the light source in question is ascertained; a value of one is accorded to every entry in the row for the light source in question equivalent to compatible chromaticities, while a value of zero is allocated to the rest of the elements.

Multiplication of the correlation matrix by a vector with elements equivalent to the same (r, g) bins employed in the correlation matrix enables approximation of the illuminant chromaticity. A value of one or zero is respectively accorded to the vector elements, depending on the occurrence or lack of occurrence of the equivalent chromaticity in the image. The number of chromaticities compatible with the light source is given by the ith element of the generated vector. Ideally, there will be compatibility between all image chromaticities and the real light source, giving the latter maximal correlation. Similar to Gamut mapping techniques, multiple potential light sources can be had. In the present case, the average of the totality of candidates near the maximum is used.

The other version of Colour by Correlation involves determination of the likelihood that the training illuminants are responsible for the observed chromaticities based on the correlation matrix. Subsequently, a maximum probability estimate or another estimate can be used to select the optimal illuminant. The surface reflectance database is consulted again to obtain the set of (r,g) for every illuminant to enable computation of the correlation matrix. A record is afterwards made regarding how frequently every discrete (r,g) occurs. Weighting of the counts is done depending on the availability of extra previous information regarding the occurrence likelihood of these reflectances. Support for illuminant priors is provided as well. Uniform statistics are employed in the current implementation. According to the particular illuminant, there is proportionality between the assembled counts and the likelihood of occurrence of a certain (r,g). The logarithms of these likelihoods for a particular illuminant are retained in an equivalent row of the correlation matrix.

The posterior distribution can be computed through a straightforward implementation of Bayes' rule. In the case of a series of occurring chromaticities C, the likelihood of scene illuminant can be calculated as follows:

$$P(I \mid C) = \frac{P(C \mid I)P(I)}{P(C)} \tag{2.10}$$

The above expression can be redefined based on the premise of homogeneous priors for I and $P(C)$ normalisation:

$$P(I \mid C) \propto P(C \mid I) \tag{2.11}$$

If the occurring chromaticities are assumed to be autonomous, the likelihood of observation of individual chromaticities c gives rise to $P(C \mid I)$, with the illuminant I:

$$P(C \mid I) = \prod_{c \in C} P(c \mid I) \tag{2.12}$$

With logarithms, the expression takes the form:

$$\log(P(C \mid I)) = \sum_{c \in C} \log(P(c \mid I)) \tag{2.13}$$

Selection of a response according to the posterior probability distribution is the last step. The computation of this final amount is achieved by applying the correlation matrix to the vector of chromaticity occurrences. The logarithm of the posterior probability associated with the ith illuminant is represented by the ith element of the generated vector. The three options proposed by the original study [FHH97]are maximum probability, mean probability, and local area mean. A comprehensive discussion of these options has been extended in [**?**] in relation to a relevant Bayesian approach to colour constancy.

The technique presented above may have a number of possible limitations. One such limitation is that, even if the training set includes the actual or a similar illuminant, zero likelihood can be generated by an observed set of chromaticities for all illuminants, as a result of noise and other factors causing incompatibility between the model and the real world. Another limitation is that the illumination could be the fusion of two sources of light (e.g. random combination of direct sunlight and blue sky), and under optimal circumstances, the technique should provide an intermediate response. The issue can be addressed in several ways. One solution to ensure that a potential illuminant is close to the real illuminant is for the algorithm to make sure that the (r, g) space is covered by the set of illuminants. Furthermore, to prevent any holes occurring in the distribution and make up for the noise, a Gaussian filter $(\rho = 10)$ could be used to smooth the frequency distribution of observed (r, g) during the assembly of the correlation matrices.

Bayesian Methods

Bayesian or probabilistic algorithms are statistical algorithms [BF94, Sap98, FHH99, BMF00, TEW01, RLM03] that are concerned with obtaining data regarding the probability of observation of a particular *RGB* response under potential scene illuminants as statistical priors. The probability that every potential scene illuminant is the actual illuminant in a certain scene is computed for a specific *RGB* image with a light source intended for classification. The scene light source approximate is usually considered to be the illuminant demonstrating maximum probability.

The *Color By Correlation* technique is the most straightforward application of the Bayesian approach to colour constancy [FHH99]. A general overview of the Bayesian principle model is provided in the following part. Restricting scene illuminants in this algorithm encode prior data regarding likely scene lights into a discrete limited set of potential lights τ_c. Despite the fact that, practically, the same likelihood is allocated to all lights, a random likelihood of observation can be allocated to every candidate scene illuminant. To be able to encode prior data regarding surfaces, a distribution of image chromaticities is defined for every potential scene illuminant. There are two possible actions that can be undertaken by the chromaticity distribution, namely, acquisition of data regarding the gamut of potential surfaces (by stating the observation or lack of observation of a specific chromaticity under a particular illuminant) or encoding of a measure indicating the likelihood of occurrence of a certain chromaticity under every illuminant. In the first instance, apart from the introduction of a restraint on the set of potential scene illuminants, the algorithm is the same as the Gamut mapping method, while in the second instance, a measure indicating how likely it is for each potential illuminant $\underline{\rho}_w$ to be the scene light source is determined:

$$L(\underline{\rho}_w) \propto \sum_{i=1}^{n} Pr(\underline{\rho}_i \mid \underline{\rho}_w) \tag{2.14}$$

The light with maximum probability is the approximated scene illuminant:

$$\hat{\underline{\rho}}_w = \arg max_{\underline{\rho}_x \in \tau_c}[L(\underline{\rho}_x)] \tag{2.15}$$

Although the same principle is at the basis of all Bayesian methods, there are differences in their approach to encoding of prior information regarding lights and surfaces as well as in the particulars of their application and how complex it is. These methods are advantageous primarily because they integrate the maximum possible amount of information in the problem definition and, from a statistical perspective, this information is used effectively.

The assumption that the likelihood of occurrence of a particular *RGB* in an image is not dependent on additional observed *RGB*s constitutes a possible shortcoming of the Colour by Correlation algorithm. Taking into account any dependency in practice is one way of dealing with this issue. The algorithm is also limited by the fact that, similar to the previously addressed neural networks, its efficiency depends on the level of compliance of real-image data with the algorithm-encoded statistical priors.

Combining Algorithms

Existing evidence suggests that satisfaction of algorithm premises by the image content ensures that illumination can be acceptably accomplished by algorithms. Nevertheless, an algorithm for illumination estimation that can account for all image data is yet to be developed. This prompted the development of a different series of learning algorithms integrating various approaches to attempt to enhance the efficiency of estimation. Several studies suggested the integration of the outputs of different algorithms [CF99, SHF05, BGS08]. The Grey World, Max RGB and a neural network-based approach were the algorithms for illuminant estimation considered in [CF99]. It was demonstrated by, compared to the individual performance of each algorithm, estimation could be more precisely made through an optimised (from the perspective of least-squares) weighted average of the estimations generated by all three algorithms. Meanwhile, in [BGS08], the outputs of five illumination estimation algorithms (i.e. Grey World, Max RGB, Grey-Edge, Colour by Correlation [FHH99], and Colour in Perspective [Fin96]) were integrated. Estimation was indeed enhanced by the integration of the different algorithms, but not significantly. Similar to the general Minkowski norm formalism, the estimation improvement may be explained in terms of the fact that the algorithm was refined by the inclusion of a greater number of parameters.

In other studies, selection of illumination estimation algorithm for a particular scene content, based on pre-established image classes, was done with the help of high-order statistics of the spatial distribution of image edges [GGVdW07, GGVDW11]. Two numbers regulating the form of a Weibull distribution were used to parameterise the distribution of edges. Subsequently, a specific constancy algorithm was indexed by the Weibull parameters, similar to the Grey-Edge method.

Using Semantic Information

The choice of optimal colour constancy algorithms in keeping with image classes was based on certain semantic information in other approaches. For instance, decision forests were used in [BCCS10] to formulate an indoor-outdoor classifier for image content, while in other studies, a stage classifier was employed to differentiate medium-level semantic categories [LGG+09b, LGG+09a, NSRG10]. Furthermore, a different study demonstrated the utility of semantic content probability [VdWSV07]. Other approaches sought to improve illumination estimation based on so-called "memory colour" (i.e. colours related to certain classes of objects) [RNK+09].

2.6.4 Physics-based Algorithms

The simplified Lambertian image formation model (see Equation (2.1)) constitutes the foundation of the majority of illumination estimation algorithms. By contrast, physics-based algorithms determine the illuminant by using a known physical process assumed to exist (e.g. specularities or mutual illumination) [Tom96, FDH91, Lee86, NBE58]. The assumption underpinning algorithms using specular highlights is that the dichromatic reflectance model can be used to model the reflectances occurring in the world. The expression of the dichromatic reflectance model is as follows:

$$\rho_k(x) = m_b(x)S(\lambda)E(\lambda)Q_k(\lambda) + m_s(x)E(\lambda_b)Q_k(\lambda_b) \quad (2.16)$$

In the above, the relative surface amount is denoted by m_b, while the relative amount of specular reflectance is denoted by m_s. These two factors contribute total amount of light reflected at position x.

Physics-based algorithms are geared towards identifying pixels with an m_b value of zero or close to zero, as they are indicative of the reflection of the greatest part of light in a scene (Equation (2.9) suggests that the colour of pure specularities is identical to that of the light source). As the colours of these pixels is close to or the same as the colour of the light source, they can be used to produce the estimation.

On the downside, identification of specular highlights in images is difficult. For instance, the image pixels with the highest brightness level may be specular but they could just as well not be; furthermore, in the case of pixels that are specular, there is the additional matter of lack of purity. There are algorithms for extracting "pure" specular colour based on merging information from multiple surfaces [Sha85] but their efficiency is usually restricted to the laboratory environment.

To generate additional information about the colour of the light source, mutual illumination has been employed in some other physics-based methods [Tom96, FDH91]. The incidence of light reflected from one surface on another surface is what yields mutual reflection [FDH91]. Investigation of the changes undergone by the sensor responses in the area of mutual reflection revealed that the use of the additional information available enabled a satisfactory estimation of the surface reflectance function. However, due to the difficulty involved in detecting areas of the image with occurrence of mutual reflection hinders the use of these approaches on a wide scale outside the laboratory environment.

There are considerable differences between the dichromatic reflectance model and the topic under consideration, and therefore current thesis will focus solely on the simplified Lambertian model and adopt its assumption in method development.

2.7 Image Datasets for Assessment of Computational Colour Constancy Algorithms

The benchmark image datasets employed in this thesis to analyse how colour constancy algorithms perform are the Simon Fraser calibrated datasets A and B [BMFC02**?** , BMCF02], the Grey Ball dataset [CF03], and Gehler's Cambridge dataset [GRB$^+$08].

2.7.1 Simon Fraser University

The images in the datasets of the Simon Fraser University (SFU) have been produced in a laboratory environment [BMFC02, BCF02, BMCF02] and include thirty-one scenes of coloured objects taken under eleven distinct illuminants. The format of the images is the standard 8-bit TIFF, although a 16-bit TIFF format is provided as well to enable testing with extra dynamic range.

The group A and group B (synthetic data) consisting of 321 and 220 images, respectively, are the two image groups usually employed for this dataset. Images from this dataset are exemplified in Figure (2.8).

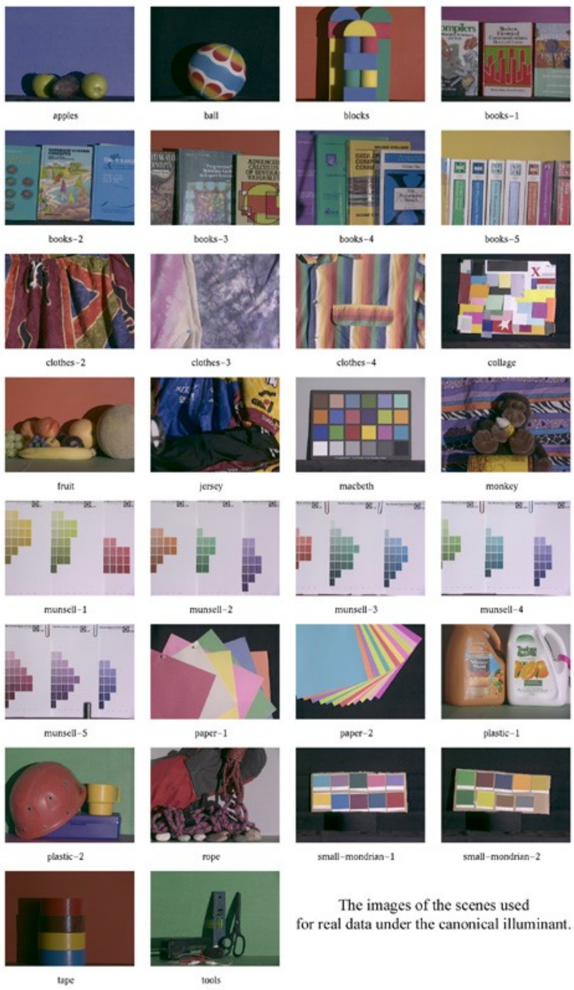


Fig. 2.8 Examples of images in the Simon Fraser University (SFU) dataset.[FL00]

2.7.2 The Ciurea and Funt Grey Ball Set

The over 11,000 images making up the Grey Ball dataset [BMF00] have been taken under a wide range of lighting conditions. More specifically, the images were derived from a video that was recorded over a period of two hours in different conditions, both outside and inside. Every scene shows a small grey sphere that was fastened to the camera and subsequently enabled the measurement of the light RGB in each scene. Figure (2.9) illustrates some images from this dataset.

Fig. 2.9 Examples of images in Ciurea and Funt's Grey Ball dataset.[BMF00]

As far as the current researcher is aware, no other dataset for illumination estimation is as extensive as the Grey Ball. It presents the drawback that the images in the same sequence are highly correlated because they are derived from a video. Different studies have suggested that a representative subset should be employed rather than the whole set; for instance, a subset of 150 images (10 images for every scene) was used in [AGA07], being deemed sufficiently representative of the entire set. Conversely, a subset of 1135 images was suggested by another study [BCCS10].

2.7.3 Gehler's MacBeth Colour Checker Set

The 568 images making up Gehler's dataset [BF97] have been captured under varying lighting conditions and include scenes with rich content, depicting people, landscapes, buildings, and objects in both indoor and outdoor environments. Each scene contains a Gretag MacBeth Colour Checker Chart, which enables the determination of the light source of the scenes based on the examination of the achromatic patches. The dataset is advantageous primarily because of the realism and high quality of its images. Some of the images included in the dataset are exemplified in Figure (2.10). A linear version of the originally non-linear Gehler's dataset has been developed by Shi and Funt [FL00] and is publicly accessible [SF10].

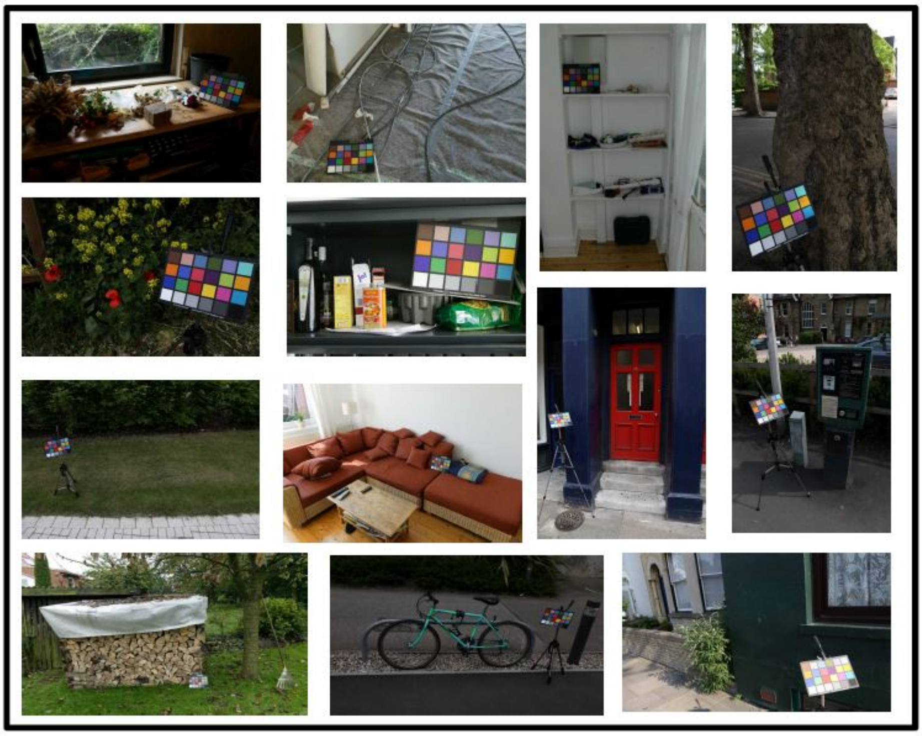


Fig. 2.10 Examples of images in the Gehler dataset.[BF97]

2.8 Conclusion

The concept of colour constancy and related computational colour constancy algorithms have been reviewed in the present chapter. Colour constancy refers to the ability of the visual system to eliminate the effect of the light source. Efforts have been dedicated to replicate this ability in contemporary colour imaging and reproduction. In this context, the purpose of computational colour constancy algorithms is to approximate the illuminant and use finite image data to generate colour invariant object estimations. Illuminants, objects and the equivalent imaging system are the three factors influencing the development of colour signals in images. The features and categories of these factors have been addressed in this chapter. In addition, the von Kries model and different computational colour constancy algorithms, including Grey World coupled with the Retinex theoretical framework, statistics-based methods, Gamut mapping algorithms, and learning-based methods, have been discussed in the chapter in terms of their background, underlying assumptions, implementation and shortcomings. Last but not least, a series of commonly used image datasets for assessment of illumination estimation algorithms have been introduced. Based on the foundation provided by this review of the latest methods and approaches, the development of a new learning-based algorithm for approximating the light source will be attempted, as outlined in the next chapters.

Chapter 3

Illuminant Classification based on Random Forest

3.1 Introduction

From Equation (2.1), we can see the colour response ρ of a device to a given surface depends on both the reflectance properties of the surface and the spectral power distribution of the current illuminant. When illumination changes, so too does the colour response ρ. However, the human vision system, as a robust colour-based system, can ensure the perceived colour of objects remains relatively constant under varying illuminant conditions, because it can recognize the object colour changing caused by the illuminant change, then recover the influence made by illuminant change. This ability to account for the court of the light source is called colour constancy and helps humans identify objects [L$^+$77]. However, from a computer vision perspective, this illumination dependence is a problem since the digital vision system of devices like cameras only honestly record the colour that they receive, so in such a changing illuminant scene, the same object will be recorded as many different colours. This implies that using colour as a cue to help solve fundamental vision tasks, such as scene segmentation, object recognition and tracking might run into problems if the scene illumination is changing. Solving colour constancy is also an important step in a digital camera's processing pipeline since it is well-established that, if image colours are not corrected to account for the scene illuminant colour, the image will look quite different to how an observer remembers the scene. For these reasons, solving the colour constancy problem is one of the most challenging and important problems in the field of computer vision research.

If we pose the colour constancy problem as the problem of estimating the illuminant in a given scene, then the problem can be formulated as how we can invert Equation (2.1) to recover the scene illuminant SPD $E(\lambda)$. For a single surface we have only three measurements (the elements of λ), and both $E(\lambda)$ and $S(\lambda)$ are unknown. This implies that the problem is under-constrained. If we assume that $E(\lambda)$ is constant throughout an imaged scene, adding more surfaces provides us with additional constraints on the illuminant. However, it also introduces further unknowns: the reflectance functions of the additional surfaces. Thus, we cannot solve this problem in this way, so to deal with the under-constrained nature of the problem, we have to make additional assumptions about the world. This has prompted many studies in the related literature to address colour constancy methods works [AAKA06, VDWG05, VdWSV07, For90, GRB$^+$08].

In this chapter, under the assumption that images taken under the same illuminant should have similar colour appearance, we present a novel learning-based technique for colour constancy. Based on a powerful machine learning model, Random Forest, we treat colour constancy as a problem of inferring the illuminant class of the input image. Random forest is a tree based data structure; our technique is therefore intrinsically fast and efficient.

There are two main contributions we make in this work: firstly, we develop a new image representation feature suitable for developing classification or learning based colour constancy, and secondly, we develop a novel colour constancy framework that can classify or recognise illuminants based on Random Forest. We will present experimental results on a number of publicly available testing databases and show that our new technique outperforms state-of-the-art techniques.

3.2 A Novel Random Forest Approach to Colour Constancy

The basic idea of our Random Forest-based approach is to treat the problem of estimating the unknown illuminant of the input image as an illuminant classification problem. Random Forest is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees [Bre01]. It is one of the most accurate learning algorithms and for many data sets it produces a highly accurate classifier [LW02]. Random Forest is a generic classification technique, and here we present an approach to adapting it for tackling the colour constancy problem.

3.2.1 The Selection of Colour Values in Illuminant Estimation

An efficient framework for modelling real-world data is highly important, just like with any issue of pattern categorisation. A major difficulty for colour constancy algorithms is extraction of colour features from images to perform effectively. In this regard, the main approach that is adopted is concerned with selection of colour values.

The majority of algorithms rely on image colour signals to estimate illuminants. The input usually consists of raw data from digital cameras (e.g. three-channel *RGB* signals). In some cases, conversion of the colour signals to certain data without device dependence (e.g. XYZ or $L*a*b*$) is performed [FR98]. However, since the precision of diagonal transformation is diminished in device-independent spaces, device outputs are employed directly in the majority of techniques for illumination estimation. Gamut mapping techniques afford greater importance to colour space due to its impact on gamut characterisation and, implicitly, on the performance of the techniques. However, the choice of colour spaces in Gamut mapping techniques has been shown to have an influence on illumination estimation. CRULE was the original Gamut mapping technique in which characterisation of the 3D gamuts was undertaken based on the (R,G,B) space [For90]. However, the technique presented shortcomings regarding the application of absolute colour values, so it was improved by a different study [Fin96] based on projection of the (R,G,B) colours to a 2D relative colour space as $(R/B,G/B)$. Colour by Correlation is another widely used Gamut mapping technique that is based on the chromaticity space (r,g). Similarly, the (r,g) image signals are employed by the neural network technique [CFB02] as the input information for the neural network system. Meanwhile, the sensor correlation technique [TEW01] is carried out in the absolute colour space (R,B) and its underlying assumption is that a greater amount of information about the light source is contained in colours with a higher level of lightness.

The choice of colour spaces has been the focus of ample debate, particularly with regard to whether their composition should include absolute or relative colour values and how many dimensions they should have. 3D and 2D colour spaces differ in that the former rely completely on the three-channel colour information from cameras, whereas the latter make it easier for the gamut calculation to be performed by leaving out one information channel. Essential for reflection of the illuminant attributes, the lightness information is incorporated in the colour spaces employing absolute values [FHH97]. By contrast, lightness information is not included in colour spaces using relative values and image data processing is two-dimensional. In these spaces, gamuts are primarily indicative of low-lightness colours and therefore they do not differ significantly from various illuminants. Furthermore, signals with small values are more affected by noise, which means that sensitivity to noise is greater among colours spaces using relative values.

Colour constancy research started using probability distributions due to the introduction of a Bayesian model-based approach to illumination estimation [BF94, BF97]. The premise underpinning this approach is that a previous probability distribution determines how likely the occurrence of a specific illuminant or surface is in a scene. Representation of this prior probability distribution takes the form of multivariate normal distribution for illuminant power distributions as well as for surface reflectance functions. Colour indexing is a different technique for detecting objects [FF95] that considers the inherent attribute of objects to be their colour probability distribution. It achieves object detection by comparing the chromaticity histograms of objects in images and in the object database. Colour indexing has prompted the use of chromaticity histograms in probability distribution representations, despite not being classified as a colour constancy algorithm. In addition to offering key statistical data for images, chromaticity histograms are indicative of the light source under which the images were captured as well.

In the case of the neural network technique, the chromaticity values of an image are used as input data for the neural network system [CFB02]. This is in fact the chromaticity histogram of the image. Meanwhile, in the case of Colour by Correlation, the potential colour distributions under various light sources are exploited by the correlation matrix. The chromaticity histograms of a given reflectance database usually supplies the information.

In our technique, we first derive two opponent colour signals, Red-Green (RG) and Blue-Yellow (BY), from the original RGB signals according to Equation (3.1):

$$RG = R - G, \quad BY = (R+G)/2 - B \tag{3.1}$$

The RG, BY value is in the range $(-255, 255)$, we translated this range to $(0, 255)$ by two following steps:

Firstly, add 255 to each value then the range is translated to $(0, 510)$; Secondly, divided by 2 of each value and round up to an integer, then the value is translated to the range $(0, 255)$.

We then treat each of the two opponent signals as an 8-bits per pixel greyscale image and construct pixel histograms H_{RG} and H_{BY} for the RG image and the BY image respectively. Here H_{RG} and H_{BY} are 256 dimensional vectors. $H_{RG}(i)$, i = 0, 1, 2,$\cdots$, 255, is the frequency of the pixels in the RG image having a quantized value of i, and similarly, $H_{BY}(i)$, i = 0, 1, 2,$\cdots$, 255, is the frequency of the pixels in the BY image having a quantized value of i.

The two histograms are obviously dependent on the contents of the image. As our task is estimating the colours of the illuminant, our representation should be relatively independent of content; otherwise the image contents will cause estimation errors. For example, a yellow leaf in daylight may be estimated wrongly as in Tungsten light because of the high frequency of yellow colour in the image. To avoid this problem, we only take into account the existence

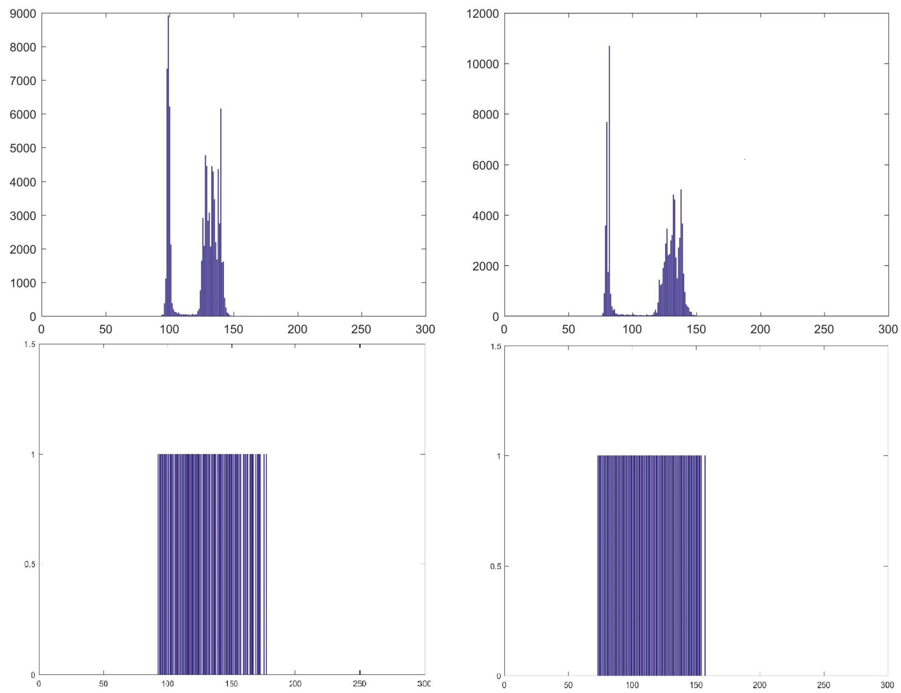


Fig. 3.1 The figure is shown the binary histogram which is content independent, only labelled the colour appeared. The first row is the original RG,BY histogram, and the bottom row is the one after binary by Equation 3.2.

of a colour rather than the number of times it appears in the image or the number of pixels having that colour. A simple way to achieve this is to binaries H_{RG} and H_{BY} to ensure that each chromaticity is counted only once. Our image representation feature is therefore two 256 dimensional binary vectors obtained as Equation (3.2):

$$h_{RG}(i) = \begin{cases} 1, & H_{RG}(i) > 0 \\ 0, & otherwise \end{cases} ; h_{BY}(i) = \begin{cases} 1, & H_{BY}(i) > 0 \\ 0, & otherwise \end{cases} \tag{3.2}$$

And finally, we cascade the two binary vectors h_{RG} and h_{BY} into one 512 dimensional feature vector $F_{RGBY} = (h_{RG}, h_{BY})$ for the construction of our Random Forest-based recognition or classification based colour constancy algorithm. The RG, BY binary histogram is shown in figure 3.1.

3.2.2 Constructing the Illuminant Recognition Random Forest

To generate a decision tree, the first step is the choice of the split function. In our case, we assumed the images under the same illuminant would have similar feature vector F_{RGBY}. So the split function will be based on F_{RGBY}. In building a random tree, we randomly select a dimension of the feature vector of the input image and then based on the value of that chosen dimension distribute the image to either the right child or the left child of a binary tree node. For example, if dimension i of F_{RGBY} has been selected as the testing dimension, the training images will be split according to Equation (3.3):

$$\begin{cases} Image\ goes\ to\ the\ left\ child, & F_{RGBY}(i) = 0 \\ Image\ goes\ to\ the\ right\ child, & F_{RGBY}(i) = 1 \end{cases} \tag{3.3}$$

After each split, the input images at the current node I_n will be divided into two groups: I_l and I_r; these are then further split until it has reached the leaf nodes.

For a typical decision tree, we should generate multiple hypothesised tests and pick the best split dimension and threshold value as the split function of each tree node. At each non-leaf node, N different dimensions in the feature vector F_{RGBY} are randomly selected and then we perform N different splits according to Equation (3.4). For each split, we calculate the maximum information gain to evaluate the merit of the split. At each node, after splitting the samples to the left node or the right node, we can compute their corresponding illuminant class histograms. The class information of the training dataset is the known illuminant of each training image. A good split means the class histogram of the left node should be quite different from the class histogram of the right node. The score function is:

$$Score(split) = \Delta E = -\frac{|I_l|}{|I_n|}E(I_l) - \frac{|I_r|}{|I_n|}E(I_r) \tag{3.4}$$

where $E(I)$ is the Shannon entropy of the class distributions in the set of samples I. $|I_l|$ is the number of samples contained in the left child and $|I_r|$ is the number in the right. I_n is the set of training sample in node n. After comparing the scores of the splits, the split with the best score will be selected and the dimension i related with this split will be recorded at the current node as the split dimension.

The decision tree generation process will be stopped when certain tree construction criteria such as maximum tree depth have been met. Each leaf node is then associated with an illuminant distribution histogram, $H(k)$, k = 1, 2, K, where K is the total number of possible illuminant classes, and $H(k)$ records the probability of the input image being under the k-th illuminant when it falls onto that leaf node.

3.2.3 Illuminant Estimation based on Random Forest

For a test input image, after extracting the h_{RG} and h_{BY} features, we let the image features go through all the trees. When the image reaches a leaf node of the m-th tree, we save the illuminant distribution histogram of the leaf node $H_m(k)$, k = 1, 2,K, where K is the total number of possible illuminant classes. Suppose we have M trees in the random forest, we sum all M illuminant histograms together:

$$H(k) = \sum_{m=1}^{M} H_m(k) \tag{3.5}$$

The l-th illuminant is estimated as the illuminant of the input image if $H(l) \geq H(k)$ for all k.

3.2.4 Recover the Colour Under Canonical Illuminant

After labelling the l-th illuminant as the estimate illuminant for the input image, we can simply use the diagonal model of illumination change to recover the image colour like taken under a canonical illuminant. The diagonal model is independent of the colour channel; for example, the colour signal values of a white patch taken under a white, canonical illuminant is (R_c,G_c,B_c), and the response under a unknown illuminant is (R_e,G_e,B_e). Then the mapping from the unknown illuminant to canonical illuminant can be achieved by scaling the tree channels by (R_c/R_e,G_c/G_e,B_c/B_e), and the same scaling works for other, non-white patches. This is indicated by Equation (3.6):

$$\begin{bmatrix} R_O \\ G_O \\ B_O \end{bmatrix} = \begin{bmatrix} \frac{R_c}{R_e} & 0 & 0 \\ 0 & \frac{G_c}{G_e} & 0 \\ 0 & 0 & \frac{B_c}{B_e} \end{bmatrix} \begin{bmatrix} R_I \\ G_I \\ B_I \end{bmatrix} \tag{3.6}$$

R_I is the input pixel red channel value and R_O is the output after the correction, same as the other two colour channels. Under ideal condition, a white patch under canonical illuminant, the colour channel values (R_c,G_c,B_c) should be (255,255,255). However, as our work only focuses on estimating the unknown illuminant for the input image, so the error we tested is based on our estimated illuminant value and the ground truth illuminant value, as well as other algorithms we used for comparison purposes.

3.3 Algorithms Evaluation

In the previous sections, we have described a general Random Forest framework for illuminant classification. We now evaluate its performances for various parameter settings and compare it with that of other leading colour constancy methods.

We evaluate our algorithm performance on two public benchmark datasets: The first dataset consists of 321 images of constructed scenes taken under 11 different illuminant sources in the Lab SFU [BMFC02]. For each illuminant source the spectral distribution is known. Thus the dataset is of high quality but does not represent the full variation of typical scenes as outdoor scenes are missing. The second database [BMF00] is a large database containing 11346 images in several indoor or outdoor scenes. The images were actually frames in a video; each image has a grey ball at the right down corner to calculate the ground truth of the illuminant.

Following a common practice in the literature, we calculate the angular error according to equation 3.7 from [BCF02] as the performance indicator:

$$Error_{angular} = cos^{-1}((T \cdot E)|T|^{-1}|E|^{-1}) \tag{3.7}$$

where T is the ground truth illuminant value and E is the estimated value by the colour constancy algorithms. The error is calculated by the degree distances between the two colour vectors. Note that results with Lab SFU dataset are averaged over 50 different random trails and results with grey ball dataset are averaged over 30 trails. Futhermore, the dataset is 80% for training and 20% for test.

3.3.1 Random Forest Parameter Setting Evaluation

There are several important parameters that could influence both the computation cost and the algorithm performance, so before initiating the experiments, we need to set these parameters. All the values of the parameters we set are from the test based on Lab SFU datasets and there are 11 classes in this dataset. In each class, 80% of the image is used for training and the remaining 20% for test. The original images are under *sRGB* colour space.

Different Colour Features

In addition to using the *RG* and *BY* colour signals to drive the image representation features as described in Section 3, we have also tested using other colour signals to derive the binary histogram of equation (3) for building the Random Forests, including, the original R, G and B signals to build a 3x256 dimensional binary feature vector; $r = R/B$ and $g = G/B$ to build a 2x256 dimensional binary feature; and *rg* and *by* to build 2x256 dimensional binary feature, where $r = R/(R+G+B)$, $g = G/(R+G+B)$, and $b = B/(R+G+B)$, $rg = r - g$, $by = ((r+g)/2 - b$.

Lastly, we have also tested the *RGBY* colour features in a 2D way: F'_{RGBY} = (h_{RG}*h_{BY}), then the feature F'_{RGBY} will be a N*N matrix. We tested this feature still in a one-dimensional way: we reshaped the F'_{RGBY} from (N,N) to $(1,N*N)$, to match the input of the forest.

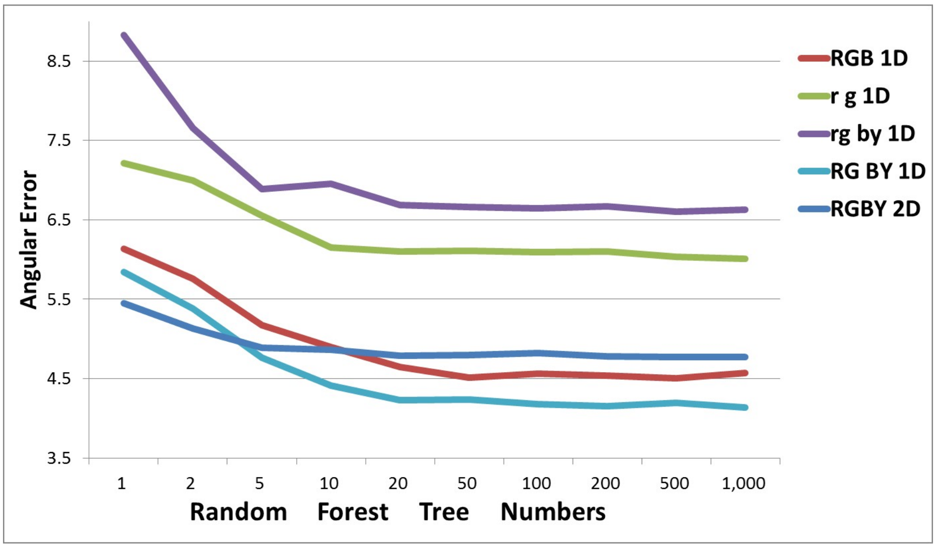


Fig. 3.2 The algorithm performance when different colour features are used to construct the image representation features. The X-axis is the tree number, and the Y-axis is angular error result. There are results for the Lab SFU dataset.From the result we can see the *RG* and *BY* colour signals worked under 1D framework produce the best result compared to others.

The results are shown in Figure (3.2). As the number of trees increases, algorithms performance with all features is getting better. *RG* and *BY* feature have the best result when the forest scale is increasing over 5 trees. The *RG* and *BY* features under 2D framework have a more stable performance, but not the best. *r* and *g*, *rg* and *by* combinations perform worse than the original *R*, *G* and *B* signals.

There are two reasons why we developed our opponent colour feature, and both of them have been confirmed by experiment: Firstly, an opponent colour space model is more similar to human visual system[WN97], human have two different retina cells: cones and rods, deal with illuminants and chromaticity. Our opponent colour signals are doing this in the same way: *RG* channel is more sensitive with illuminant intensity as a bandpass function, and *BY* is a lowpass function which is sensitive to absolute chromaticity value [Pit00]. This could provide more information to train the forest. Secondly, one of the reasons that we transfer *R*, *G*, and *B* colour features to *RG* and *BY* opponent chromaticity colour signals is trying to reduce the feature from 3D to 2D thus simplifying the algorithm. It has been shown that using the opponent chromaticity signals can not only capture the illuminant information well but also can relax the conditions imposed on the color constancy models [Fin96]. Furthermore, experiment results have shown that these features can indeed extract illuminant information from images more efficiently.

Random Forest Scape (Number of Trees)

The random forest size is another important parameter, with an appropriate size of the forest; the algorithm could have the maximum performance and reduce the computation cost. Figure (3.3) shows the angular error results for different sizes of the Random Forests for both datasets. The algorithm can reach an acceptable result even with only one tree, and then as the number of trees increases, the performance reaches a saturation point rapidly, meaning that so long as we use a sufficient number of trees in the forest, our method will have stable performance. Our results show that using 100 trees is sufficient, and we set 100 as our Random Forest scope.

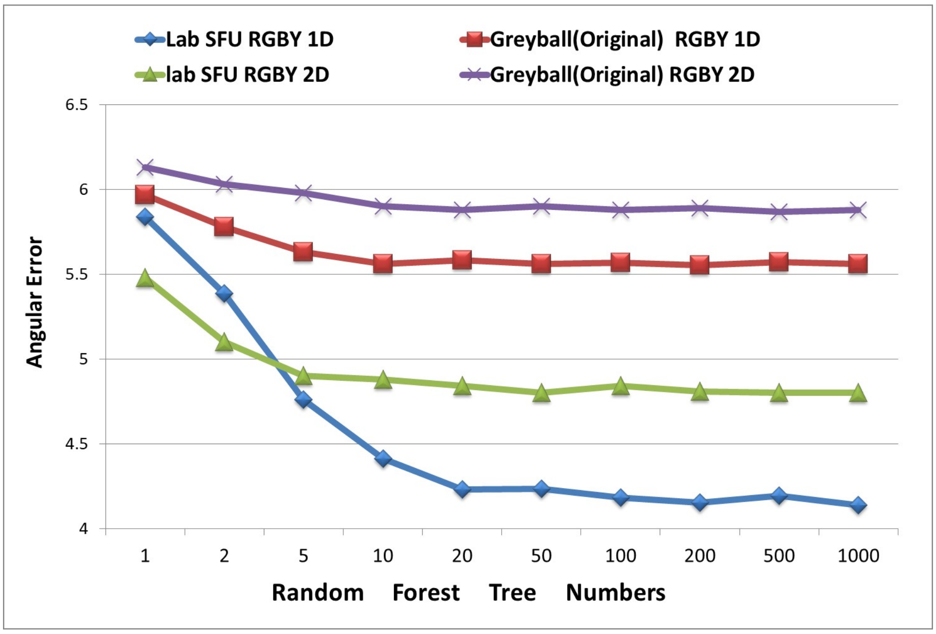


Fig. 3.3 The algorithm performance for different forest sizes. The X-axis is the tree number, and the Y-axis is angular error result. When the tree numbers reached 100, the performance is pretty stable and sufficient for both datasets.

3.3.2 Comparison of Colour Constancy Algorithms

The second experiment is the evaluation of different algorithms for the two data sets with three ground truths (The grey ball dataset's ground truth values have been modified under a linear colour space since the original data set is captured under a non-linear colour space [BMFC02]). The forest size here is 100 and the percentage for training is 80%. The results are averaged over 50 different trials in Lab SFU dataset and 30 in grey ball data set. In this comparison, we use two different ways of constructing features for training, the 1D 1*512 dimension vector F_{RGBY}, and the 1D 1*65536 vector (reshaped from 2D 256*256 matrix, will be labelled as Random Forest 2D in evaluation). The algorithms included for comparison are classic algorithms like the grey world [Buc80], the max-RGB algorithms [L+77], shades of grey [FGJ98], grey edge [GGVdW07], and the gamut mapping algorithms with different gamut weighting methods[For90, GGVDW10, CK02]. We also include some learning-based methods, such as: Using Natural Image Statistics [GG11] and Exemplar-based [JD12], which are some of the state-of-the-art methods with the best performance. All the results for these algorithms are from the website www.colorconstancy.com, an academic evaluation platform for colour constancy. Results of the Gamut mapping requires training data. The following results are obtained using 31 images (all images recorded under the "syl-50MR16Q"-illuminant) to construct the canonical gamut. This canonical gamut is then used to estimate the illuminant for all images, including the training images. This is done to be able to make a comparison with other methods. Interestingly, most of the learning-based methods were not tested on the Lab SFU data-set, the missing result was labelled as a "*". One possible reason could be that the images in this data-set are somehow insufficiently natural for the learning model as the images were taken under lab condition.

Table (3.1) shows the results of the Random Forest algorithm performance in two benchmark datasets. In all the experiment results, the Random Forest with 1D feature vector always has better results than the 2D version. In the Lab SFU data set, our algorithm is better than all of the statistical methods and also performs similarly to the gamut mapping methods. However, it must be noted that the gamut mapping results contain the estimation of the training images, so we believe the "true" performance could be a little bit worse. As grey edge and gamut mapping also need to set some parameters in different datasets (Minkowski-norm ρ is varied between 1 and 15, and the smoothing value σ is varied between 1 and 12), the results are also with the reference parameter-setting labelled from the www.colorconstancy.com, pursued as the best performance of the algorithms.

The Natural Image Statistics used Weibull parameterization, related to the image attribute (e.g., grain size and contrast), and applied an MoG-classifier to determine the correlation and weighting between the Weibull-parameters and the image attributes (number of edges,

Algorithms	Lab SFU	Grey-ball(Original)	Grey-ball(Linear)
Grey world	9.8	7.9	13
Max-RGB	9.1	6.8	12.7
Shades of Grey	6.4	6.1	11.6
Grey edge 2nd order	5.2	6.1	10.6
pixel-based Gamut mapping	3.7	7.1	11.8
Edge-based Gamut mapping	3.9	6.8	12.8
Intersection-based Gamut mapping	3.6	6.9	11.8
Natural Images	*	5.2	9.9
Exemplar-based	*	4.4	8.0
Random forest 2D	4.8	6.7	10.6
Random forest 1D	4.2	5.6	8.2

Table 3.1 Comparison of Colour Constancy Algorithms

amount of texture, and semantic information of the scene). Then, based on these image characteristics, the proper colour constancy algorithm (or best combination of algorithms) is selected for a specific image. Meanwhile, the Exemplar-based method focuses on surfaces in the image. The first step is segmentation of images by surfaces, then for each surface, the nearest neighbour models in a training image must be found and its illumination is estimated 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. In the "grey ball" dataset, in which all the images are video frames, these two algorithms could provide better results than ours (except with the linear ground truth of Natural Image Statistics, ours is better), but obviously the structures of these two algorithms are more complex. Even so, from Table (3.1) we can still see that the performance of the Random Forest framework is still close to the Exemplar-based method, thus proving the competitive position of our method. Figure (3.4,3.5) shows some visual examples of colour correction results of an image based on illuminants estimated by different algorithms. The image is randomly selected from each dataset and it is seen that the visual result of the random forest method is also better with lower angular errors compared with statistical methods and gamut mappings (for some algorithms we mentioned in our experiment, the code is needed to generate the compared photos, but this is difficult to obtain).

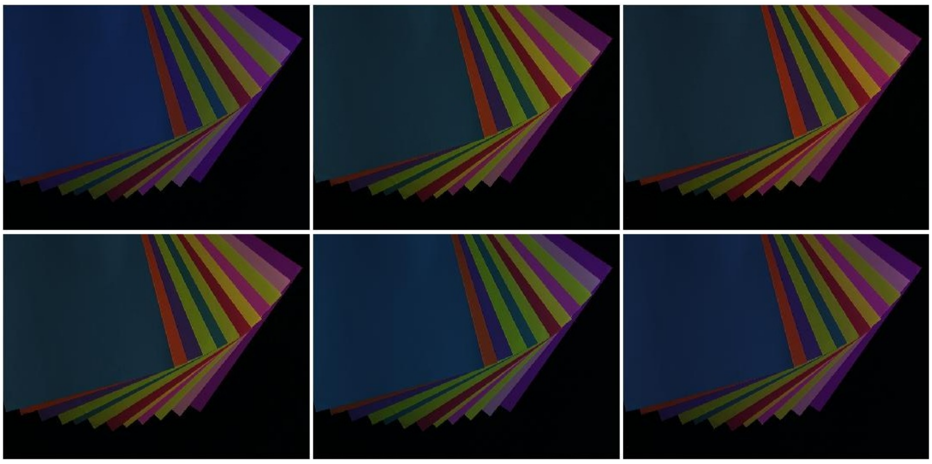


Fig. 3.4 Colour constancy colour correction results of grey world, max RGB, grey edge and Random Forest in Lab SFU dataset. Top row, left: original, middle: corrected with ground truth data, right: Random Forest (angular error 4.6). Bottom row, left: grey world (angular error 6.9), middle: max RGB (angular error 11.8), right: grey edge (angular error 15.5).

3.4 Conclusions

In this chapter, a new colour constancy algorithm has been introduced. Based on the assumption that images taken under the same illuminant should have similar colour appearance, treating the unknown image illuminant estimation task as an illuminant classification problem, we adopted a new colour features to construct the forest: *RG* and *BY* colour signals, and used random forest method as a strong classifier to achieve the goal of colour constancy: estimate the unknown illuminant for input image. Compared with other algorithms on two benchmark datasets, random forest is more efficient and can outperform other state of the art algorithms.

However, as a classification method, Random Forest can only provide a few discrete illuminant estimations based on how the datasets have been classified. In the two test datasets, the SFU has the most specified illuminant classes (11 strongly controlled lab-based illuminants). For each illuminant, images under the same classes have a very similar ground truth, so we could say in this case the SFU dataset has an ideal classification for our Random Forest algorithm. In the second grey ball dataset, the challenge is that the dataset only briefly divided into 15 categories, and it is not even based on the illuminant but the scenes. Under this situation, even in the same class, the ground truth variance is more diffused than the SFU lab dataset. Compared with continuous methods, like grey world, if the illuminant situation is very complicated, the performance of random forest would get worse, due to the lack of efficient classification of the dataset. How to divide the dataset into an appropriate number of classes could be an issue for further improvement. This issue will also have a negative

Fig. 3.5 Colour constancy colour correction results of grey world, max RGB, grey edge and Random Forest in Grey ball dataset. Top row, left: original, middle: corrected with ground truth data, right: Random Forest (angular error 3.3). Bottom row, left: grey world (angular error 1.1), middle: max RGB (angular error 11.3), right: grey edge (angular error 9.3).

influence on the final recovered image. Providing a few discrete illuminant estimations means providing only a few discrete recover results. If an image is taken under a non- standard illuminant, or an illuminant does not exist in the current training dataset, the angular error could be significantly high. Therefore, based on the current foundation, future work should focus more on Random Forest regression development, to provide a similar continuous function.

Chapter 4

Experimental Work Based on Synthesised Data

4.1 Introduction

The limitation of the classification framework for illumination estimation has been addressed in the preceding chapter. Ideally, the natural illuminance distribution ought to be a continuous function; however, the number of illuminant categories supplied by the datasets gives the possible number of estimation results if the data classification is not detailed enough or has an insufficient number of classes. Two possible options are available to enhance the algorithm and address this issue, namely, employment of a more classified training dataset or application of the clustering technique for a more specific classification of the data.

The proposed implementation approach based on synthetic dataset is outlined in the present chapter. This dataset contains data on illuminant spectral distribution, surface reflection, and camera characterisation. A comparison of the proposed approach with several important algorithms is conducted and the outcomes of experimental work employing the synthetic dataset are presented. The significance of such experimental work lies in the fact that there is knowledge of the ground truth, ambiguities related to camera characterisation and pre-processing are unlikely to occur, and accurate manipulation of each of the different factors influencing colour constancy is possible, thus affording a good insight into those factors. Furthermore, compared to other data, synthetic data could supply a greater number of classes for our classification framework training because they can be employed to produce a greater amount of illuminant spectral information.

4.2 Synthetic Dataset

The synthetic dataset was first introduced in [BMFC02] and can be accessed over the Internet (http://www.cs.sfu.ca/ colour/data). It has high relevance for computational colour constancy research. The 321 images of scenes captured under eleven carefully selected light sources constitute the essential part of the dataset. They have been employed in the earlier test. In addition, the dataset includes a number of standardised sets of spectra for synthetic data empirical work, as well as data for fluorescent surfaces.

Given that the present thesis is concerned primarily with unique illuminant settings, the premise regarding the homogeneity of illumination across the whole (synthetic) scene is particularly important.

4.2.1 Illuminant Spectral Data

Illuminant spectral data measured in the human visual range constitute the first element of the dataset. In keeping with the format employed by the PhotoResearch PR-650 spectrophotometer, the spectra are made up of 101 measurements in the range 380-780 nm in 4 nm steps. A sample of illuminant spectral distribution measured with the spectrophotometer is illustrated in Figure (4.1).

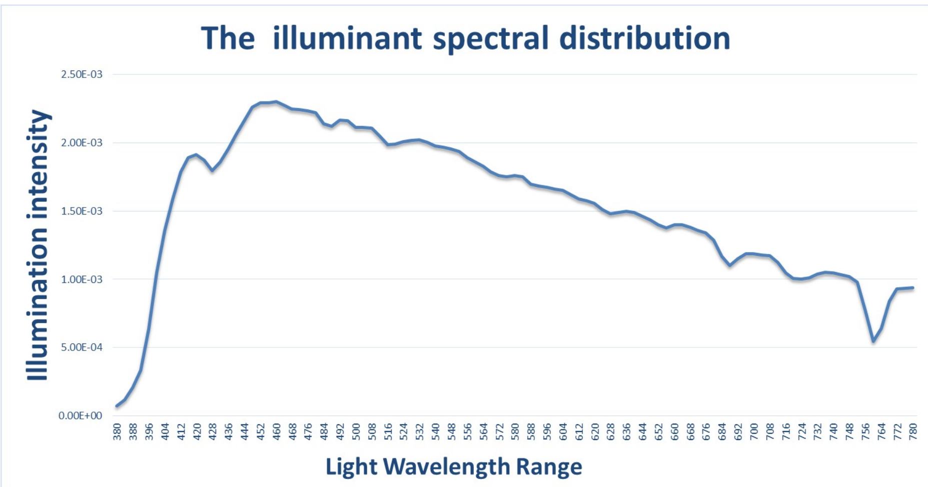


Fig. 4.1 The spectral distribution of the illuminant in visual light wavelength range; the X-axis and Y-axis respectively indicate the wavelength and the intensity of illumination reaching the spectrophotometer.

As previously mentioned, the image data in the Simon Fraser University (SFU) dataset are associated with eleven sources of illuminant spectra (see Chapter 2.6), which were chosen in such a way so as to provide maximal coverage of the range of chromaticities of both natural and artificial illuminants, whilst taking into account stability over time and physical appropriateness. The sources comprise three fluorescent lights (Sylvania warm white, Sylvania cool white, and Philips Ultralume), four different incandescent lights with power of 12 V, and four lights employed together with a blue filter (Roscolux 3202). There is a close similarity between the spectrum of the incandescent source Sylvania 50MR16Q and an ordinary incandescent lamp. The spectra of the other three are comparable to daylight of three distinct colour temperatures (Solux 3500K, Solux 4100K, Solux 4700K). The bulbs cover the outdoor illumination range to an acceptable degree when they are employed alongside the blue filter.

4.2.2 Surface Reflection Data

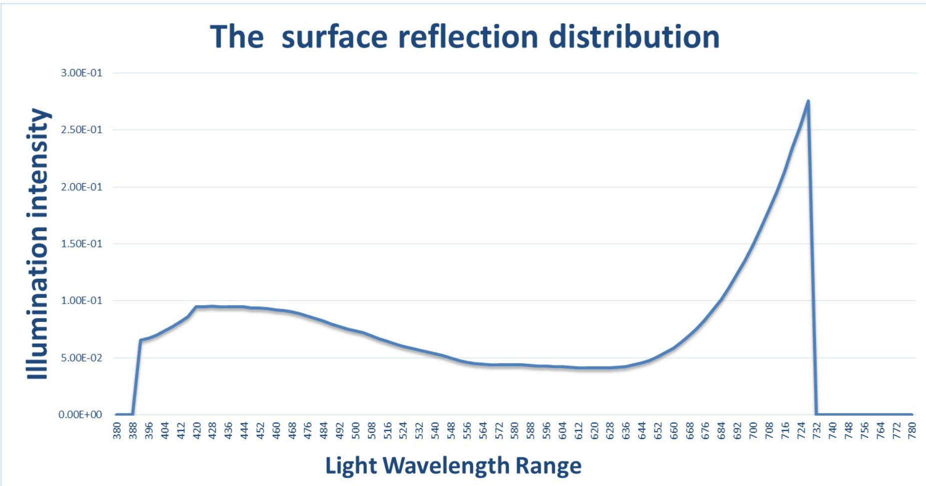


Fig. 4.2 A sample of the intensity distribution of surface reflection; the X-axis and Y-axis respectively indicate the wavelength and intensity of the illumination reaching the spectrophotometer.

The surface reflection intensity distribution data constitute the second element of the synthetic dataset. A number of sources have provided the 1995 surfaces in the dataset, which comprise 24 Macbeth Colour Checker patches, 1269 Munsell chips, 120 Dupont paint chips [VGI94], 170 natural objects [VGI94], 350 surfaces from the Krinov dataset [Kri53], and 57 extra surfaces whose measurement was undertaken in [BCF02]. In addition to being selected as a superset of the reflectance sets employed in other colour constancy studies, this set helps to model the world to calibrate (train) the algorithms and to assess the algorithms. A sample of surface reflection intensity distribution is illustrated in Figure (4.2), with the

Sylvania 50MR16Q serving as the canonical light source and reflecting the light reaching the spectrophotometer.

4.2.3 Camera Characterization Data

The canonical illuminant employed in the current test is the Sylvania 50MR16Q, because this is the illuminant that can ensure optimal camera balance. In other words, the camera response to pure white is more or less identical for each of the three channels under this illuminant. The individual measurement of the camera response function by the three colour channels is depicted in Figure (4.3).

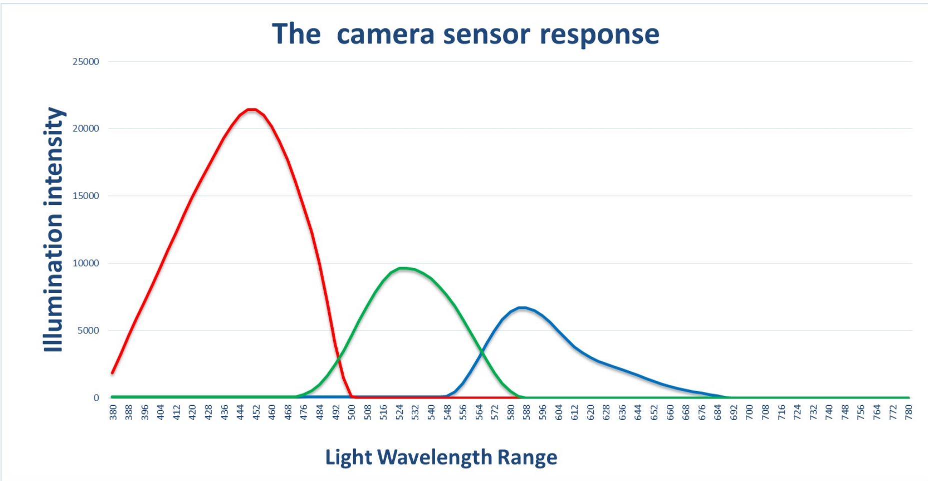


Fig. 4.3 The response function of the camera sensors under Sylvania 50MR16Q as the canonical light source; the X-axis and Y-axis respectively represent the wavelength and intensity of the illumination reaching the spectrophotometer and the three colour channels (R, G, B) each being used for the measurement of the response function.

4.3 Production of Synthetic Data

The benefit of using synthetic images in experiments is that they enable control of the entire environment. Furthermore, the reliability of results from algorithm assessment is ensured through the production of images in a random extensive amount. Synthesised images also enable users to establish the number of patches that constitute a scene, every chromaticity occurring in the image being allocated a patch.

Under a given light source, the calculation of the RGB colour of a patch (surface) is carried out with the formula below, based on its surface reflectance S^j, the spectral distribution of the illuminant E^k and the spectral sensitivities of camera sensors ρ:

$$R = \sum_i E_i^k \cdot S_i^j \cdot \rho_i^R, G = \sum_i E_i^k \cdot S_i^j \cdot \rho_i^G, B = \sum_i E_i^k \cdot S_i^j \cdot \rho_i^B \tag{4.1}$$

The wavelength domain is spanned by the index i, equivalent to wavelengths in the 380-780 nm range. The sensor sensitivities of a SONY DXC-930 CCD video camera are used, with disabled gamma correction and with calibration for tungsten light.

For every illuminant, the ground truth is determined first. Ideally, the surface reflectance in the *RGB* channels would be perfect reflectance, and therefore Equation (4.2) can be applied to obtain the illuminant ground truth:

$$R = \sum_i E_i^k \cdot \rho_i^R, G = \sum_i E_i^k \cdot \rho_i^G, B = \sum_i E_i^k \cdot \rho_i^B \tag{4.2}$$

The next step is selection of the illuminant spectra for every research section in such a way as to cover more or less homogeneously the (r, g) chromaticities of the standard conditions of illumination. Normalisation of every illuminant spectrum is performed to ensure a maximum response of 255 among the RGB channels for the pure white captured by the camera under every light source.

Figure (4.4(a)) presents the chromaticity of every one of the eleven illuminants in relation to the camera. These illuminants are placed in the relevant cells with a width of 0.02 units that were created through the division of the (r, g) space.

The 97 spectra making up the second illuminant were measured within and around the SFU campus during different hours of the day and in different climatic conditions. Although abnormal lighting (e.g. near neon advertising lights) was not included, certain reflected light was included, as long as it was within a reasonable range. Figure (4.4(b)) illustrates the chromaticities associated with this illuminant.

Arbitrary linear combinations of spectra are employed from the two sets in order to achieve the targeted coverage density. What warrants this is the fact that light from at least two sources is frequently merged in illumination. Furthermore, the developed illumination spectra will demonstrate behaviour of physical sources with chromaticities identical to the developed ones, provided that the diagonal model is maintained. The chromaticities of the training set derived through this approach are presented in Figure (4.4(c)).

The same procedure is applied to obtain the illuminant set for testing, with the exception that the density of the (r, g) space is four-fold greater. Figure (4.4(d)) illustrates the chromaticities obtained.

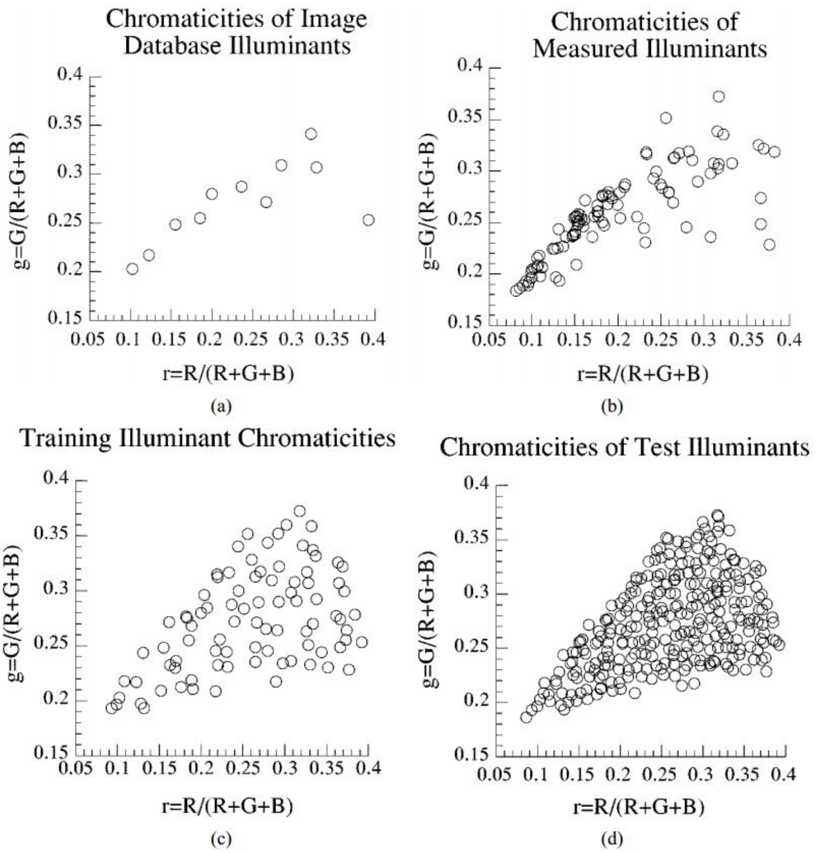


Fig. 4.4 The chromaticity distributions of the different illuminant sets employed in the present paper [BMFC02]; (a) the eleven illuminants used in the development of the test images; (b) the plotting into space of the chromaticities of an extra set consisting of multiple sources, including several illuminations measured within and around the SFU campus; (c) the training set developed on the basis of these sources; (d) a comparable set employed for testing with a greater density of population in the chromaticity space.[BMFC02]

Figure (4.5) shows the (r, g) chromaticities of the reflectance dataset captured with the Sony DXC-930 camera under the canonical light source.

4.4 Experiments

As we mentioned before, the advantage of using synthesised data is generating any arbitrary number of illuminant classes for our Random Forest framework training, and for each class, a large number of training samples (images) could be generated as well. So in our case, two experiments are produced. The first experiment involves comparative analysis with the prior experiment carried out in [BF02]. In this experiment, to make the comparison fair

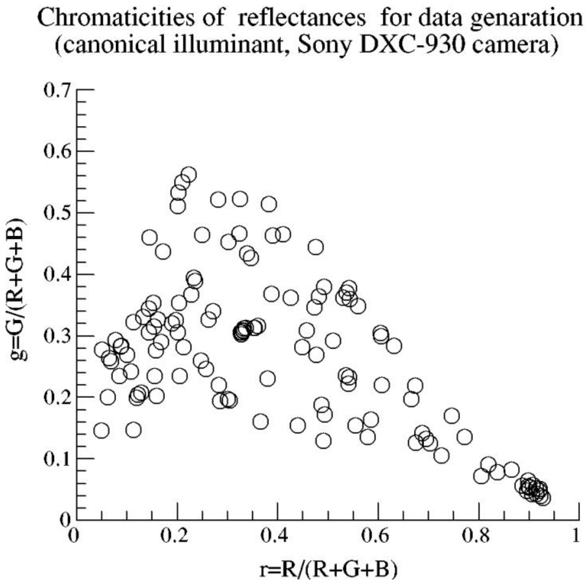


Fig. 4.5 The chromaticities of the reflectance dataset captured with the Sony DXC-930 camera under the canonical light source. [BMFC02]

enough, we follow every step in [BF02] strictly; while the second experiment involves use of synthetic data for training and real-world image datasets for testing, which is our original target, to evaluate the efficiency of the synthesised data using for estimating the real world illuminants.

4.4.1 Error Measures

Three major error measures are usually employed in colour constancy research.

In the first error measure, the illuminant ground truth (R,G,B) and the equivalent estimate are considered vectors in (R,G,B) space and the angle between them is determined in degrees. The angular error is calculated according to the equation below, if the target illuminant and the estimate are respectively $\mathbf{T} = (R_T, G_T, B_T)$, and $\mathbf{E} = (R_E, G_E, B_E)$:

$$\cos^{-1}((T \bullet E)T^{-1}E^{-1}) \tag{4.3}$$

The second error measure represents the vector distance in (r,g) space of the illuminant chromaticity and the corresponding estimate. It is calculated with the equation below, if $(r_T, g_T) = (R_T/S_T, G_T/S_T)$ and $(r_E, g_E) = (R_E/S_E, G_E/S_E)$ with the target illuminant chromaticity and the estimate respectively being $S_T = R_T + G_T + B_T$ and $S_E = R_E + G_E + B_E$:

$$\sqrt{(r_E - r_T)^2 + (g_E - g_T)^2} \tag{4.4}$$

Although the two measures can be interchanged to a greater or lesser extent, the first measure is more relevant for (R, G, B) algorithms, while the other measure is nearer the amount targeted for reduction by a number of chromaticity algorithms.

The third error measure is concerned with the error in the end-result of colour constancy, representing how much the corrected image differs from the precise target image captured under the canonical light source. Such a result necessitates registered images with identical geometry for every illuminant, which is why it is not easy to obtain with image data. To fulfill the prerequisite of identical geometry, every illuminant must derive from the same source alongside filters, which means that a general set of illuminants as employed in the present study is unsuitable. Although in the restricted case of chromaticity mappings, this issue is less pronounced, geometry fluctuations are still problematic. Synthetic data facilitate examination of results and in this thesis the *RMS* error is employed across synthetic scene surfaces in (r, g) between the target data and the corresponding mapped estimate. More to the point, the calculation of the *RMS* error can be performed through the equation below if the observed response for channel or chromaticity component k for pixel i and the equivalent amount for the target image are respectively considered to be ${\rho_i}^{(k)}$ and ${\tau_i}^{(k)}$:

$$\left(\frac{1}{N} \sum_{i}^{N} \frac{1}{K} \sum_{k}^{K} \left({\rho_i}^{(k)} - {\tau_i}^{(k)} \right)^2 \right)^{\frac{1}{2}} \tag{4.5}$$

In the above, K and N respectively denote the number of channels of chromaticity elements, which is generally 2, and the number of synthetic surfaces or image pixels.

4.4.2 Production of Data for Training and Testing

In the previous work in [BF02], the experiment steps are:

1. Randomly select 4 synthetic surfaces from 1995 surfaces.

2. Randomly select one illuminant from training dataset in Figure 4.4 (d).

3. Generate a scene (image) with 4 selected surfaces in step 1 taken by camera in Figure 4.3, under the selected illuminant in step2.

4. Repeat step 1 to 3 1000 times, then 1000 generated scenes with 4 colours are acquired.

5. Start again from step 1 with 8, 16, 32, 64, 128, 256, 512, and 1024 synthetic surfaces, then each series of surfaces being associated with 1000 generated scenes with arbitrary selection of surfaces and illuminant from the reflectance database and the illuminant test database.

In our experiment, for getting a comparative result, we generated the same test database following the steps above. The difference is how we generate the training dataset:

1. Randomly select 1 illuminant A from training illuminant shown in Figure 4.4 (c).
2. Randomly select 1024 synthetic surfaces from 1995 surfaces.
3. Generate a scene (image) with 1024 selected surfaces in step 2 taken by camera in Figure 4.3, under the selected illuminant in step1.
4. Repeat step 1 to 3 1000 times, then 1000 generated scenes with 1024 colours are acquired under the selected illuminant A.
5. Start again from step 1, generate another 1000 scenes with 1024 colours under illuminant B, C, D ... until all 97 illuminants in training set has been selected. Then a training dataset with 97*1000 images is generated.

Since the data is created by using synthetic data, so the actual images are no need to generate, only the colour signals. So each image in training dataset is a 3*1024 two dimension matrix.

For every experiment, the algorithms will be assessed as to how they perform with 4, 8, 16, 32, 64, 128, 256, 512, and 1024 synthetic surfaces, each series of surfaces being associated with 1000 generated scenes with arbitrary selection of surfaces and illuminant from the reflectance database and the illuminant test database, respectively. The root mean square (*RMS*) error is calculated across 1000 results for every error measure, algorithm and number of scenes. The *RMS* error can be calculated with the equation below, with the error for the *i*th synthesised scene being denoted by E_i for a certain error measure:

$$\left(\frac{1}{N}\sum_{i}^{N} E_i^2\right)^{1/2} \tag{4.6}$$

In the above, $N = 1000$.

The *RMS* has been selected due to the fact that it affords an approximation of the standard deviation of algorithm estimates around the target, based on the premise that errors with mean zero have a normal distribution more or less (this roughly holds for the majority of algorithms) [BF02]).

4.4.3 Experiment with Matte Data

Solely matte surfaces are employed for training purposes because the proposed algorithm functions in Lambertian space based on the premise of lack of specular reflection. Two Grey World techniques are among the algorithms selected for detailed examination [Buc80, VDWG05]. More specifically, GW is the initial Grey World version, while DB-GW is a more refined version that depends on database training to obtain the average colour parameters [VDWG05]. The White Patch algorithm employed is the SCALE-BY-MAX algorithm [L+77].

The other two algorithms used in the experiment are several versions of Forsyth's Gamut mapping technique [For90, Fin95, FH98], the neural network technique (NEURAL-NET) [CFB02], and Colour by Correlation [FHH99]. Initially, Forsyth's technique was known as CRULE ("coefficient rule"), while ECRULE ("extended CRULE") denoted the illumination constraint set applied to CRULE. The Gamut mapping technique was enriched with two new concepts by the Colour in Perspective algorithm (CIP) developed by Finlayson [Fin96], namely, compatibility with the chromaticity space and additional constraint of the diagonal maps through their restriction to maps equivalent to anticipated illuminants. As the framework of the Gamut mapping technique, the creation of the solution set can be approached in a number of ways. In the present case, the initial maximum volume heuristic is indicated by MV; the averaging of the constraint set is indicated by AVE, with the application of a convex estimation to the illumination constraint if need be; the numerical integration of the constraint set due to its non-convex nature is indicated by ICA ("illumination constrained average").

For the Colour by Correlation technique (C-by-C-01), the original version of the algorithm proposed by Finlayson and colleagues [Fin96], the illuminant with maximum posterior probability is denoted by MAP, the chromaticity estimate determined by taking the average weighted by the posterior distribution is the MMSE estimate, and the MLM estimate is determined by identifying the maximum from the combination of the posterior distribution and a Gaussian mask. Additional details on the algorithm is provided in [BMF00].

Two minimal colour constancy techniques are incorporated in the results as well, namely, NOTHING and AVE-ILLUM. The former is based on the premise that the visual system does not require calibration for the actual illuminant, in other words, that the actual illuminant is the canonical or target illuminant. The latter is based on the premise that the illuminant is the average of the illuminants in the dataset that have been subjected to normalisation.

Figure (4.6) shows the plotting of the error for the chosen algorithms according to the number of matte surfaces. The algorithms typically perform better with the rise in the number of surfaces in the scenes, as this expands the amount of information available to them. Table (5.1) presents the error measure results derived from three techniques of error measures

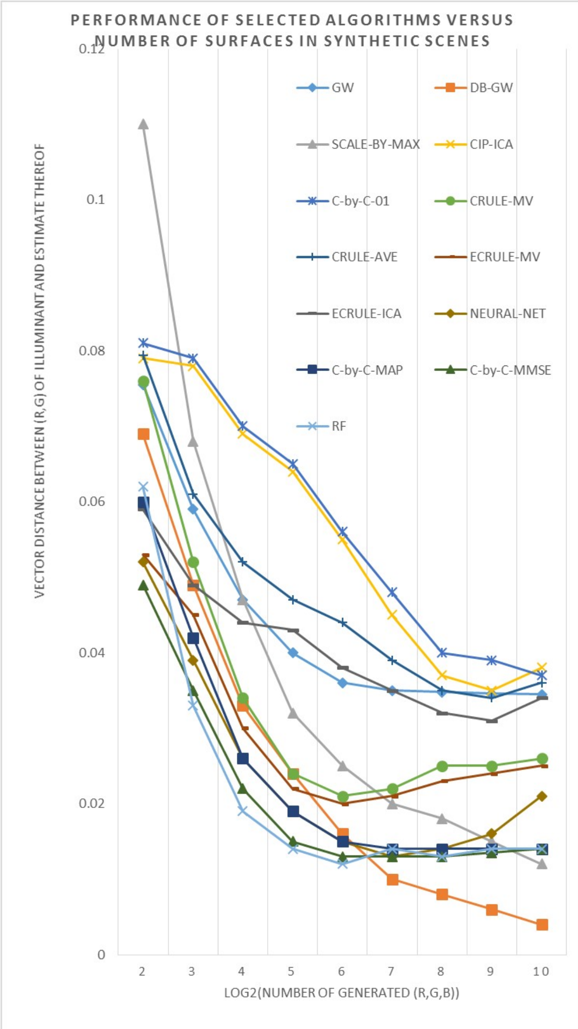


Fig. 4.6 Error in (r, g) chromaticity in relation to the number of surfaces for several algorithms with the highest performance

applied for a number of eight surfaces of a scene. This particular number has been chosen because the author of the dataset deems it is most similar in difficulty to the data from a broad range of images captured with a real camera [Bar99, BF02]. It is not usually possible to compare the absolute errors detected with synthesised and captured data, but the research still seeks to examine how relative performance is altered across the two conditions. The numbers need to be more or less identical to ensure maximum validity, which can be done by indicating eight synthetic surfaces that could be compared with the results from the image data.

The error of both NOTHING and AVE-ILLUM remains the same regardless of the number of surfaces, because neither technique is dependent on the scene. Consequently, neither of them is included in Figure 4.6 as the error is not within the range. By comparison to

Algorithms	angular error	median error	max error	rg error	RMS(rg)
NOTHING	16.45	13.67	37.65	0.114	0.113
AVE-ILLUM	11.79	8.72	37.65	0.086	0.089
GW	8.00	5.34	31.32	0.058	0.062
DB-GW	6.51	4.38	32.75	0.048	0.054
SCALE-BY-MAX	9.03	6.91	30.12	0.067	0.072
CIP-MV	26.27	18.43	67.83	0.200	0.240
CIP-AVE	18.12	12.75	40.87	0.130	0.141
CIP-ICA	10.51	8.32	45.34	0.077	0.081
NEURAL-NET	5.23	3.02	39.91	0.038	0.045
C-by-C-01	10.79	7.65	46.01	0.078	0.082
C-by-C-MAP	5.63	4.02	35.66	0.042	0.048
C-by-C-MLM	5.63	3.95	38.95	0.039	0.045
C-by-C-MMSSE	5.25	3.35	47.12	0.034	0.041
CRULE-MV	6.75	4.89	31.23	0.052	0.058
CRULE-AVE	8.39	6.31	59.31	0.061	0.066
ECRULE-MV	6.04	3.97	37.12	0.046	0.052
ECRULE-AVE	7.22	5.05	31.21	0.051	0.054
ECRULE-ICA	7.15	4.23	36.41	0.051	0.053
Random Forest SYN	4.31	2.51	32.67	0.033	0.039

Table 4.1 The performance of different algorithms in relation to the number of error measures. The indicated values represent the *RMS* across 1000 scenes produced in a synthetic manner. For each scene, there are eight surfaces chosen arbitrarily from the reflectance dataset. The scenes were captured under an illuminant chosen arbitrarily from the dataset of test illuminants. Note that the minimum error is not shown in the table because it is always 0. As for the max errors, even some of the latest algorithms can have relatively high worst errors as shown in the public academic website for colour constancy (www.colorconstancy.com).

NOTHING, AVE-ILLUM has greater efficiency as a minimal algorithm due to the distribution of the test illuminants all through the dataset and the location of the canonical illuminant near the edge of the set (its redness is above average).

A similar explanation helps to shed light on why the CIP-MV algorithm performs so poorly. This algorithm possesses bias, as demonstrated in [Bar99]. More to the point, the bluest illuminant corresponding to the observed chromaticities is selected by the maximum volume constraint in the $(R/B, G/B)$ chromaticity space. Therefore, the use of these parameters could have an adverse impact and a large number of surfaces is needed to increase the performance of CIP-MV above that of NOTHING and AVE-ILLUM. On the other hand, the performance of the CIP-ICA algorithm was consistently higher not only than the performance of the minimal algorithms, but also the performance of the other two CIP techniques.

Meanwhile, the performances of the C-by-C-01 algorithm and CIP0-ICA were similar. The calculation of the canonical gamut in the five distinct Gamut mapping techniques involves determination of the entire reflectance set under the canonical light source (i.e. Sylvania 50MR16Q), followed by calculation of the convex hull of that set. The experiment revealed that the ECRULE-ICA technique performed better all the time. In fact, compared to MV and AVE, the ICA technique has less bias.

In order to maximise the camera response function, every illuminant ground truth has been subjected to normalisation (all three colour channels possess maximal value with perfect pure white reflection). This explains why SCALE-BY-MAX performs so well. However, the error is not entirely null, due to the fact that the perfect reflectance in the dataset is excluded from the experiment. Furthermore, whereas the GW algorithm has a general estimator, the DB-GW has a specified estimator that is closely correlated with the dataset. The high performance of the DB-GW algorithm with a high number of surfaces can be explained in terms of the fact that an average grey value resulted from the production of synthetic data from the dataset with increase in the number of surfaces. Moreover, due to being optimal for RMS error, the performance of the C-by-C-MMSE algorithm exceeded that of the C-by-C-MAP algorithm. In the case of the neural network technique, the error occurred between C-by-C-MAP, C-by-C-MMSE and C-by-C-MLM.

By comparison to other learning-based algorithms relying on the distribution of input chromaticities, the proposed algorithm has produced the optimal chromaticity estimates. As indicated by the same trend in the earlier experiment, there was a fast decrease in the number of errors of Random Forest technique as the number of surfaces was increased. Alongside the results in the table, this provides evidence that, even when synthetic data are used, the Random Forest demonstrates efficiency and reliability as a classifier for the issue of processing extensive amounts of data.

4.4.4 Experiment with Real-World Images

Algorithm testing by using synthetic training data is the goal of the second experiment. The algorithm must perform satisfactorily to enhance the practical significance of this experiment. The training data remain unchanged from the earlier ones, but a larger number of illuminant classes are used compared to the preceding two datasets and synthetic data are used to train the SYN algorithm.

The results reveal that the algorithm trained with synthetic data performed much more poorly than anticipated. However, none of the three tests showed any flaw. The results could potentially be explained in terms of the fact that synthetic data and real-world images continued to have significant differences, despite the careful approach to gathering data.

Algorithms	Lab SFU	Grey-ball(Original)	Grey-ball(Linear)
Grey world[5]	9.8	7.9	13
Max-RGB[5]	9.1	6.8	12.7
Grey edge 2nd order[6]	5.2	6.1	10.6
pixel-based Gamut mapping[20]	3.7	7.1	11.8
Edge-based Gamut mapping[20]	3.9	6.8	12.8
Intersection-based Gamut mapping[20]	3.6	6.9	11.8
Random Forest 1D	4.2	5.6	8.2
Random Forest SYN	6.8	9.8	13.8

Table 4.2 The performance of the Random Forest algorithm with real-world images, as trained by synthetic data

Another possible explanation is that the illuminants were not entirely realistic in the real world, as some of them used a random linear combination technique. This could elucidate why no other learning-based problem has been approached with the use of a synthetic training dataset. For instance, only discrete estimation results can be generated by the Colour by Correlation, but just a few studies have indicated the use of data in this manner.

4.5 Conclusions

The present chapter proposed the use of a synthetic dataset, which helps algorithm training by supplying a random large number of images and illuminants, to address the problem of the classification framework for illumination estimation, namely, that the possible number of estimation results will be determined by the number of illuminant classes supplied by the datasets if classification of data is not comprehensive enough or if it contains an insufficient number of classes. Comparative analysis with several popular algorithms was carried out and the results of experiments conducted with synthetic data were presented. Moreover, unlike most other studies, we attempted to assess the realistic relevance of the dataset by testing the algorithm trained with synthetic data on real-world images. The results of the experiments indicated that the two datasets remained considerably different, therefore highlighting the need to improve the manner in which synthetic data are produced. Nevertheless, the experiment confirmed that the proposed Random Forest algorithm performed effectively and demonstrated robustness. In the following chapter, the second option is discussed, namely, the application of the clustering technique to enhance the specificity of data classification.

Chapter 5

A Clustering Classification Framework for Colour Constancy

In the previous chapter, we specified our approach to trying to improve the Random Forest classification framework with a large classified synthetic data-set. However, the result of the experiment displayed a bit of regression. A possible reason for this is the unknown difference between real images captured by camera and synthetic data. For this reason, we attempt a different approach, which is using cluster methods to classify the training data more specifically.

In this chapter, we introduce a novel clustering classification colour constancy framework (referred to as the 4C method). In the new 4C method, we assume that the training data contain the RGB values of the white point of the scene, referred to as $R_wG_wB_w$, which is obviously the colour of the illuminant under the imaging system. It is always possible to obtain these measurements by placing a neutral white patch in the scene when collecting training images, and some of the data-sets in this work [BMF00, BF97] use this approach to obtain the $R_wG_wB_w$ values of the white point (patch) of the scene. Based on the assumption that similar illuminants will have similar illumination colours (colours of the white patch of the scenes), we first use a clustering algorithm to group similar $R_wG_wB_w$'s of the training samples into the same cluster. We then treat the $R_wG_wB_w$'s in the same cluster as belonging to the same illumination source and each cluster as one class of illuminant. The colour constancy problem, i.e., that of estimating the unknown illuminant of an image, becomes that of identifying which illuminant class (cluster) the illuminant of the image belongs to. To achieve this, we use a classification algorithm to classify the image into the illuminant class (cluster). We present experimental results on publicly available testing data-sets and show that our new method is competitive with state-of-the-art methods.

5.1 Random Forest Regression

This approach is motivated by a desire to attempt to provide more prediction illuminant labels when the data-set can only offer a few. Providing a few discrete illuminant estimations means providing only a few discrete recover results. However, the Random Forest could achieve our goal by using regression method.

In our last work, for a test input image, after extracting the h_{RG} and h_{BY} features, we let the image features go through all the trees. When the image reached a leaf node of the m-th tree, we saved the illuminant distribution histogram of the leaf node $H_m(k)$, k = 1, 2,K, where K is the total number of possible illuminant classes. Suppose we have M trees in the Random Forest, we sum all M illuminant histograms together:

$$H(k) = \sum_{m=1}^{M} H_m(k) \tag{5.1}$$

The l-th illuminant is estimated as illuminant of the input image if $H(l) \geq H(k)$ for all k.

Thus, we will first try to adopt a weighting function by using the prediction table $H(k)$, instead of just labelling the input image with the highest scoring illuminant label (since samples in the same leaf node can be treated as king of relatively near neighbour). For each label k, there will be reference to a (R,G,B) ground truth value with the labeling illuminant k, so the final predicted (R_e,G_e,B_e) can be calculated by:

$$R_e = \frac{\sum_k^N R_c^k H(k)}{\sum_k^N H(k)}, G_e = \frac{\sum_k^N G_c^k H(k)}{\sum_k^N H(k)}, B_e = \frac{\sum_k^N B_c^k H(k)}{\sum_k^N H(k)} \tag{5.2}$$

N is the number of top ranking values used in this weighted average function. We simply performed a quick test with the same experiment condition as shown in Chapter 3. The result is shown in Table (5.1), with $N = 1$ being our original version.

Algorithms	Lab SFU	Grey-ball(Original)	Grey-ball(Linear)
Random forest N=1	4.2	5.6	8.2
Random forest N=2	5.0	6.7	9.6
Random forest N=5	6.8	7.5	10.1
Random forest N=10	8.0	7.8	10.3

Table 5.1 Random Forest performance test with weighting function of the ranking list in the same leaf node

However, the performance becomes increasingly worse as N is increasing. We will analyze the reason for this in the following section.

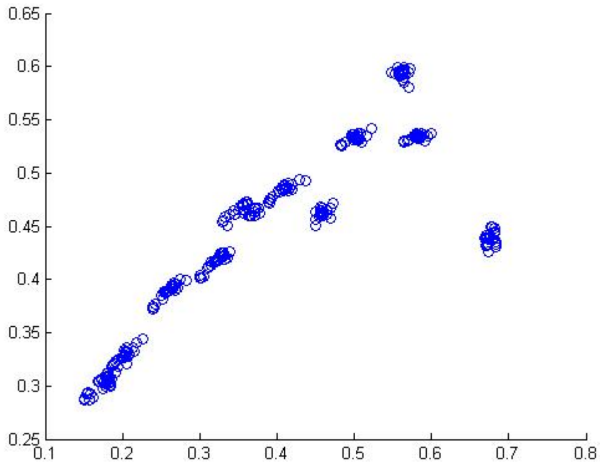


Fig. 5.1 The ground truth distribution from the SFU data-set. From the figure we can see the value points are gathered into 11 groups, the distance between each group is relatively far away. [BMFC02]

Let us see the ground truth distribution of the three public data-sets. The first one is the SFU data-set with 11 illuminants that have been carefully chosen and the images in the same class (under the same illuminant) have a very close ground truth value. Therefore, in this case, the data-set is very suitable for our classification framework. If the estimation produced an incorrect result, since the distance between clusters is relatively greater than a natural ground truth distribution, it could have a negative effect on the final performance.

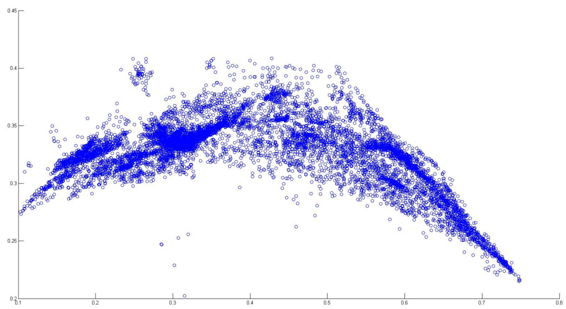


Fig. 5.2 The ground truth distribution from the grey ball data-set. The data are abundant and almost cover the whole colour space. However, the data are classified by scene instead of ground truth. [CF03]

In the second "grey ball" data-set, since it is a large data-set, it should be suitable for regression, the distribution of ground truth almost covering the whole colour space. Nevertheless, why is the regression result still worse than the original result? The answer is that, from the data-set, the image was not labelled with illuminants, as we mentioned in Chapter 3, when we used this data-set for training; furthermore, the class number 15 is based on the scene, not the ground truth, and even in the same scene, the illuminant can still vary. In Figure (5.3), we selected several images from the same scene in the grey ball data-set and we can see the grey balls in the bottom right corner are not strictly the same in these images. Nonetheless, we treated these images as one class for training in the previous implementation.

Fig. 5.3 The images selected from the grey ball data-set [CF03]. All the images were taken from the folder Granville Island Market2, which we used as one class in our previous implementation. Form the images we can see that, as the grey ball in the bottom right corner was used to compute the ground truth, it still looks varied in the same scene.

There is another benchmark data-set we have not used for training, namely, the Colour Checker data-set [GRB+08]. This data-set consists of 568 images of scenes under different lighting conditions, with a large variety of content: people, landscapes, buildings, objects, indoors, and outdoors. Each scene's illuminant is obtained using a Gretag MacBeth Colour Checker Chart that was placed in the scenes (through analysis of the achromatic patches). The main advantage of this database is that it contains more realistic images and of higher quality. Figure (5.3) shows examples of images in this data-set, while in Figure (5.4), we can see the ground truth distribution of this data-set, also very discretely.

Therefore, without a pre-process of the training data, our algorithm will have limited performance. In this case, we will adopt the clustering method to generate enough classes for training.

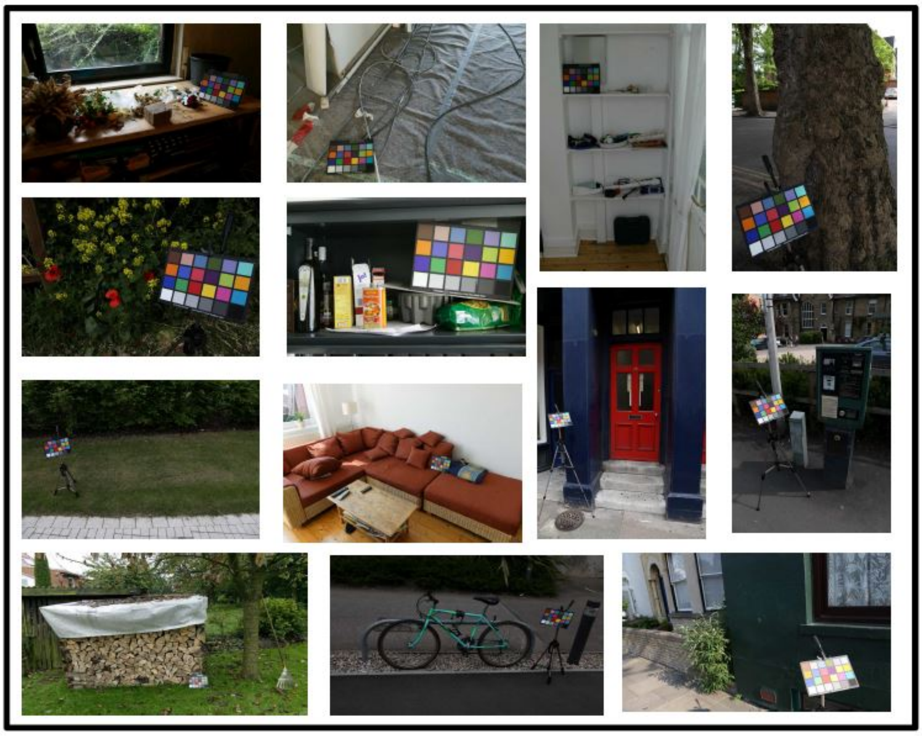


Fig. 5.4 Examples of images the Colour Checker data-set [GRB+08]

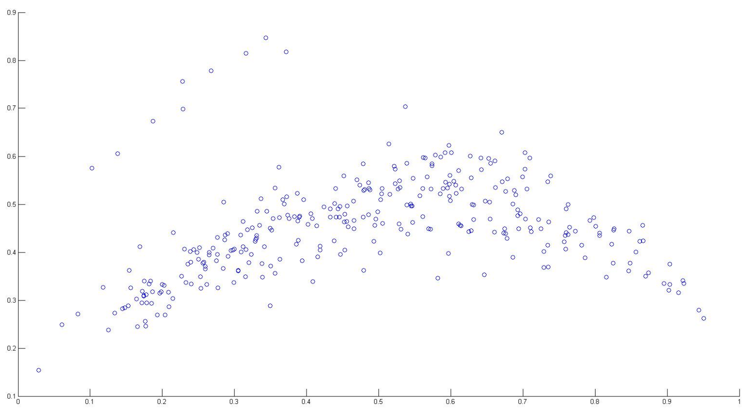


Fig. 5.5 The ground truth distribution from the Colour Checker data-set [GRB+08]. From the figure we can see that, without a good training method, our Random Forest algorithm cannot yield a good performance.

5.2 The Clustering Classification Colour Constancy Method

The clustering classification colour constancy (4*C*) method is illustrated in Fig (5.6). In the left column, we illustrate the clustering stage. From the training images, we first extract the $R_wG_wB_w$'s of the white patches that have been placed on the scenes during training data collection. We then use a clustering algorithm to group similar $R_wG_wB_w$'s into the same cluster. Each cluster is then treated as one class of illuminant source. In the middle column, we illustrate the classification stage of the 4*C* method. Firstly, we extract the features of the training images, which are then applied to a classifier that in turn will classify the input image into one of the illuminant source classes. In the right column, we illustrate how a trained classifier can be used to estimate the illuminant class of the testing image. For an image with unknown illuminant source, we first extract the image feature, which is then fed to the classifier that will output an estimate of the class of the illuminant source.

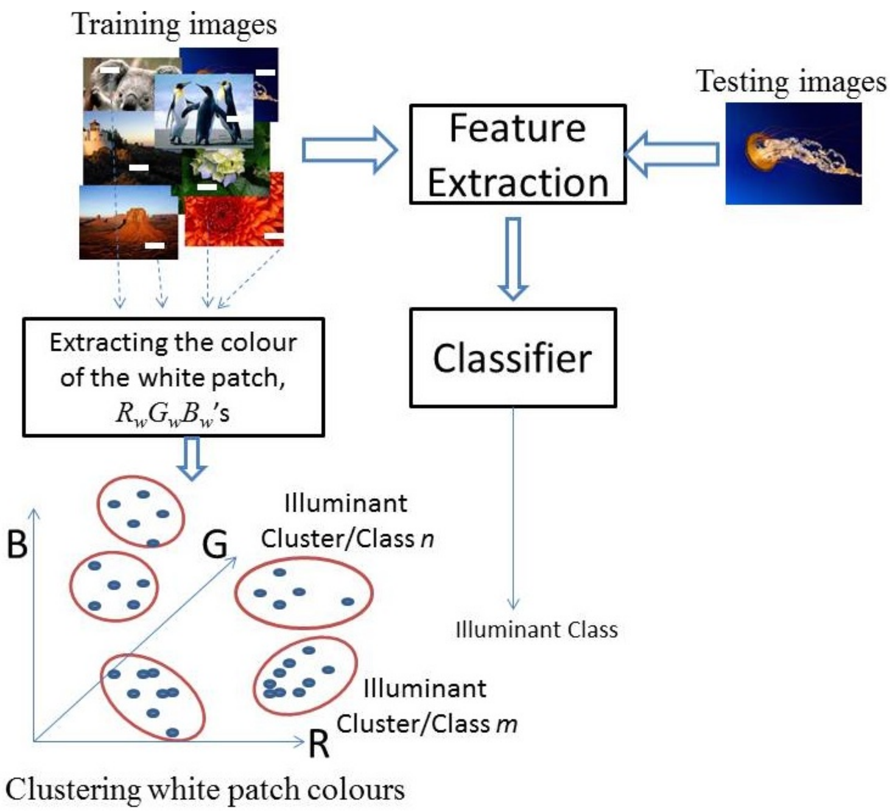


Fig. 5.6 The 4*C* colour constancy method.

The two challenging issues are how to cluster the images into illumination groups appropriately and how to build the classifier for classifying the input image. For clustering, there are many standard techniques available. In our work, we use the popular *K*-means and

FSCL algorithm. The challenge is to determine the value of K. Similarly, there are many classification techniques we can use to build the classifier, and we propose the use of the Random Forest-based classifier and Support Vector Machine method to tackle the problem. The challenge here is what image features to use to build the classifier.

5.2.1 Illuminant Colour Clustering

We firstly use K-means clustering to cluster the colours $R_wG_wB_w$'s extracted from the white patches of the training image into K clusters. Although simple to implement, the basic K-means algorithm can be unreliable as the final clusters are highly dependent on the initial centres. To alleviate this problem, we have implemented a random initialization step to choose the initial cluster centres. To initialize one cluster centre, we randomly chose 10% of the training samples and take their average to form the initial values of the cluster centre. We found this method worked quite robustly and was pretty stable. Another issue is that we do not know what K should be. Fortunately, in our case, this is not critical and we will show that our framework can work stably for a wide range of K (from 5 to 50). For each cluster (class), we average the $R_wG_wB_w$'s of the images in the cluster and use it as the colour of the white point colour of that illuminant class. The clustering result is shown in Figure (5.7).

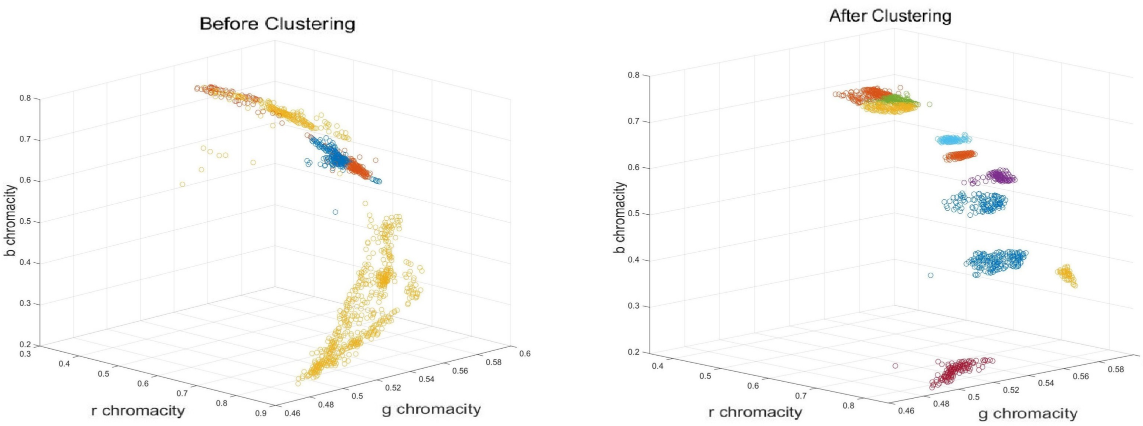


Fig. 5.7 The sample classes chosen in the same *rg* chromaticity space scale, before and after clustering. The left side is 3 original different classes of totally 15 classes from Greyball dataset, from the image we can see the samples in the original class, especially samples in the yellow classes, are very sparse in the space. The right side is the result after clustering by our method, 10 were selected from 50 classes (K value). It is clearly shown that the data are more closer than before and more suitable for training.

Secondly, we will test another clustering method, which is an extension of K-means clustering, frequency sensitive competitive learning (*FSCL*) [AKCM90], to compare the performance of different clustering methods. The classical K-means algorithm has the "dead units" problem. That is, if a centre is initialised far away from the input data set in comparison with other centres, it may never win the competition, so it may never update, becoming a dead unit. To solve the "dead units" problem, the $FSCL$ algorithm adopts a punish strategy: each centre counts the number of times when it has won the competition and reduces its learning rate consequently. If a centre has won too often, "it feels guilty" and it pulls itself out of the competition. The $FSCL$ algorithm will be designed in the following way:

1. Initialise the illuminant classes, $C_i(0), i = 1, 2, \ldots, I$, with the prototypes $w(i)$ (in our case it is a three-dimensional vector in (R, G, B) space), set the counters associated with each class to 1, i.e., $n_i(0) = 1$;

2.Present the training sample, $X(t)$, where t is the sequence index, and calculate the distance between $X(t)$ and the classes, $D_i(t) = D(X(t), C_i(t))$, and modify the distance according to $D'_i(t) = n_i(t)D_i(t)$;

3.Find j, such that $D'_j(t) \leq D'_i(t)$ for all i, update the class and counter:

for each cluster C_i, where $i = 1, 2, \ldots, I$, update the prototype $w(i)$ to be the centroid of all samples currentl in C_i,

$N_j(t+1) = N_j(t) + 1$

4. Repeat step 2 until prototype $w(i)$ does not change significantly or cluster membership no longer changes.

In addition, we will still test for the same range of K as above.

5.2.2 Illuminant Classification by Support Vector Machine

The implementation of the Random Forest-based classifier has already been shown in Chapter 2. Therefore, in this chapter, for comparison purposes, we will adopt another one of the best classifiers, namely, support vector machine (SVM), which will be used to test our framework performance. Vapnik's [SS04] Support Vector Machine theory has been applied successfully to a wide variety of classification problems [CGM02, HL02, LL03]. SVMs have also been used to address classification/regression problems including financial market forecasts, travel time prediction, power consumption estimation, and highway traffic flow prediction [YCK02, WHL04, DZJ02]. Futhermore, it is no surprise that researchers use the SVM method to tackle the illuminant classification problem; a previous study [FX04] applied support vector machine regression to estimate a continuous-valued function that encoded the fundamental interrelation between a given input and its corresponding output in the training data. This function could then be used to predict outputs for given inputs that were not included in the training set. This work is similar to a neural network. However, a neural network's solution is based on empirical risk minimisation. In contrast, SVR introduces structural risk minimisation into the regression, as one of the character of SVM, and thereby achieves global optimisation [CFB02].

Colour Feature to Represent Illuminant Estimation

Most state-of-the-art methods prefer using binarised 2D chromaticity space histograms (r, g) to represent the input image data. The chromaticity histograms have the potential advantage that they discard intensity shading that varies with the surface geometry and viewing direction, which is most likely unrelated to the illumination's spectral properties. However, as we have proved in our framework, our new RG,BY colour signals could perform better with the illuminant intensity. For this reason, we will still use these features in the implementation of our SVM classification method. The colour feature extraction is the same as shown in Chapter 3. Finally, the two binary vectors we cascaded are still one 512 dimensional feature vector F_{RGBY}.

The Support Vector Machine Framework Contraction

Fundamentally, in the simplest scenario in a 2D space, the purpose of SVM is to identify a linear function to categorise points in two distinct classes according to their features. In a space of greater dimensionality, data points are considered by SVM as p-dimensional vector and divided with a $(p-1)$-dimensional hyperplane. The SVM training algorithm can be identified as a non-probabilistic binary linear classifier because, for a series of training examples categorised into one of two groups, it creates a model for allocating new examples into one of the two groups. Examples are represented by an SVM model as points in space and their mapping is done in such a way as to ensure that a maximum gap exists between the divided groups. Subsequently, the mapping of new examples is done in the same space and predictions are made regarding the group they fall into according to the side of the gap they are found. On the downside, the linear hyperplane is sometimes impossible to locate. Furthermore, SVM is capable of undertaking not only linear classification, but also non-linear classification based on the so-named "kernel" function, with data point inputs being automatically mapped in features spaces of great dimensionality.

In the case of nonlinear SVM, kernel functions characterise the feature space. Knowledge exists about a number of these functions, but a theoretical framework for their selection is yet to be developed. The Radial Basis Function (RBF) kernel is chosen in this study because it performs well with regard to algorithm convergence and robustness. It is expressed as follows:

$$K(X,X^{'}) = exp(-\frac{\left\|X-X^{'}\right\|^{2}}{2\sigma^{2}}) \tag{5.3}$$

The empirical risk and the norm of the weights are reduced at the same time to derive the nonlinear classification function in *SVM*, which in this case is the C-support Vector Classification. Referred to as structural risk minimisation, the process is expressed as:

$$sgn(\omega^{T}\phi(x)+b) = sgn(\sum_{i}^{l} y^{i}a^{i}K(X,X^{'})+b) \tag{5.4}$$

In the above, the training vectors x_i are mapped into a space of greater dimensionality by $\phi(x_i)$, while the label of the indicate vector is denoted by y^i . The experiment parameters have been established in keeping with earlier studies due to the existence of a number of prior assessments of the parameter setting in relation to illumination estimation [FX04, AGA07, CB00], as well as to make it easier to undertake comparison under the same condition. *SVM* implementation is underpinned by a code derived from a well-known *SVM* library that is compatible with a range of programming languages [CL11].

5.3 Algorithm Evaluation

We evaluate our new $4C$ method on three public benchmark data-sets: Lab SFU [BMFC02] consisting of 321 images, Grey Ball [BMF00] contains 11346 images and Colour checker [BF97] consisting of 568 real-world images. Following a common practice in the literature, we calculate the angular error according to equation (5.5) from [BCF02] as the performance indicator:

$$Error_{angular} = cos^{-1}((T\cdot E)|T|^{-1}|E|^{-1}) \tag{5.5}$$

where T is the ground truth illuminant value and E is the estimated value. The error is calculated by the degree distances between the two colour vectors of the same as before, all the results with Random Forest are averaged over 50 different random trials (each with different 80% training and 20 % testing samples). The tree number in the random forest is 100 (we have varied the number of the trees and results are saturated at 100). Furthermore the *SVM* experiments are also the average values of 50 trials (each trail applies a *FSCL* clustering then randomly selects 20% for test in each cluster.)

5.3.1 Comparison of Colour Constancy Algorithms

Table 5.2 Comparison of Colour Constancy Algorithms with the 4*C* Framework

Algorithms	Data-sets				
	Lab SFU	Greyball (Original)	Greyball Linear	Colour checker Original	Colour checker Linear
Grey world	9.8	7.9	13	9.8	6.4
Max-RGB	9.1	6.8	12.7	8	7.6
Grey edge 2nd order	5.2	6.1	10.6	7.1	5.1
Pixel-based Gamut mapping	3.7	7.1	11.8	6.9	4.2
Edge-based Gamut mapping	3.9	6.8	12.8	7.7	6.5
Intersection-based Gamut mapping	3.6	6.9	11.8	6.9	4.2
Exemplar-based	*	4.4	8.0	5.2	2.9
Natural Images	*	5.2	9.9	6.1	4.2
SVR	*	*	13.1	*	8.1
Bayesian	*	*	*	6.7	4.8
4*C* RF K=5	4.48	4.98	9.11	4.9	4.3
4*C* RF K=10	4.40	4.69	8.62	4.88	4.31
4*C* RF K=15	4.48	4.68	8.31	4.75	4.23
4*C* RF K=50	4.57	4.64	8.01	4.78	4.23
4*C* RF K=100	4.60	4.68	8.01	4.73	4.29
4*C* RF FSCL K=5	4.44	4.86	8.97	4.85	4.27
4*C* RF FSCL K=10	4.40	4.54	8.54	4.78	4.31
4*C* RF FSCL K=15	4.34	4.45	8.20	4.80	4.19
4*C* RF FSCL K=50	4.40	4.47	8.04	4.81	4.24
4*C* RF FSCL K=100	4.44	4.50	8.03	4.80	4.23
4*C* SVM FSCL K=50	3.73	4.17	7.6	4.42	4.21
NetColorChecker	*	3.90	*	*	3.10
Deep learning	*	*	6.60	*	2.16

In this experiment, we finally include all three data sets we mentioned in the literature, as the adoption of a clustering method. The experiment is the evaluation of different algorithms for the three data sets with five ground truths (The grey ball data-set's ground truth values have been modified under a linear colour space since the original data set is captured under a non-linear colour space [BMFC02], as well the colour checker data set [SF10]). The forest size here is 100 and the percentage for training is 80%. The results are averaged over 50 different trials in the Lab SFU data-set and 30 in the grey ball data set. The compared algorithms included are classic algorithms like the grey world [Buc80], the max-RGB algorithms [L^{+}77], shades of grey [FGJ98], grey edge [GGVdW07], and the gamut mapping algorithms with different gamut weighting methods[For90, GGVDW10, CK02]. We also included some learning-based methods, such as: Using Natural-Image Statistics [GG11], Exemplar-based [JD12], support vector regression [FX04], Bayesian method [GRB$^+$08], and two deep learning based colour constancy methods [LGHL15, OK16]. Deep Learning methods are expected to have the best performances.

These two versions of deep learning methods have tackled colour constancy problems in different ways: The NetColorChecker [LGHL15] approach formulate a regression approach to estimate the illuminant by using Convolutional Neural Networks (CNN) and training with the ImageNet dataset. In the previous work, algorithms of colour constancy mostly constrain to image features like chromaticity information, illuminant intensity information, and other low level features, and the ground truth are hand labelled. In this method, the first training step is to compose a multi-scale image feature include pixels, edges, object parts and object models. After that, the ImageNet dataset, a much larger dataset compared to the current colour constancy dataset, is used for training. The ground truth is obtained by using current colour constancy algorithm gray-shades. The last step is retraining the parameters obtained in the previous step by using current benchmark datasets (like Greyball and ColourChecker dataset).

The Deep Learning method of [OK16] bears some similarity to our 4C method. Firstly they also treated colour constancy as an illuminant classification problem, using K-means method to cluster the training images. They chose $K = 25$ and the employed a classic CNN architecture [KSH12]. The CNN consists of five feature extraction layers of convolution and the maxpooling, followed by three fully-connected layers. The sparse connections in the layers 3-5 for multi-GPU implementation are replaced with dense connections for the single GPU system, and the final output of the CNN is a vector of length K, which K refers to the illuminant cluster number (25 in this case). Finally, the chromaticity of the estimate illumination was determined by the average value of all images ground truths in cluster K, same as ours. It is worth pointing out that we developed our 4C independently.

All the results for these algorithms are from the website www.colorconstancy.com, an academic evaluation platform for colour constancy. Results of Gamut mapping algorithms requires training data. The following results were obtained using 31 images (all images recorded under the "syl-50MR16Q"-illuminant) to construct the canonical gamut. This canonical gamut is then used to estimate the illuminant for all images including the training images. This is done to be able to make a comparison with other methods. Interestingly, most of the learning-based methods were not tested on the Lab SFU data set, the missing result was labelled as a "*". One possible reason could be that the images in this data-set are insufficiently natural for the learning model as the images were taken under lab condition.

Moreover, for some learning-based methods, the performance on nonlinear ground truth of the natural images was not tested, as the colour feature r,g is more suitable for a linearized ground truth.

We have evaluated our 4C method with different Ks, and for each K the result is the mean value of 50 trials. The result is very impressive. From Table (5.1), we can see that for the Grey Ball data set, our new technique outperformed all other methods except for the latest deep learning neural network methods. For the Colour Checker data-set, our method also performed better than others except for the deep learning methods and one of the latest and much more complex technique [JD14] with the linear ground truth. For the Lab SFU data-set, our new Random Forest method achieved a very good performance, but not as good as the best techniques in the literature. We notice that Lab SFU data are all man-made objects under artificial illumination sources whilst the other two data-sets are natural scenes under natural illumination sources. Therefore, it is possible that the 4C method works best for scenes containing rich colours like the natural scenes in the Grey Ball and Colour Checker data-sets. Another reason is related to the ground truth distribution, the distance between clusters is very great and samples in each cluster are very close, which means that, if the prediction is wrong, the error will be relatively high than a more evenly distributed situation. It is also seen that the number of clusters does not affect greatly the performance of our method (in fact, in the latest deep learning method, they used the same approach to cluster the training images, which shows that the K determine problem is still unsolved). In addition, our SVM with $FSCL$ clustering beat the previous SVR algorithm, with regression model, demonstrating that our new 4C method is a stable and robust technique.

5.4 Conclusive Remarks

In this chapter, we have developed a new colour constancy technique called the clustering classification colour constancy or 4C method. We first used a clustering technique to cluster the training images into groups based on the colours of the white points (patches) of the training images. We then treated each cluster as an illumination class, and developed a classification method to classify the images into these illuminant classes. We have tested our new method on publicly available benchmark datasets and showed that our method is was competitive to with state-of-the-art methods.

Chapter 6

Video Illuminant Recognition and Tonal Stabilization

The colour constancy problem does not exist only in still image, but also in videos. These fluctuations are typically caused by the camera automatically adjusting of its tonal settings while shooting. In this chapter, we presents a method for video illuminant recognition, and furthermore, try to reduce undesirable tonal fluctuations in videos, which are easily noticeable when the sequence is viewed.

We address the colour constancy problem from still image to video. The video tonal stabilization problem is still a unsolved problem, and current algorithms [FL11, BTS$^+$15, BBC$^+$12] are only focusing on keeping the tonal stable during the video playing, not really trying to recover the incorrect illuminant. From this point, we are trying to combine these two problems together, as firstly keeping the tonal stable and secondly recovering the frame colour to canonical illuminant, and propose an algorithm for improving the performance of the video. Our approach operates on a continuous video shot by first cutting the video into one or more scenes based on features presenting the characteristics of the illuminants. We then correct a sequence of frames in the same scene by using the Random Forest illuminant estimation framework: for every scene, we extract the features of each frame in the same scene and return an illuminant estimation result, then we will use the most-voted result as the estimation of this scene and correct the tonal information. Finally, a smooth function between different scenes will be applied to adjust the flick and flash occurring at the boundary of the neighbour scenes. Finally, we evaluate our method on some test videos to determine its performance.

6.1 Introduction

In recent times, there has been an unprecedented growth in captured video content, as devices for video capture have become more and more affordable and video sharing websites have become increasingly popular. For instance, on the YouTube website, video content with a combined duration of around 24 hours is uploaded every minute [CDL08]. To a great extent, this content is produced unprofessionally by amateur videographers with the help of basic video cameras.

A feature that sets most amateur video footage apart from professionally-made videos is that the movement of the camera is not controlled adequately and the exposure and colour balance vary considerably, as amateur videographers lack the skills or gear needed to avoid these issues. As shown in Figure (6.1), the camera automatic exposure and white balance control are responsible for the variations in tone: modifications in lighting and frame compositions cause ongoing slight adjustments to the tonal settings. However, it is not advisable to turn off auto-exposure because it makes avoidance of over- and under-exposure challenging as the capture that the camera is capable of with a fixed exposure setting is not usually as wide as the dynamic range of the scene. By contrast, it is possible to turn off automatic white balance, but this option is not available with all cameras. Nonetheless, tonal variations due to illuminant modification still occur even if the automatic white balance is not on.

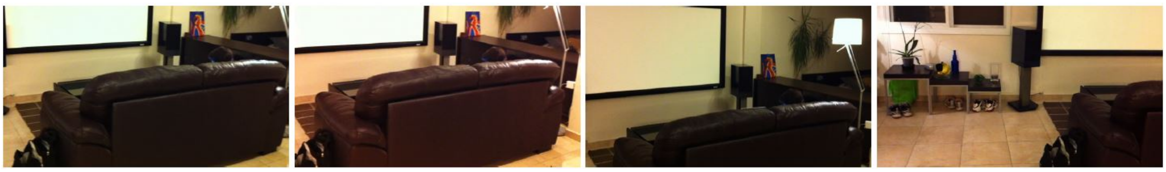


Fig. 6.1 Multiple frames from a video sequence taken with an iPhone; marked colour variation is due to the auto white balance of the camera.[FL11]

A range of approaches have been proposed to eliminate the camera shaking effect and thus stabilise the video motion. Two such approaches have been recently put forth in [LGJA09] and [MOT$^+$06]; meanwhile, some approaches are geared towards eliminating tonal variation or improving tonal stabilisation [FL11, BTS$^+$15, BBC$^+$12]. The variation is addressed by most approaches based on identification of a mapping to the anchor frame. However, given the complexity of the issue, this solution will be unsuccessful if the colour of the anchor frame is not correct.

A new solution for dealing with video tonal stabilisation is put forth in the present chapter. Based on a continuous video shot, this solution involves using the characteristics of the illuminants to separate the video into one or multiple scenes; subsequently, the Random Forest framework for illumination estimation is applied to ensure the correctness of the frame sequence in the same scene. An estimate for illumination will be generated based on the extraction of the features of every frame in the same scene and the estimate with the highest number of votes will be considered the scene estimation and used to rectify the tonal information. Furthermore, the flick and flash effect occurring between adjacent scenes will be adjusted through application of a smooth function between the scenes. Following alignment of the sequence, several clips taken with different test videos will be used to demonstrate that the proposed solution is viable. Furthermore, the performance of the approach will be demonstrated in relation to illuminant detection of video frames and retrieval of frame tonal information will be undertaken under straightforward illuminant fluctuating conditions. The premise applied is that there is a gradual change in scene lighting conditions and that they do not undergo spatial fluctuation over the frame.

6.2 Related Work

The fields of computational photography, image and video processing, and computer graphics have addressed various related problems.

6.2.1 Camera Response Recovery

The modelling of the digital video capture pipeline involves the following steps [Poy12]: conversion of the analogue linear *RGB* values reaching the camera sensor into digital values, division of the values through luma chroma separation, processing for brightness and colour adjustment, and encoding to the desired digital video format. Non-linear operations may be included in the procedures of conversion from analogue to digital and the following processing. A tonal effect may arise in this pipeline due to the likely variation in the response function of the camera between different cameras with different operational settings. Sequence stabilisation could be achieved through response function inversion if the camera response at every frame were known.

Parametric [MP94, MN99, TRK01], semi-parametric [CM$^+$05], and non-parametric techniques [DM08] are among the techniques developed to model and retrieve the camera response. In general, these techniques are compatible with still, geometrically registered images with only exposure variation, and therefore a series of suitably large precise equivalences between all frame pairs are needed for the techniques to be applicable to video and the computation of such equivalences might prove challenging. Furthermore, a result demonstrating numerical stability might not be obtained due to the insufficiently large exposure change between sequential frames, even if there is availability of the necessary equivalences. Additionally, these techniques would have to be extended to manage more general camera parameter alterations.

6.2.2 Colour Transfer

The transfer of colour from the anchor to the rest of the frames may seem enough to accomplish tonal alignment. In fact, over time, many different colour transfer techniques have been introduced [AP10]. One study [RAGS01] prompted various efforts to achieve compatibility between different universal colour statistics of two images, like mean and variance in certain colour spaces. However, camera and object motion cause considerable variation in frame statistics, which means that these techniques are inapplicable for purposes of tonal stabilisation. The relative rapidity with which such modifications can take place signifies that further modifications would be triggered by endeavours to attain compatibility between the universal statistics. Meanwhile, local techniques have been proposed to match image regions at local level and introduce an equivalent offset [TJT05, KMHO09]. Despite the great potential of such techniques, it is still enormously difficult to match regions in a reliable way within the presence of camera and scene motion.

A scribble interface has been employed in several recent studies [LLW04, LAA08] with notable recolouring results. In another study, on the basis of a robust non-linear parametric transfer model, the scribble interface was applied in a system created particularly for inter-image colour transfer managed by users [AP10]. However, the present thesis aims to demonstrate that a more straightforward transfer model is appropriate for accomplishing tonal stabilisation, without having to rely on user interaction.

6.2.3 Colour Constancy and White Balance

Over time, a range of white balance algorithms have been put forward and more detailed information about them can be found in [AAKA06] and [Hor06]. The principle underlying the majority of such algorithms is matching pixel values with particular scene features (e.g. average reflectance or colour of the illuminant). Tonal variations are generated when such an approach is used on a frame-by-frame basis, due to the deviation between frames displayed by statistical estimates of the scene features in question. Furthermore, it is impossible to reverse white-balance corrections because they are orthogonal to certain universal colour manipulations (e.g. saturation). For some tasks, different techniques need to be combined, but this is not the case here, as the tonal issue in videos is the sole concern.

6.2.4 Current framework

The approach presently applied to achieve video tonal stabilisation [FL11] is based on the tonal alignment algorithm, which is designed to make it seem that, for a particular anchor frame, identical tonal settings were used to shoot a sequence of neighbouring frames by adjusting the frames accordingly. Alignment is usually established among the frames mediating a pair of sequential anchors and the fusion of two resulting sequences in keeping with how far each frame is from each anchor gives the end-result. This fusion ensures that the tonal settings at the anchors that exhibit differences are smoothly interpolated.

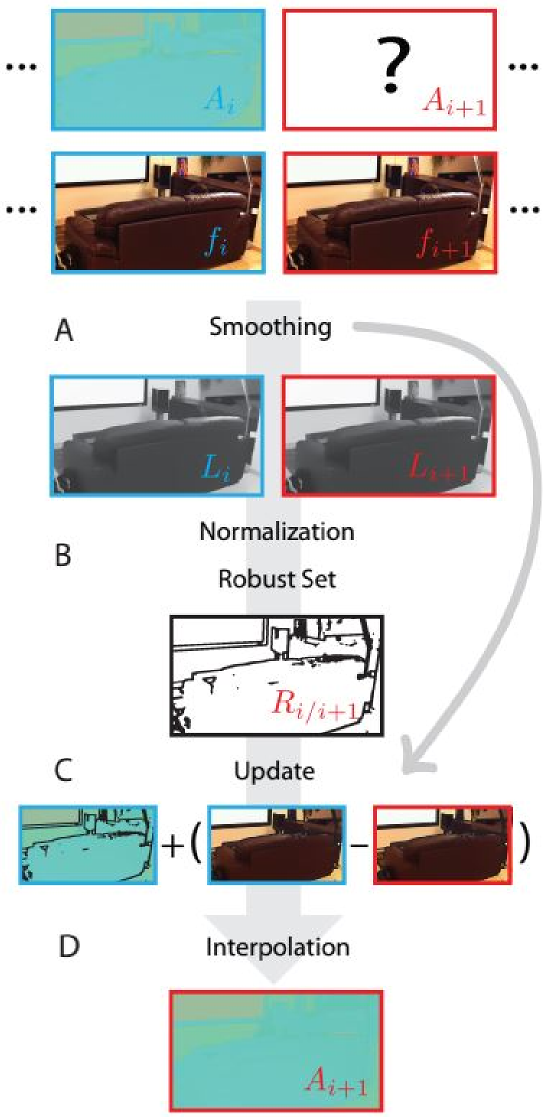


Fig. 6.2 The flow of the algorithm of adjustment map update.[FL11]

An adjustment map A_i is created for every frame f_i, delineating how the adjustment of the colour channels for every pixel should occur to achieve the target aligned value. Following creation of the adjustment maps, they can be applied to the frame to obtain the aligned sequence. The anchor map (i.e. identity mapping) is known, and therefore the aim is to generate a technique that will enable the maps to be effectively and reliably propagated along the sequence of frames. At a more formal level, an approach is sought by the algorithm to determine A_{i+1} based on knowledge of f_i, A_i and f_{i+1}. Figure (6.2) presents the flow of this algorithm.

Among the first endeavours to solve this issue, the technique could yield reliable results in achieving tonal stabilisation in the majority of cases. Nevertheless, the right anchors must be found for the technique to perform effectively, and what is more, anchor identification does not guarantee that the colour of the anchor is appropriate. Given these considerations, the implementation of the proposed still image colour constancy framework will be approached by first developing a technique for video illuminant recognition and subsequently attempting to retrieve the frame colour in several uncomplicated scenarios.

6.3 Algorithm Framework

The core of our method is that, compared with the tonal alignment algorithm, instead of adjusting a sequence of adjacent frames to the anchor frame, we assume our still image illuminant estimation framework could give a correct result, then for each frame, the anchor frame is the same scene taken under the canonical illuminant. Therefore, the key problem of the technique is how to extend the illuminant estimation framework from still images to consistent video frames.

The pipeline is shown in Figure (6.3):

1. For an input video, the video will be initially cut into different scenes $S(i), i = 1,2,3,\cdots$ and buffer area $B(i), i = 1,2,3,\cdots$ between two scenes.

2. For each scene $S(i)$, we estimate the illuminant class $C(j)$ for each frame $f(k)$ in the same scene $S(i)$, and the $C(K)$ appearing most frequently will be selected as the estimated illuminant for the scene $S(i)$. Furthermore, all the video frames in the same scene will be colour corrected under the canonical illuminant from illuminant $C(K)$.

3. For each buffer area $B(i), i = 1,2,3,\cdots$ between two scenes, given a single frame $f(k)$, we calculate the distances between the current frame to the last frame of the last scene $DL(k)$ and the first frame of next scene $DN(k)$. The last frame of the last scene has the correction parameter (R_L, G_L, B_L), and the first frame of next scene has the correction

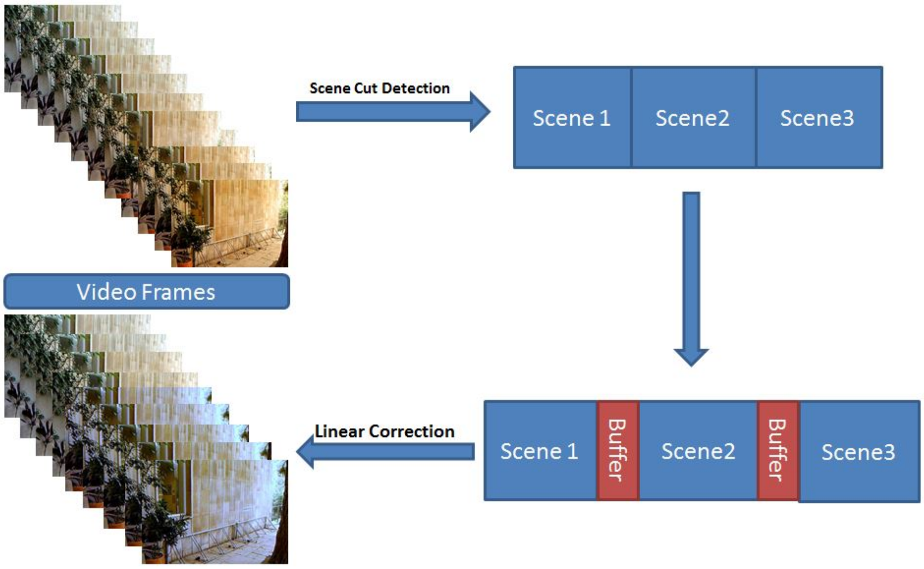


Fig. 6.3 The video tonal stabilization framework: firstly, the input frames are cut into different scenes by scene detection. After that, a buffer area is set between different scenes and finally a linear correction model is applied to smooth the boundary of the scenes.

parameter (R_N, G_N, B_N). Therefore, the correction parameter for the current frame $f(k)$ will be:

$$
\begin{aligned}
R_k &= \frac{DN(k)}{DL(k)+DN(k)} R_L + \frac{DL(k)}{DL(k)+DN(k)} R_N \\
G_k &= \frac{DN(k)}{DL(k)+DN(k)} G_L + \frac{DL(k)}{DL(k)+DN(k)} G_N \\
B_k &= \frac{DN(k)}{DL(k)+DN(k)} B_L + \frac{DL(k)}{DL(k)+DN(k)} B_N
\end{aligned}
\tag{6.1}
$$

There are three key challenges: how to properly divide the video into different scenes based on illuminant, how to adequately correct the image colour to canonical light, and how to smooth frames in the buffer areas to avoid flickers during scene change.

Our contribution is that, compared with the current work only focusing on alignments between frames, we combine the tonal alignment and colour correction tasks together. In other words, it means the anchor frame in our case is the frame taken under the canonical light, which we think it is the correct colour for objects shown in the frame.

6.3.1 Video Illuminant Recognition and Scene Detection

A film is divided into basic temporal units known as shots through scene/shot transition detection [CNP06]. A shot is composed of several interconnected and sequential images that are captured contiguously by one camera as a continuous action in temporal and spatial terms. Video post-production relies significantly on this procedure, which is also important for automated indexing and video recovery according to content or summarisation applications. Vast video archives are made accessible by the latter; for example, to generate a visual overview of the entire film, a representative picture may be selected by an application from every scene and the processing of such indexes enables a search engine to produce results for queries about all the films containing a particular scene.

For computers, cut detection is not such a straightforward procedure as it is for people, because information regarding the time of capture and type of capturing camera is not available for every frame of a video. At the moment, it seems that a cut detection algorithm must have extensive artificial intelligence to be capable of accurately identifying all the cuts.

There are two stages in the process of cut detection, namely, scoring, whereby a particular score is allocated to every pair of sequential digital video frames indicating how similar or different the two frames are, and decision, which involves assessment of the allocated scores, with a high score being indicative of cut detection.

Errors are likely to occur with this approach for two reasons. One reason is that a decision hit will be triggered by a score even if it is only slightly greater than the threshold value. Therefore, to increase the average difference between the "cut" and "no cut" score as much as possible, a broad distribution of values is essential in the first stage. The other reason is that a careful approach must be adopted in the selection of the threshold, statistical techniques being particularly helpful in this regard.

Scene Difference Scores

Differences in visual content can be accessed through a wide range of scores, the most frequently used ones including:

Sum of absolute differences (SAD). The most straightforward of existing algorithms, this technique involves pixel-by-pixel comparison of two sequential frames, with the absolute values of the discrepancies of each two equivalent pixels being added up. The positive number obtained in this way serves as the score. False hits may occur due to the high sensitivity that SAD exhibits to the slightest of scene alterations, including rapid camera motions, explosions, and even turning on of a light in a dark scene. By contrast, soft cuts rarely elicit a reaction from SAD. Nonetheless, because of its detection with highest probability of every visible hard cut, SAD is employed on a regular basis in the creation of a fundamental set of "possible hits".

Histogram differences (HD). This algorithm shares close similarities to SAD, with the exception that it is designed to determine the discrepancy between the histograms of two sequential frames. A histogram represents a table listing the number of pixels shaded in every colour included in a frame. False hits are less likely to occur with HD because its sensitivity to slight scene alterations is not as high as that of SAD. Nonetheless, HD does have a limitation, in that the visible content of two images can be significantly different, despite the histograms of the images being identical. For instance, the histogram of an image depicting the sea and beach and the histogram of an image depicting a cornfield and sky can be the same. Furthermore, detection of hard cuts is not absolute with HD.

Edge change ratio (ECR). This algorithm undertakes the comparison of the content of two frames based on the following procedure: conversion of the two frames to edge pictures; extraction of likely outlines of objects in the pictures; and dilatation-based comparison of the edge pictures to determine how likely it is that the objects in the first frame are also present in the second frame. For the purposes of score generation, few algorithms are more effective than ECR. It not only demonstrates high sensitivity to hard cuts, but it is also capable of detection of numerous soft cuts. However, detection of some soft cuts (e.g. wipes) is not possible when a basic version of ECR is used because the fading-in objects are classified as common objects that move through the scene. Nonetheless, particular types of soft cuts can be detected by ECR if manual extension is applied to it.

A higher performance can be achieved by merging two or more of such algorithms.

Decision Approaches

Different methods can also be employed in the decision stage of cut decision, including:

Fixed threshold: This method involves comparison of the scores to a pre-established threshold, with cut detection occurring if the threshold is lower than the scores.

Adaptive threshold: This method involves comparison of the scores to a threshold, with different scores in the video being considered to adjust the threshold to the features of the current video. As with the fixed threshold, cut detection occurs if the threshold is lower than the scores.

Our Framework

In our case, as the objective is to detect the illuminant changing in the videos, we ignore the content change and use illuminant relative features to recognise the scene change. For this reason, the change we are trying to recognise will be soft cut, not hard cut, since the illuminant change is a consequence changing process, and therefore some methods suitable for hard cut (e.g. sum of absolute differences) are not be applied in this case. Furthermore, because the content change is ignored, the edge change ratio will not be considered neither.

Despite deciding to choose histogram differences, extraction of the proper feature to illustrate the characteristics of different illuminants is still a challenging task. The feature we use in this case is *RGBY* colour feature, which is the one used in still image colour constancy model in the previous work.

Fig. 6.4 The illuminant changing range in video "entrance", the illuminant started to change around frame 60 until 75, a totally different illuminant.[FL11]

We start from a sample of video clips named "entrance". It contains 261 frames and during the whole video, there is one illuminant changing duration between frames 60 and 75. We visually hand-labelled this change area and the frames are shown in Figure (6.4).

Then, for each frame in an input video, we first derive two opponent colour signals, Red-Green (RG) and Blue-Yellow (BY), from the original *RGB* signals according to Equation (6.2).

$$RG = R - G, \quad BY = (R+G)/2 - B \tag{6.2}$$

We then treat each of the two opponent signals as an 8 bits per pixel greyscale image and construct pixel histograms H_{RG} and H_{BY} for the RG image and the BY image respectively. Here H_{RG} and H_{BY} are 256 dimensional vectors. $H_{RG}(i)$, $i = 0, 1, 2, \cdots, 255$, is the frequency of the pixels in the RG image having a quantized value of i, and similarly, $H_{BY}(i)$, $i = 0, 1, 2, \cdots, 255$, is the frequency of the pixels in the BY image having a quantized value of i.

Lastly, we calculate the H_{RG} and H_{BY} histogram differences FD between frames:

$$FD = \sum_{i}^{255} H_{RG}(i) - H_{BY}(i) \tag{6.3}$$

For a shot video with N frames, we will get the histogram differences feature $FD(N-1)$, with each element in this feature showing the differences between two neighbour frames.

As we focus on illuminant change in videos, and most cases of illuminant change in the video are a continuous process, referring to a soft cut in the shot detection area, we define two different situations and deal with each with different strategies. The first condition is the section with a low-level FD value, indicating the relative frames are under the same illuminant. Since our feature is illuminant sensitive, we define this situation as a scene. The second condition is the section with a high level FD value caused by dramatic changing of neighbour frames illuminants. We define this condition as the buffer area. How to decide the boundary between scene and buffer area is depended on the FD value, and we set the mean value of FD as a threshold value. If the current frame FD value is lower than the threshold and the next frame difference value is higher, then that next frame will be labeled as the new buffer area, otherwise it will be labelled with the current frame with the same illuminant. In contrast, if the current frame FD value is higher than the threshold and the next frame difference value is lower, then this will be the end of the current buffer area.

Firstly, we have tested our $RGBY$ feature with the most commonly used RGB colour signals (calculate the difference histogram by all three channels). Three comparisons of histogram distances of features RGB and $RGBY$ are performed after extracting the feature from test videos "entrance", "greycard", and "sofa" [FL11]. We visually hand-labelled the illuminant changing area to test the performance of both features.

The first video is "entrance". The illuminant changes from frame 60 to 75. The X-axis is the frame number and the Y-axis is the FD value.

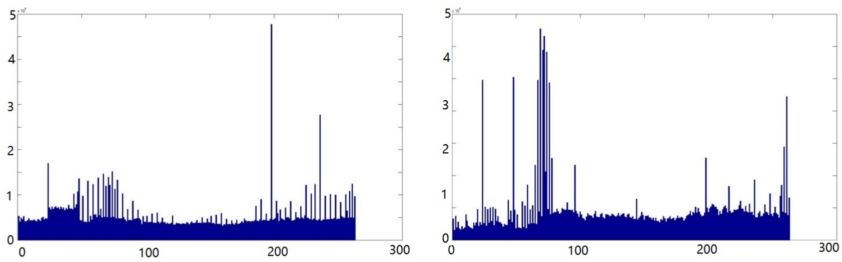


Fig. 6.5 The illuminant feature comparison in the video "entrance": the illuminant starts to change around frame 60 until 75.The left is RGB feature and the right is $RGBY$.

The second video is "greycard", it is the most complicated case in the test data course of the change of both illuminant intensity and chromaticity. The change occurs from frames 55 to 63, 126 to 140, 336 to 366, 458 to 471, 550 to 555.

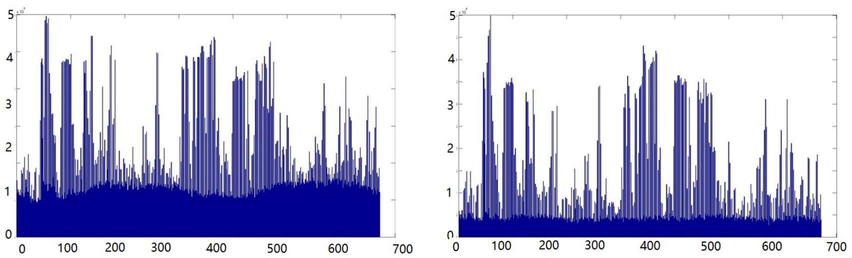


Fig. 6.6 The illuminant feature comparison in the video "greycard". The left is *RGB* feature and the right is *RGBY*.

The third video is "sofa" and the illuminant change in this test video is also very dramatic. The change occurs from frames 55 to 65, 177 to 184, 237 to 245, 304 to 309, 353 to 365, and 592 to 595.

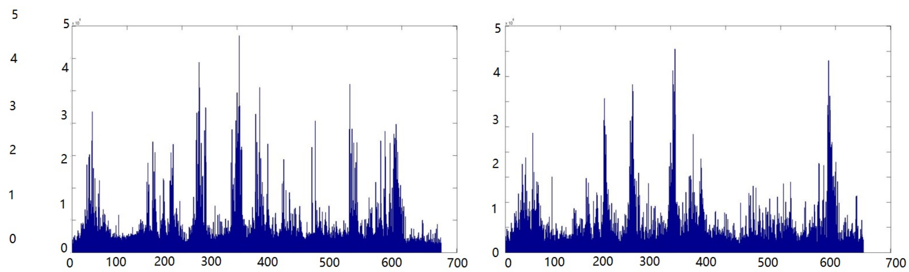


Fig. 6.7 The illuminant feature comparison in the video "sofa". The left is *RGB* feature and the right is *RGBY*.

Comparison of the three test videos reveals that the *RG*,*BY* features in the labelling area always exhibit a peak, which could indicate they are more illuminant-sensitive than the *RGB* features. However, from the figures we can still observe some noise occurs in the histogram: for example, in some areas, the occurrence of the causes our algorithm to create a wrong scene cut, and to avoid this disadvantage, we apply a median filter to smooth the histogram. The filter result is shown below, we can see the noise is greatly reduced. After this procedure, the video is cut into M scenes and $M-1$ buffer areas.

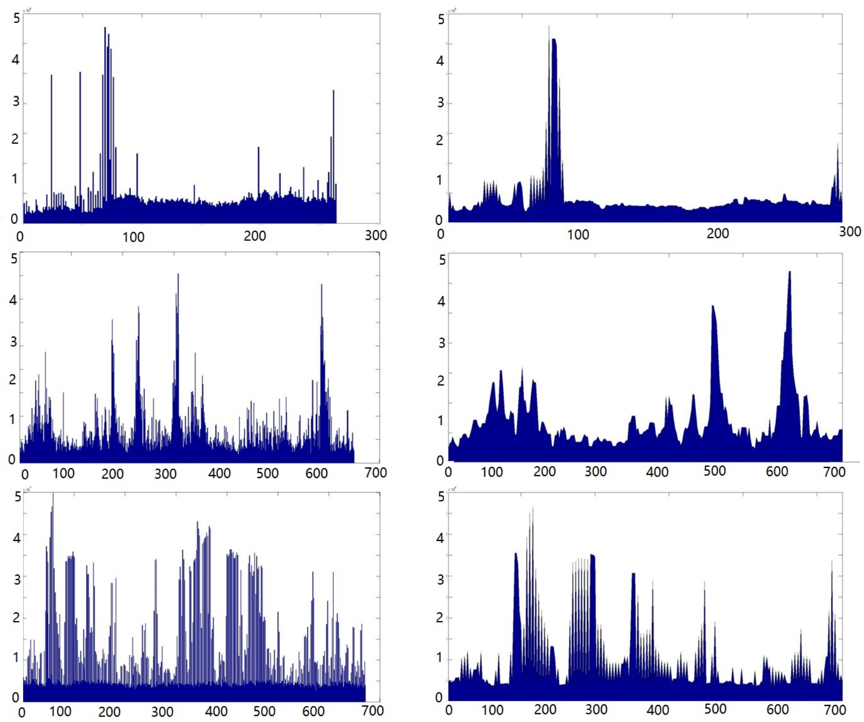


Fig. 6.8 The illuminant feature after application of the median filter, the illuminant change area is more obvious and the noise has been successfully reduced.

6.3.2 Scene Illuminant Estimation based on the $4C$-Method

In the previous section, we have shown our illuminant scene detection with *RGBY* colour signals could recognise the illuminant change in the video successfully, and the next step will be an attempt to recover the frame colour into canonical illuminant. Since our algorithm need training first, we apply the Greyball data-set with 11346 images for training, The use of this dataset is justified because it has a large number of training samples and because these images are also actually video frames, similar to the test videos. The number of training classes is 50, which is the K value of $K - means$.

For all the frames in the same scene, we will generate same colour correction parameters. However, since our algorithm can not perfectly predict the illuminant, to increase the accurate rate, we will use a voting strategy to decide which illuminant will be the prediction result. For example, all the frames in the same scene will have their own predicted illuminant label, and the most frequent result will be the final label for the current scene.

6.3.3 Linear Correction

We will create a linear correction framework [FL11], similar to the literature, to smooth the buffer area illuminant change. For each buffer area $B(i), i = 0, 1, 2, \cdots, M-2$ between two scenes, given a single frame $f(k)$, we calculate the distances between the current frame to the last frame of the last scene $DL(k)$ and the first frame of next scene $DN(k)$. The last frame of the last scene has the correction parameter (R_L, G_L, B_L), and the first frame of the next scene has the correction parameter (R_N, G_N, B_N). Therefore, the correction parameter for the current frame $f(k)$ will be:

$$\begin{aligned} R_k &= \frac{DN(k)}{DL(k)+DN(k)} R_L + \frac{DL(k)}{DL(k)+DN(k)} R_N \\ G_k &= \frac{DN(k)}{DL(k)+DN(k)} G_L + \frac{DL(k)}{DL(k)+DN(k)} G_N \\ B_k &= \frac{DN(k)}{DL(k)+DN(k)} B_L + \frac{DL(k)}{DL(k)+DN(k)} B_N \end{aligned} \tag{6.4}$$

6.4 Experiment And Recover Results

The previous section has already shown that our colour feature can recognise the illuminant change in the videos. In this section, we will show some video-recovered results from the test datasets. The first video sample is "entrance". Figure 6.9 shows the original input frames. We selected several frames that covered all of the scenes in the video. From the frame, we can see that, when the illuminant is changing from the shadow area to the outdoor sunshine, it has a significant influence on the frames. Meanwhile, the frames captured under the sunshine obviously have an incorrect illuminant.

We firstly tried to apply our previous still image colour constancy algorithm to recover the video directly. Each frame has been recovered separately, as independent images. The result is shown in Figure (6.10), from which we can see the algorithm cannot output a good prediction result for all frames since there is no image colour constancy algorithm that could perfectly recover image colours and the video problem has different issues with the still image, which is the second dimension constancy: timeline.

Moreover, we have conducted another test on the colour-recovering strategy, after the scene and buffer area detection step; for each scene, we only generated one feature for estimating the illuminant. The super feature combines all the frame features in the same scene by "or" math operation of all the binary RGBY histogram features (see Chapter 3 for binary process). It makes sense in two ways: all the frames in the same illuminant should

Fig. 6.9 Original input of video sample "entrance", the frames are selected from different parts of the duration.[FL11]

Fig. 6.10 Correct directly by separate frame

have the same correction parameter and with more information provided, the4C prediction would be more accurate.

However, the result shown in Figure (6.11) verifies our questioning: none of the colour constancy algorithms is perfect; in this case, for each scene, the chance of prediction would be just one, and if the prediction is not correct, it will negatively influence the final result and also the buffer area recovery, since the function is strongly depended on the parameters of the neighbour illuminants.

Figure (6.12) shows the recovered result using the voting strategy. From the video, we can see the illuminant has been recovered correctly. Based on this result, the buffering area can also correctly transfer the colour to canonical illuminant. The comparison is shown in Figures 6.13 and 6.14.

Fig. 6.11 Correct illuminant with one super feature

Fig. 6.12 Correct illuminant with voting strategy

However, the limitation of the algorithm is also very clear; even though the voting strategy does improve the accuracy of the prediction, it can still go wrong sometimes. When this happens, the video correction will yield an incorrect result. Furthermore, the disadvantage of the classification method is that provision of only a few classes is not enough to cover the entire illuminant change situation. If the ground truth is not included in the training dataset, or the frames have an illuminant that is too unusual (as the frames shown in Figure 6.15), the nature of the video will worsen this disadvantage, as the second dimension of constancy is timeline. The next few figures are the frames from "sofa" and "greycard"; in both of them, the illuminant changes are more complicated, and therefore our algorithm can only recover parts of the scene correctly, while other parts with very unusual illuminant fail to provide a satisfying recover output. Nevertheless, as an extension research topic for current still image colour constancy algorithms, the work warrants further research; for example, a combination of current tonal stabilisation algorithms to provide anchor frames with correct colour is a viable research direction.

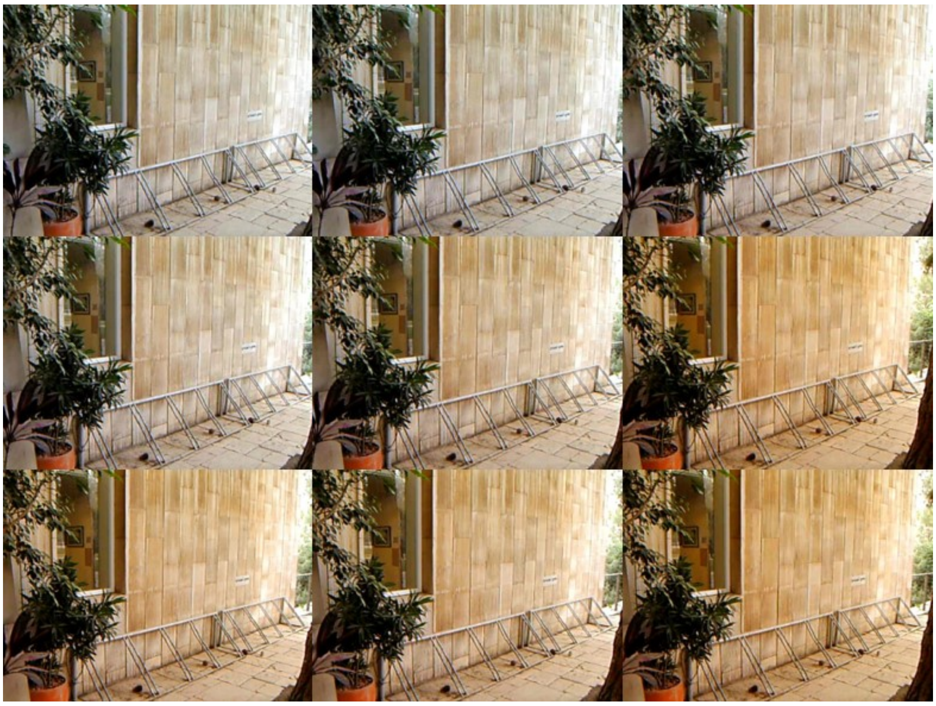


Fig. 6.13 Buffer area before linear correction.[FL11]

Fig. 6.14 Buffer area after linear correction.

Fig. 6.15 The original input video frames of "sofa".[FL11]

Fig. 6.16 The failed correction of input video frames of "sofa".

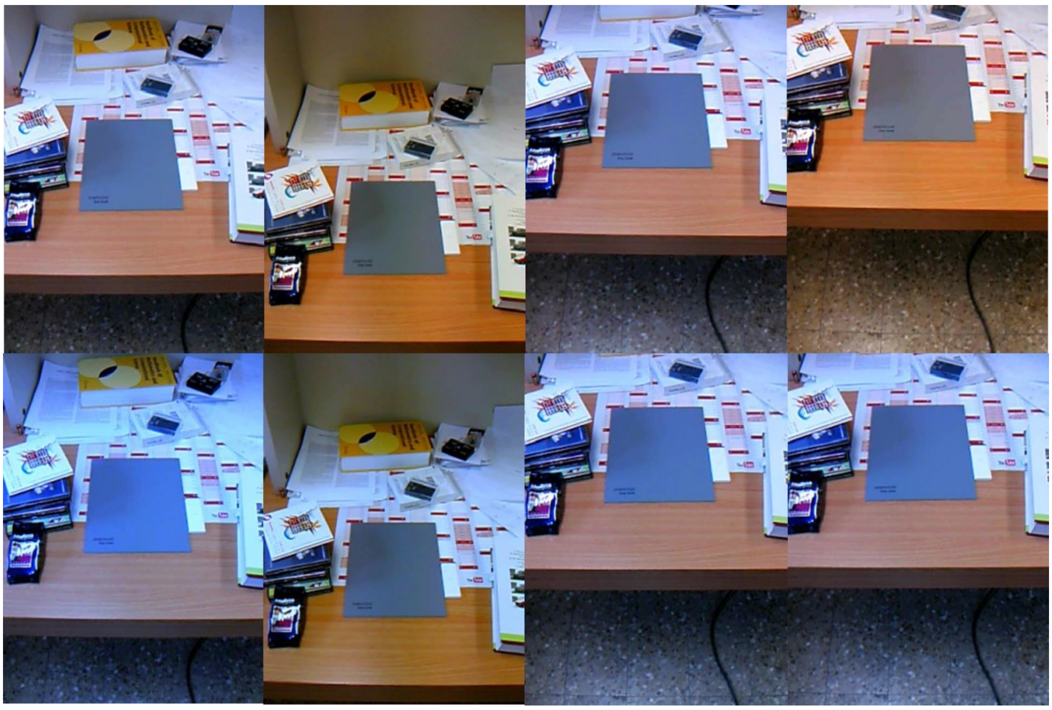


Fig. 6.17 Comparison between the video frames of "greycard"; the top line is the original input frames, the bottom line is the relative "corrected" frames achieved with our method.[FL11]

6.5 Conclusion

In this chapter, we have designed a video illuminant recognition framework with our illuminant sensitive colour feature *RG*, *BY* to provide a successful scene illuminant detection result. Furthermore, we implemented our still image colour constancy algorithm to tackle the two-dimensional video colour constancy problem, instead of finding the anchor frame, which is the most widely used video tonal stabilisation method. We tried to recover all the frames colour to the canonical illuminant. We adopted the voting strategy to increase the accurate rate of prediction and recover the illuminant change buffer area with a linear function. We tested our method on several videos, with results showing that the method could perform well on videos without dramatic illuminant changing. However, limitation occurred when the illuminant change situation was very complicated and there were not enough classes for training; in such a situation, the algorithm could not produce a satisfying result, but as a new way to try to jointly deal with the two video problems (i.e. tonal stabilisation and video colour correction), the work conducted so far warrants further discussion and research.

Chapter 7

Conclusions

The principal aim of this thesis was to propose an efficient statistical algorithm for illuminant estimation. First, we presented a review of the state-of-the-art illuminant estimation algorithms, highlighting the relevance and potential of the framework based on the learning method. We then presented our new algorithm and experiments on different image datasets. The results showed that the new approach delivered good illuminant estimation, most of the time outperforming the state-of-the-art algorithms and sometimes achieving as good an estimation as much more complex algorithms. Furthermore, we improved our still image illuminant estimation framework into video frames, to tackle the video tonal stabilisation problem, although the result still left room for improvement, showing good potential for further research.

In this Chapter, we summarise the main contributions of this thesis together with a discussion of the possible applications of the work.

7.1 Contributions

In Chapter 3, our work started from developing a new novel colour constancy algorithm: we treated the illuminant estimation problem as an illumination classification problem, then applied a strong and robust classifier, Random Forest, as our core technology to develop the algorithm. This yielded the first contribution of our work, namely, creation of a new colour signal RG, BY, which could indicate the illuminant character more obviously than other colour features, and the outcome of the comparison conducted demonstrated the advantage of our new colour feature. Secondly, we used our new colour feature to build a successful illuminant classifier by implementing Random Forest to the illuminant estimation problem. We evaluated our algorithm on two datasets, for which the data had been already classified (or could be treated as classified, as the grey ball data were used as the same scene with the same illuminant assumption), and the performance has shown our technique was robust and competitive with other state-of-the-art methods. After that we discussed the disadvantage of our method: as a classifier, the data need to be well-classified before training, and to reduce the negative effect and improve the performance of our algorithm, we came up with two potential approaches, namely, using a large dataset to generate more classes for training and classifying the training data more specifically.

In Chapter 4, we specified our approach to the implementation with synthetic dataset. To deal with the disadvantage of the classification framework for illuminant estimation, the synthetic dataset was used as it could provide an arbitrary large number of images and illuminants for training. The synthetic dataset had all the necessary physical based components of creating colours: the illuminant spectral data, surface reflection data, and camera characterisation data. By using the data provided, we generated a much larger training dataset to construct our Random Forest, and compared in the same way with the previous experiment. The result showed that our method still provided good performance with synthetic data, and furthermore, we also tried to test the algorithm on real-world image dataset to evaluate the realistic meaning of the dataset, which has not been comprehensively studied before. However, the experiment showed there was still a huge gap between these two datasets, meaning that better ways need to be found to optimise the generation of the synthetic data. Therefore, future work should focus on classifying the data more specifically by using the clustering method.

In Chapter 5, the ground truth of the illuminants was mostly collected by the white point of the scene, referred to as $R_wG_wB_w$, which was obviously the colour of the illuminant under the imaging system. It is always possible to obtain these measurements by placing a neutral white patch in the scene when collecting training images, and some of the datasets adopted this approach. These white points could be used to cluster the training images into different clusters based on similar illuminant. Starting from this, we introduced a novel clustering classification colour constancy framework referred to as the $4C$ method. Based on the assumption that similar illuminants will have similar illumination colours (colours of the white patch of the scenes), we first used a clustering algorithm to group similar $R_wG_wB_w$'s of the training samples into the same cluster. We then treated the $R_wG_wB_w$'s in the same cluster as belonging to the same illumination source and each cluster as one class of illuminant. The colour constancy problem, i.e., that of estimating the unknown illuminant of an image, became that of identifying which illuminant class (cluster) the illuminant of the image belonged to. To achieve this, we used a classification algorithm to classify the image into the illuminant class (cluster). In our work, we have tested the classic K-means and frequency sensitive competitive learning ($FSCL$) as the clustering method and the Random

Forest and support vector machine (SVM) as the classification method. Furthermore, the experiments were conducted on three benchmark datasets, and one of them was "Colour Checker", which required clustering of the images before training. After comparison with a large range of classic statistical methods, gamut mapping methods and leading machine learning methods (include the very latest deep learning neural network colour constancy algorithm), the result showed that our $4C$ framework was competitive with state-of-the-art methods.

In Chapter 6, to explore further potential development in colour constancy relative problem, we addressed the colour constancy problem from still image to video. The video tonal stabilisation problem is still an unsolved problem, and current algorithms are only focusing on keeping the tonal stable during the video playing, not really trying to recover the incorrect illuminant, since applying colour constancy algorithms directly to recover video frames was not realistic (the experiments conducted supported this). However, even with this disadvantage, we were still trying to combine these two problems together, as firstly keeping the tonal stable and secondly recovering the frame colour to canonical illuminant, and proposed an algorithm for improving the performance of the video. Our approach operated on a continuous video shot by first cutting the video into one or more scenes based on features presenting the characteristics of illuminants. We then corrected a sequence of frames in the same scene by using the Random Forest illuminant estimation framework: for every scene, we extracted the features of each frame in the same scene and returned an illuminant estimation result. Subsequently, we used the most-voted result as the estimation of this scene and corrected the tonal information. Finally, a smooth function between different scenes was applied to adjust the flick and flash that occurred at the boundary of the neighbour scenes. Through a series of experiments conducted in different ways, we demonstrated that each component in our framework could improve the quality of the video to a certain extent. However, there was a limitation in that, when the illuminant change situation was very complicated and there were not enough classes for training, the algorithm could not produce a satisfying result. Nevertheless, as a new approach to try to deal with the two problems in videos together (i.e. tonal stabilisation and video colour correction), the work warrants further discussion and research.

7.2 Future Work

As outlined in the following part, a number of viable lines of inquiry could contribute to the expansion of colour constancy research.

Dichromatic Reflection Model

As previously indicated, the **Mondrian World** imaging model [L+77] is the simplified reflection model that underpins the present work. The assumptions associated with this model regard the spatial homogeneity of the illuminant, the flatness and coplanarity of the objects, and the Lambertian nature of surface reflectance. To minimise the complexity of the colour reflection framework, the spatial distribution of the illuminant, scene geometry and the impact of the direction of incident and reflected light on object spectral reflectance are not considered by the **Mondrian World** either. However, the assumptions of this model are always incompatible with real-world scenes, despite evidence from studies [Bra04, For90, FCB96] attesting to the richness of **Mondrian World** that make it a suitable framework for preliminary analysis.

To characterise the division of the specular and diffuse reflection components, the **Dichromatic Reflection Model** was proposed by Shafer in 1985 [Sha85], based on the assumption that the scene had just one illuminant, with no ambient light or object inter-reflection. The light, $L(\lambda,\Theta)$, is characterised by this model as being reflected from a point on a dielectric, heterogeneous material and consisting of the light $L_s(\lambda,\Theta)$ and the light $L_b(\lambda,\Theta)$, respectively reflected from the material surface and the material body. The incident and emitted light angles and the phase angle in relation to the surface normal are defined by the vector $\Theta = (\theta_i, \phi_i, \theta_r, \phi_r)$, while the wavelength parameter is denoted by λ and the surface or specular reflection component is denoted by L_s. Being similar to the illumination in terms of spectral power distribution, L_s takes the form of a highlight or gloss on the object. Meanwhile, the body or diffuse reflection component L_b gives the defining object colour of the surface while also comprising the object shading attributes:

$$L(\lambda,\Theta) = L_b(\lambda,\Theta) + L_s(\lambda,\Theta) \tag{7.1}$$

The spectral and geometric reflection attributes of L_s and L_b are separated by the **Mondrian World** and are modelled as outcomes of spectral power distributions, $cs(\underline{U})$ or $cb(\underline{U})$, and geometric scale factors, $ms(\underline{B})$ or $mb(\underline{B})$, characteristic of the reflected light intensity.

Segmentation of colour images and extraction of photometric invariant features are the two main applications of the **DRM** model. In the case of object segmentation, scene incidental events (e.g. shading and specularities) pose considerable difficulty. For these events, it is possible to anticipate how colour distributions will behave based on the model supplied by **DRM**. A line is generated by the object body reflectance from the colour space source with an angle influenced by the colour of both the object and the illuminant. Another line in the direction of the colour of the illuminant is generated by specularities. The issue stems from the fact that information on body reflectance and illuminant colour is necessary for accurate image segmentation and the other way around [KSK90, MS97]. Therefore, a relevant line of inquiry that could be pursued in future research and contribute to the computer vision field is the extension of the proposed **Mondrian World** to the dichromatic reflection model.

Colour Constancy for More Than One Illuminant

The working premise underpinning most colour constancy algorithms is that the illuminant has a homogeneous spectral distribution throughout a scene. However, the occurrence of more than one illuminant means that this premise seldom holds true in practice. This topic has been the focus of many studies [GLG12, FFB95, JD14]. The first technique that was introduced was applicable exclusively to images of the outdoors and could differentiate shadow areas from non-shadow ones [KIT05] . Subsequently, additional techniques capable of this differentiation were developed, but the output images they produced were visually distinct from the initial input image [FH01, LD05]. A different technique was designed to replicate the attributes of opponent and double-opponent cells by employing retinal mechanisms and adjustment [SS02] . The Exemplar-based technique discussed in the experiment part is the most recent technique that takes into account multiple light sources [JD14]. The work employed the approach of unsupervised learning of a suitable model for every training surface in training images as a way of dealing with the colour constancy issue. The process began with image segmentation into various surfaces, followed by identification of the closest neighbour models for every surface in a test image and approximation of its illuminant on the basis of the comparison of the statistics of pixels associated with the closest neighbour surfaces and the target surface. The fusion of the approximated illuminants across surfaces into a single estimate gives the final illuminant estimation result. At the moment, this is the main approach used to address the issue of multiple illuminants. Therefore, it would be worth investigating in a future study how the proposed framework could be applied from one illuminant to multiple illuminants.

Digital Camera White Balance

In the field of digital photography, most digital cameras have means to carry out colour correction based on the type of scene illumination, where automatic illuminant selection could benefit the illuminant estimator presented in this work. The camera white balance is actually the colour constancy problem. For this reason, future work should focus on the commercialisation of the algorithms presented in this thesis into real processing pipelines.

References

[AAKA06] Vivek Agarwal, Besma R Abidi, Andreas Koschan, and Mongi A Abidi. An overview of color constancy algorithms. *Journal of Pattern Recognition Research*, 1(1):42–54, 2006.

[AGA07] Vivek Agarwal, Andrei V Gribok, and Mongi A Abidi. Machine learning approach to color constancy. *Neural Networks*, 20(5):559–563, 2007.

[AGK+09] Vivek Agarwal, Andrei Gribok, Andreas Koschan, B Abidi, and M Abidi. Illumination chromaticity estimation using linear learning methods. *Journal of Pattern Recognition Research*, 4(1):92–109, 2009.

[AGKA06] Vivek Agarwal, Andrei V Gribok, Andreas Koschan, and Mongi A Abidi. Estimating illumination chromaticity via kernel regression. In *2006 International Conference on Image Processing*, pages 981–984. IEEE, 2006.

[AKCM90] Stanley C. Ahalt, Ashok K. Krishnamurthy, Prakoon Chen, and Douglas E. Melton. Competitive learning algorithms for vector quantization. *Neural Networks*, 3(3):277 – 290, 1990.

[AP10] Xiaobo An and Fabio Pellacini. User-controllable color transfer. In *Computer Graphics Forum*, volume 29, pages 263–271. Wiley Online Library, 2010.

[Bar99] K. Barnard. A practical color constancy. *Ph.D. dissertation*, 27(3):147–151, 1999.

[BBC+12] Ronan Boitard, Kadi Bouatouch, Remi Cozot, Dominique Thoreau, and Adrien Gruson. Temporal coherency for video tone mapping. In *SPIE Optical Engineering+ Applications*, pages 84990D–84990D. International Society for Optics and Photonics, 2012.

[BCCS10] Simone Bianco, Gianluigi Ciocca, Claudio Cusano, and Raimondo Schettini. Automatic color constancy algorithm selection and combination. *Pattern Recognition*, 43:695–705, 2010.

[BCF01] Kobus Barnard, Florian Ciurea, and Brian Funt. Sensor sharpening for computational color constancy. *JOSA A*, 18(11):2728–2743, 2001.

[BCF02] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *Image Processing, IEEE Transactions on*, 11(9):972–984, 2002.

[BF94] David H Brainard and William T Freeman. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pages 364–376. International Society for Optics and Photonics, 1994.

[BF97] David H Brainard and William T Freeman. Bayesian color constancy. *JOSA A*, 14(7):1393–1411, 1997.

[BF02] Kobus Barnard and Brian Funt. Camera characterization for color research. *COLOR research and application*, 27(3):153–164, 2002.

[BG09] Gertjan J Burghouts and Jan-Mark Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113(1):48–62, 2009.

[BGS08] Simone Bianco, Francesca Gasparini, and Raimondo Schettini. Consensus-based framework for illuminant chromaticity estimation. *Journal of Electronic Imaging*, 17(2):023013–023013, 2008.

[BJK07] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

[BMCF02] Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Fun. A comparison of computational color constancy algorithms. ii. experiments with image data. *Image Processing, IEEE Transactions on*, 11(9):985–996, 2002.

[BMF00] Kobus Barnard, Lindsay Martin, and Brian Funt. Colour by correlation in a three-dimensional colour space. In *Computer Vision-ECCV 2000*, pages 375–389. Springer, 2000.

[BMFC02] Kobus Barnard, Lindsay Martin, Brian Funt, and Adam Coath. A data set for color research. *Color Research & Application*, 27(3):147–151, 2002.

[Bra04] David Hč Brainard. Color constancy. *The visual neurosciences*, 1:948–961, 2004.

[Bre01] Leo Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

[BTS+15] Nicolas Bonneel, James Tompkin, Kalyan Sunkavalli, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. Blind video temporal consistency. *ACM Transactions on Graphics (TOG)*, 34(6):196, 2015.

[Buc80] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1 – 26, 1980.

[BVdW11] Shida Beigpour and Joost Van de Weijer. Object recoloring based on intrinsic image estimation. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 327–334. IEEE, 2011.

[BW86] David H Brainard and Brian A Wandell. Analysis of the retinex theory of color vision. *JOSA A*, 3(10):1651–1661, 1986.

[BW92] David H Brainard and Brian A Wandell. Asymmetric color matching: how color appearance depends on the illuminant. *JOSA A*, 9(9):1433–1448, 1992.

[CB00] Ronan Collobert and Samy Bengio. Support vector machines for large-scale regression problems. Technical report, IDIAP, 2000.

[CDL08] Xu Cheng, Cameron Dale, and Jiangchuan Liu. Statistics and social network of youtube videos. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, pages 229–238. IEEE, 2008.

[CF99] Vlad C Cardei and Brian Funt. Committee-based color constancy. pages 311–313. Scottsdale, Arizona, 1999.

[CF03] Florian Ciurea and Brian Funt. A large image database for color constancy research. In *Color and Imaging Conference*, volume 2003, pages 160–164. Society for Imaging Science and Technology, 2003.

[CFB02] Vlad C Cardei, Brian Funt, and Kobus Barnard. Estimating the scene illumination chromaticity by using a neural network. *JOSA A*, 19(12):2374–2386, 2002.

[CGM02] Artur Chodorowski, Tomas Gustavsson, and Ulf Mattsson. Support vector machines for oral lesion classification. In *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pages 173–176. IEEE, 2002.

[CK02] Hung-Shing Chen and Hiroaki Kotera. Three-dimensional gamut mapping method based on the concept of image dependence. *Journal of Imaging Science and Technology*, 46(1):44–52, 2002.

[CL11] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[CM$^+$05] Frank M Candocia, Daniel Mandarino, et al. A semiparametric model for accurate camera response function modeling and exposure estimation from comparametric data. *Image Processing, IEEE Transactions on*, 14(8):1138–1150, 2005.

[CNP06] Costas Cotsaces, Nikos Nikolaidis, and Ioannis Pitas. Video shot boundary detection and condensed representation: a review. *IEEE signal processing magazine*, 23(2):28–37, 2006.

[DJ09] Mark S Drew and Hamid Reza Vaezi Joze. Sharpening from shadows: Sensor transforms for removing shadows using a single image. In *Color and Imaging Conference*, volume 2009, pages 267–271. Society for Imaging Science and Technology, 2009.

[DM08] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM, 2008.

[DXW01] Jeffrey M DiCarlo, Feng Xiao, and Brian A Wandell. Illuminating illumination. In *Color and Imaging Conference*, volume 2001, pages 27–34. Society for Imaging Science and Technology, 2001.

[DZJ02] AiLing Ding, XiangMo Zhao, and LiCheng Jiao. Traffic flow time series prediction based on statistics learning theory. In *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, pages 727–730. IEEE, 2002.

[Ebn03] Marc Ebner. Combining white-patch retinex and the gray world assumption to achieve color constancy for multiple illuminants. In Bernd Michaelis and Gerald Krell, editors, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 60–67. Springer Berlin Heidelberg, 2003.

[Ebn04] Marc Ebner. A parallel algorithm for color constancy. *Journal of Parallel and Distributed Computing*, 64(1):79 – 88, 2004.

[Ebn07] Marc Ebner. *Color constancy*, volume 6. John Wiley & Sons, 2007.

[Fai13] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.

[FBL96] Barnard K Funt B and Martin L. Is machine colour constancy good enough? In *In Fifth European Conference on Computer Vision (ECCV 98)*, pages 445–459. Springer-Verlag, Berlin. Germany, 1996.

[FCB96] Brian Funt, Vlad Cardei, and Kobus Barnard. Learning color constancy. In *Color and Imaging Conference*, volume 1996, pages 58–60. Society for Imaging Science and Technology, 1996.

[FD01] Graham D Finlayson and Mark S Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 473–480. IEEE, 2001.

[FDF94] Graham D Finlayson, Mark S Drew, and Brian V Funt. Color constancy: generalized diagonal transforms suffice. *JOSA A*, 11(11):3011–3019, 1994.

[FDH91] Brian V Funt, Mark S Drew, and Jian Ho. Color constancy from mutual reflection. *International Journal of Computer Vision*, 6(1):5–24, 1991.

[FDL04] Graham D Finlayson, Mark S Drew, and Cheng Lu. Intrinsic images by entropy minimization. In *Computer Vision-ECCV 2004*, pages 582–595. Springer, 2004.

[FF95] Brian V. Funt and Graham D. Finlayson. Color Constant Color Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:522–529, 1995.

[FFB95] Graham D Finlayson, Brian V Funt, and Kobus Barnard. Color constancy under varying illumination. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 720–725. IEEE, 1995.

[FGJ98] Schiele B Finlayson GD and Crowley JL. Comprehensive colour image normalization. In *Fifth European Conference on Computer Vision (ECCV 98)*, pages 475–490. Springer, 1998.

[FH98] Graham Finlayson and Steven Hordley. A theory of selection for gamut mapping colour constancy. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 60–65. IEEE, 1998.

[FH00] Graham Finlayson and Steven Hordley. Improving gamut mapping color constancy. *Image Processing, IEEE Transactions on*, 9(10):1774–1783, 2000.

[FH01] Graham D Finlayson and Steven D Hordley. Color constancy at a pixel. *JOSA A*, 18(2):253–264, 2001.

[FHH97] Graham D Finlayson, Paul M Hubel, and Steven Hordley. Color by correlation. In *Color and Imaging Conference*, volume 1997, pages 6–11. Society for Imaging Science and Technology, 1997.

[FHH99] Graham D. Finlayson, Steven D. Hordley, and Paul M. Hubel. Colour by Correlation: A Simple, Unifying Approach to Colour Constancy. In *International Conference on Computer Vision*, pages 835–842, 1999.

[Fin95] Graham David Finlayson. *Coefficient color constancy*. PhD thesis, Citeseer, 1995.

[Fin96] Graham D. Finlayson. Color in perspective. *IEEE transactions on Pattern analysis and Machine Intelligence*, 18(10):1034–1038, 1996.

[Fin00] Graham D Finlayson. Computational colour constancy. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 1, pages 191–196. IEEE, 2000.

[FL00] Brian V Funt and Benjamin C Lewis. Diagonal versus affine transformations for color correction. *JOSA A*, 17(11):2108–2112, 2000.

[FL11] Zeev Farbman and Dani Lischinski. Tonal stabilization of video. *ACM Transactions on Graphics (TOG)*, 30(4):89, 2011.

[Fol90] van Dam A. Feiner S. K. Hughes J. F. Foley, J. D. Computer graphics: Principles and practice.(2 ed.). 1990.

[For90] David A Forstyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990.

[Fos03] David H Foster. Does colour constancy exist? *Trends in cognitive sciences*, 7(10):439–443, 2003.

[FR98] Adrian Ford and Alan Roberts. Colour space conversions. *Westminster University, London*, 1998:1–31, 1998.

[FS08] Gregor Fischer and Matthias Sajjaa. whitebalpr–a new method for automatic white balance. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2008, pages 202–207. Society for Imaging Science and Technology, 2008.

[FS09] Clément Fredembach and Sabine Susstrunk. Illuminant estimation and detection using near-infrared. In *IS&T/SPIE Electronic Imaging*, pages 72500E–72500E. International Society for Optics and Photonics, 2009.

[FT04] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[FX04] Brian Funt and Weihua Xiong. Estimating illumination chromaticity via support vector regression. In *Color and Imaging Conference*, volume 2004, pages 47–52. Society for Imaging Science and Technology, 2004.

[GG11] Arjan Gijsenij and Theo Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):687–698, 2011.

[GGVdW07] Arjan Gijsenij, Theo Gevers, and Joost Van de Weijer. Color constancy by derivative-based gamut mapping. In *PACV 2007-1rst International Workshop on Photometric Analysis For Computer Vision*, pages 8–p. INRIA, 2007.

[GGVDW10] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, 2010.

[GGVDW11] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Computational color constancy: Survey and experiments. *Image Processing, IEEE Transactions on*, 20(9):2475–2489, 2011.

[GLG12] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *Image Processing, IEEE Transactions on*, 21(2):697–707, 2012.

[GRB$^+$08] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[GS99] Theo Gevers and Arnold WM Smeulders. Color-based object recognition. *Pattern recognition*, 32(3):453–464, 1999.

[GvdBSG03] Jan-Mark Geusebroek, Rein van den Boomgaard, Arnold WM Smeulders, and Theo Gevers. Color constancy from physical principles. *Pattern Recognition Letters*, 24(11):1653–1662, 2003.

[HF00] Paul M Hubel and Graham Finlayson. White point determination using correlation matrix memory, March 14 2000. US Patent 6,038,339.

[HF14] John F Hughes and James D Foley. *Computer graphics: principles and practice*. Pearson Education, 2014.

[HHF+97] Paul M Hubel, Jack Holm, Graham D Finlayson, Mark S Drew, et al. Matrix calculations for digital photography. In *The Fifth Color Imaging Conference: Color, Science, Systems and Applications*, 1997.

[HL02] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002.

[Hor06] Steven D. Hordley. Scene illuminant estimation: Past, present, and future. *Color Research Application*, 31(4):303–314, 2006.

[HS94] Glenn Healey and David Slater. Global color constancy: recognition of objects by use of illumination-invariant properties of color distributions. *JOSA A*, 11(11):3003–3010, 1994.

[JD10] Hamid Reza Vaezi Joze and Mark S Drew. Improved machine learning for image category recognition by local color constancy. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 3881–3884. IEEE, 2010.

[JD12] Hamid Reza Vaezi Joze and Mark S Drew. Exemplar-based colour constancy. In *BMVC*, pages 1–12, 2012.

[JD14] H.R.V. Joze and M.S. Drew. Exemplar-based color constancy and multiple illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(5):860–873, May 2014.

[JZQ11] Guo-Quan Jiang, Cui-Jun Zhao, and Jun-Yan Qi. The research of image segmentation based on color characteristic. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, volume 4, pages 1851–1855. IEEE, 2011.

[KIT05] Rei Kawakami, Katsushi Ikeuchi, and Robby T Tan. Consistent surface color for texturing large objects in outdoor scenes. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1200–1207. IEEE, 2005.

[KMHO09] Sefy Kagarlitsky, Yael Moses, and Yacov Hel-Or. Piecewise-consistent color mappings of images acquired under various conditions. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2311–2318. IEEE, 2009.

[Kri53] Evgenij L Krinov. Spectral reflectance properties of natural formations. 1953.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[KSK90] Gudrun J Klinker, Steven A Shafer, and Takeo Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1):7–38, 1990.

[L+77] Edwin H Land et al. *The retinex theory of color vision*. Scientific America., 1977.

[LAA08] Yuanzhen Li, Edward Adelson, and Aseem Agarwala. Scribbleboost: Adding classification to edge-aware interpolation of local image and video adjustments. In *Computer Graphics Forum*, volume 27, pages 1255–1264. Wiley Online Library, 2008.

[LD05] Cheng Lu and Mark S Drew. Shadow segmentation and shadow-free chromaticity via markov random fields. In *Color and Imaging Conference*, volume 2005, pages 125–129. Society for Imaging Science and Technology, 2005.

[Lee86] Hsien-Che Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *JOSA A*, 3(10):1694–1699, 1986.

[LGG+09a] Rui Lu, Arjan Gijsenij, Theo Gevers, Vladimir Nedovi, De Xu, and Jan-Mark Geusebroek. Color constancy using 3d scene geometry. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1749–1756. IEEE, 2009.

[LGG+09b] Rui Lu, Arjan Gijsenij, Theo Gevers, Koen EA van de Sande, Jan-Mark Geusebroek, and De Xu. Color constancy using stage classification. In *ICIP*, pages 685–688, 2009.

[LGHL15] Zhongyu Lou, T Gevers, Ninghang Hu, and Marcel Lucassen. Color constancy by deep learning. In *BMVC*, pages 76.1–76.12, 01 2015.

[LGJA09] Feng Liu, Michael Gleicher, Hailin Jin, and Aseem Agarwala. Content-preserving warps for 3d video stabilization, 2009.

[LL03] Yoonkyung Lee and Cheol-Koo Lee. Classification of multiple cancer types by multicategory support vector machines using gene expression data. *Bioinformatics*, 19(9):1132–1139, 2003.

[LLW04] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 689–694. ACM, 2004.

[LM09] Michela Lecca and Stefano Messelodi. Computing von kries illuminant changes by piecewise inversion of cumulative color histograms. *Electronic Letters on Computer Vision and Image Analysis*, 8(2), 2009.

[LW02] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[MB09] R. Malik and P. Bajcsy. Achieving color constancy across multiple cameras. *in Proc. ACM International Conference on Multimedia*, pages 893–896, 2009.

[MKNM95] Babu M Mehtre, Mohan S Kankanhalli, A Desai Narasimhalu, and Guo Chang Man. Color matching for image retrieval. *Pattern Recognition Letters*, 16(3):325–331, 1995.

[MN99] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999.

[MOT+06] Yasuyuki Matsushita, Eyal Ofek, Xiaoou Tang, Senior Member, and Heung yeung Shum. Full-frame video stabilization with motion inpainting. *IEEE Trans. Patt. Anal. Mach. Intell*, pages 1150–1163, 2006.

[MP94] S Mann and RW Picard. Extending dymanic range by combining differently exposed pictures. Technical report, Technical Report 289, MIT Media Lab Perceptual Computing Section, Boston, Massachusetts, 1994.

[MS97] Bruce A Maxwell and Steven A Shafer. Physics-based segmentation of complex objects using multiple hypotheses of image formation. *Computer Vision and Image Understanding*, 65(2):269–295, 1997.

[NBE58] SM Newhall, RW Burnham, and RM Evans. Color constancy in shadows. *JOSA*, 48(12):976–984, 1958.

[NDB09] WANG Ning, XU De, and LI Bing. Edge-based color constancy via support vector regression. *IEICE transactions on information and systems*, 92(11):2279–2282, 2009.

[NPVR08] Juan L Nieves, Clara Plata, Eva M Valero, and Javier Romero. Unsupervised illuminant estimation from natural scenes: an rgb digital camera suffices. *Applied optics*, 47(20):3574–3584, 2008.

[NSRG10] Vladimir Nedovic, Arnold WM Smeulders, Andre Redert, and Jan-Mark Geusebroek. Stages as models of scene geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1673–1687, 2010.

[OK16] Seoung Wug Oh and Seon Joo Kim. Approaching the computational color constancy as a classification problem through deep learning. *Pattern Recognition*, 2016.

[Pit00] Ioannis Pitas. *Digital image processing algorithms and applications*. John Wiley & Sons, 2000.

[Poy97] Charles Poynton. Frequently asked questions about color. *Retrieved June*, 19:2004, 1997.

[Poy12] Charles Poynton. *Digital video and HD: Algorithms and Interfaces*. Elsevier, 2012.

[RAGS01] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, (5):34–41, 2001.

[RGM02] Alessandro Rizzi, Carlo Gatta, and Daniele Marini. Color correction between gray world and white patch. In *Electronic Imaging 2002*, pages 367–375. International Society for Optics and Photonics, 2002.

[RLM03] Charles Rosenberg, Alok Ladsariya, and Tom Minka. Bayesian color constancy with non-gaussian models. In *Advances in neural information processing systems*, page None, 2003.

[RMEJ05] J-P Renno, Dimitrios Makris, Tim Ellis, and Graeme A Jones. Application and evaluation of colour constancy in visual surveillance. In *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 301–308. IEEE, 2005.

[RNK$^+$09] Esa Rahtu, Jarno Nikkanen, Juho Kannala, Leena Lepistö, and Janne Heikkilä. Applying visual object categorization and memory colors for automatic color constancy. In *International Conference on Image Analysis and Processing*, pages 873–882. Springer, 2009.

[Sap98] Guillermo Sapiro. Bilinear voting. In *Computer Vision, 1998. Sixth International Conference on*, pages 178–183. IEEE, 1998.

[SF09] Matthias Sajjaa and Gregor Fischer. Automatic white balance: whitebalpr using the dichromatic reflection model. In *IS&T/SPIE Electronic Imaging*, pages 72500D–72500D. International Society for Optics and Photonics, 2009.

[SF10] Lilong Shi and Brian Funt. Re-processed version of the gehler color constancy dataset of 568 images. *Simon Fraser University*, 1(2):3, 2010.

[SG31] Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the optical society*, 33(3):73, 1931.

[SH12] Raju Shrestha and Jon Yngve Hardeberg. Computational color constancy using a stereo camera. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2012, pages 69–74. Society for Imaging Science and Technology, 2012.

[Sha85] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[SHF05] Gerald Schaefer, Steven Hordley, and Graham Finlayson. A combined physical and statistical approach to colour constancy. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 148–153. IEEE, 2005.

[SS02] Hedva Spitzer and Sarit Semo. Color constancy: a biological model and its application for still and video images. *Pattern Recognition*, 35(8):1645–1659, 2002.

[SS04] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[SSI03] Imari Sato, Yoichi Sato, and Katsushi Ikeuchi. Illumination from shadows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(3):290–300, 2003.

[SXF11] Lilong Shi, Weihua Xiong, and Brian Funt. Illumination estimation via thin-plate spline interpolation. *JOSA A*, 28(5):940–948, 2011.

[TEW01] Shoji Tominaga, Satoru Ebisui, and Brian A Wandell. Scene illuminant classification: brighter is better. *JOSA A*, 18(1):55–64, 2001.

[TJT05] Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 747–754. IEEE, 2005.

[Tom96] Shoji Tominaga. Surface reflectance estimation by the dichromatic model. *Color Research & Application*, 21(2):104–114, 1996.

[Tov96] Martin J Tovée. *An introduction to the visual system*. Cambridge University Press, 1996.

[TRK01] Yanghai Tsin, Visvanathan Ramesh, and Takeo Kanade. Statistical calibration of ccd imaging process. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 480–487. IEEE, 2001.

[VDSGS10] Koen EA Van De Sande, Theo Gevers, and Cees GM Snoek. Evaluating color descriptors for object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1582–1596, 2010.

[VDWG05] Joost Van De Weijer and Theo Gevers. Color constancy based on the grey-edge hypothesis. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–722. IEEE, 2005.

[VdWSV07] Joost Van de Weijer, Cordelia Schmid, and Jakob Verbeek. Using high-level visual information for color constancy. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[VGI94] Michael J Vrhel, Ron Gershon, and Lawrence S Iwan. Measurement and analysis of object reflectance spectra. *Color Research & Application*, 19(1):4–9, 1994.

[WB82] Gerhard West and Michael H Brill. Necessary and sufficient conditions for von kries chromatic adaptation to give color constancy. *Journal of Mathematical Biology*, 15(2):249–258, 1982.

[WB86] James A Worthey and Michael H Brill. Heuristic analysis of von kries color constancy. *JOSA A*, 3(10):1708–1712, 1986.

[WGG07] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-Based Color Constancy. *IEEE Transactions on Image Processing*, 16:2207–2214, 2007.

[WHL04] Chun-Hsin Wu, Jan-Ming Ho, and Der-Tsai Lee. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, 5(4):276–281, 2004.

[WN97] Andrew B Watson and Cynthia H Null. Digital images and human vision. 1997.

[XF06] Weihua Xiong and Brian Funt. Estimating illumination chromaticity via support vector regression. *Journal of Imaging Science and Technology*, 50(4):341–348, 2006.

[XF09] Weihua Xiong and Brian Funt. Stereo retinex. *Image and Vision Computing*, 27(1):178–188, 2009.

[XSF+07] Weihua Xiong, Lilong Shi, Brian Funt, Sung-Su Kim, Byoung-Ho Kang, Sung-Duk Lee, and Chang-Yeong Kim. Illumination estimation via thin-plate spline interpolation. In *Color and Imaging Conference*, volume 2007, pages 25–29. Society for Imaging Science and Technology, 2007.

[YCK02] Haiqin Yang, Laiwan Chan, and Irwin King. Support vector machine regression for volatile stock market prediction. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 391–396. Springer, 2002.

[Zek93] Semir Zeki. A vision of the brain. 1993.

[ZL10] Federico Zaraga and Giacomo Langfelder. White balance by tunable spectral responsivities. *JOSA A*, 27(1):31–39, 2010.

Appendix A

Publication

Liu, Bozhi, and Guoping Qiu. "A classification method for estimating the illuminant of an image." Visual Communications and Image Processing (VCIP), 2016. IEEE, 2016.

Liu, Bozhi, and Guoping Qiu. "Illuminant classification based on random forest." Machine Vision Applications (MVA), 2015 14th IAPR International Conference on. IEEE, 2015.