TWO-ILLUMINANT ESTIMATION AND USER-PREFERRED CORRECTION FOR IMAGE COLOR CONSTANCY

ABDELRAHMAN KAMEL SIDDEK ABDELHAMED

NATIONAL UNIVERSITY OF SINGAPORE

2016

TWO-ILLUMINANT ESTIMATION AND USER-PREFERRED CORRECTION FOR IMAGE COLOR CONSTANCY

ABDELRAHMAN KAMEL SIDDEK ABDELHAMED
(M.Sc., Assiut University, Egypt, 2014)

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

Master of Science

supervised by

Dr. MICHAEL S. BROWN

in
SCHOOL OF COMPUTING
NATIONAL UNIVERSITY OF SINGAPORE

SINGAPORE, 2016

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Abdelrahman Kamel
December 2016

To my parents and wife

Acknowledgements

Praise be to Allah, the Most Gracious, the Most Merciful. Prayer and peace be upon the Prophet Muhammad.

I would like to express my deepest gratitude and appreciation to my advisor *Prof. Michael S. Brown*, for his close guidance, great patience, endless support, and valuable advice throughout my whole study.

I would like to thank *Dr. Dongliang Cheng* for his continuous help throughout all stages of this work. I am thankful to *Dr. Scott Cohen*, *Dr. Brian Price*, and *Dr. Shengdong Zhao* for their valuable insights and remarks throughout this work.

Sincere thanks to Dr. Rang Nguyen, Hakki Karaimer, Dr. Li Yu, Hu Sixing, and all my lab mates at NUS School of Computing, for their help and valuable discussions. I would like to thank the whole family of School of Computing for accepting me as a student and for providing me with resources to finish this work.

Finally, I would like to express my gratitude to my family, for their patience and support, words are useless in expressing my gratitude to my wife, my kids, and my parents, to whom I simply owe everything.

Abdelrahman Kamel
2016

Abstract

Illuminant estimation and correction (white-balance) is a fundamental process in camera image processing pipelines. This thesis examines the problem of white-balance when a scene contains two illuminants. This is a two-step process: 1) estimate the two illuminants; and 2) correct the image. Existing methods addressing this problem attempt to estimate multiple illuminants to produce a spatially varying illumination map. However, their results are still error prone and the resulting illumination maps are too low-resolution to be used for proper spatially varying white-balance correction. In addition, the spatially varying nature of these methods make them computationally intensive.

Our approach is to show that this problem can be effectively addressed by not attempting to obtain a spatially varying illumination map, but instead by detecting and estimating two illuminants, namely indoor and outdoor illuminants, by performing single illuminant estimation on large sub-regions of the image. Our approach is able to detect when distinct illuminants are present in the image and accurately measure these illuminants. Since our proposed strategy is not suitable for spatially varying image correction, two user studies have been performed to see if there is a preference for how the image should be corrected when two illuminants are present, but only a global correction can be applied. The user studies show that when the illuminants are distinct, there is a preference for the outdoor illuminant to be corrected resulting in warmer final image. We use these collective findings to demonstrate an effective two illuminant estimation scheme that produces corrected images that users prefer.

Contents

List of Figures

List of Tables

List of Publications

Cheng, D.*, Kamel, A.*, Price, B., Cohen, S., and Brown, M. S. 2016. Two Illuminant Estimation and User Correction Preference. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 469–477. (*Equal Contribution)

Chapter 1

Introduction

In this chapter we are going to give an overview on the problem of white-balance. Then we will categorize the methods proposed to address this problem, present our motivation towards addressing this problem, and state our contributions achieved to resolve the problem.

1.1 White-Balance Problem

Color constancy is the ability to perceive the true colors or reflectance of objects despite the illumination falling on them. Our human visual system is equipped with a well-developed ability to perform this [McCann et al. 1976]. However, color constancy is a challenging problem facing computer vision systems that depend on colors as prominent features. Hence, computational color constancy needs to be applied to reduce the color cast caused by illumination on images and videos. Color constancy is often simplified to aim at correcting image colors such that white objects appear to be white, hence the notion of *white-balance*.

(a) 2550K (b) 3550K (c) 4550K

Figure 1.1: An example showing the effect of simulating different illuminations on an image. The values below images indicate illuminant temperatures in Kelvin. On the images, the RGB values of a white pixel are indicated, ideally the values should be such that $R = G = B$ if the image is correctly white-balanced.

Besides computer vision systems, white-balance is also required for image reproduction. Although the human viewer can compensate for the scene illumination, the illumination present in a photograph cannot be compensated. As a result, white-balance serves as the first step before image reproduction or enhancement in the camera image processing pipeline. Figure 1.1 shows an example of the same physical scene under different simulated colored illuminants. Next, we categorize the approaches to addressing the white-balance problem.

1.2 Categories of White-Balancing Methods

We can categorize white-balancing methods based on the number of estimated illuminants into three main categories as shown in Figure 1.2:

1. Single illuminant estimation methods: these methods lay the assumption that a scene is uniformly illuminated by a single illuminant and they try to estimate this illuminant and correct the image accordingly.
2. Multiple illuminant estimation methods: these methods assume that a scene is illuminated by more than one illuminant or even if there is only one illumi-

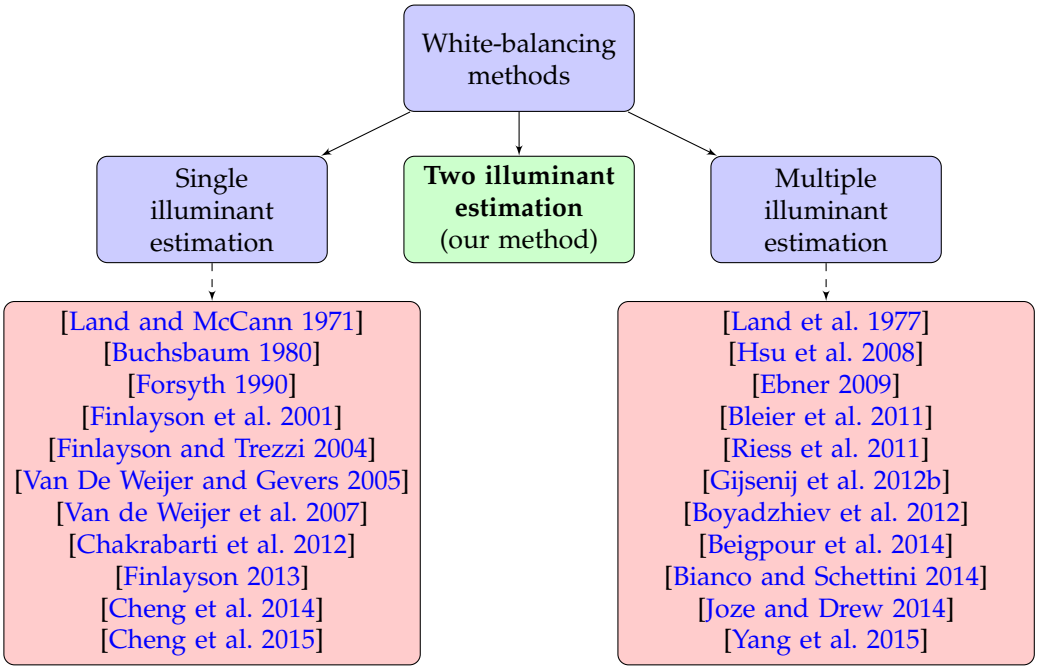


Figure 1.2: Categorization of color constancy methods showing the position of our approach, two illuminant estimation.

nant, it is not spatially uniform. Hence, these methods try to estimate either a number of illuminants or a spatially varying illuminant map, where in the extreme case there is one illuminant per pixel.

3. Two illuminant estimation method: this is our proposed method to address the white-balancing problem where we recommend estimating either one or two illuminants at most. More details come in Chapter 2.

1.3 Motivation

Most white-balance methods assume the imaged scene is uniformly illuminated by a single light source, these methods are listed under "single-illuminant estimation" in Figure 1.2, however, it is not uncommon for a scene to be illuminated

(a) RAW image. (b) Image corrected by outdoor illuminant. (c) Image corrected by indoor illuminant.

Figure 1.3: An example scene with two different illuminants (outdoor and indoor). The color of the original RAW image is biased by both illuminants. (b) and (c) show the image corrected by each of the illuminants respectively.

by more than one light as shown in Figure 1.3. That led to the proposal of many approaches that attempt to estimate multiple illuminants, these methods are listed under "multiple-illuminant estimation" in Figure 1.2. Such methods usually use a sliding window strategy or image segmentation to perform local illuminant estimation. This results in a spatially varying illumination map over the image. Such illumination maps are typically low-resolution (e.g. 15×20) and their effectiveness in subsequent white-balance correction is often not demonstrated. Moreover, these methods tend to be slow and require prior knowledge that the imaged scene contains more than one illuminant.

For the above drawbacks of multiple illuminant estimation methods, we advocated a different strategy for addressing the two illuminant estimation problem. Specifically, we found it more effective not to attempt to estimate a spatially varying illumination map. Instead, we will show that applying a single-illuminant estimation method on a relatively small number of large sub-images in the input image can not only detect if two distinct illuminants are present, but also provide accurate estimations for these illuminations.

1.4 Contributions

This thesis presents a set of contributions and findings towards the estimation and correction of images containing either one or two illuminants. First, an efficient method for accurately estimating one or two illuminants from a single image is proposed in Chapter 3. Second, two user studies were performed revealing that users do have a strong preference for a particular correction when two distinct illuminations are present in the image, this is discussed in Chapter 4. To the best of our knowledge, these are the first user studies eliciting user preferences for white-balancing. Third, we demonstrate how to combine findings from (first) and (second) into a framework for correcting images containing scenes with two illuminations, this is beneficial as there is still no clear demonstration of white-balance correction in the case of two or more illuminants, this is discussed in Section 4.3. Finally, most prior works use synthetically generated two-illuminant images as test cases. As part of our work, we provide a new image data set extracted from existing illumination and image processing data sets in which the ground truth for the two illuminants has been manually identified. This data set can be used in further evaluations of two-illuminant or multi-illuminant estimation algorithms. We believe the findings in this thesis will be beneficial in helping to develop further approaches for multi-illuminant estimation and subsequent image correction.

1.5 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 provides a fundamental background along with related work on the white-balancing problem. Our approach for two-illuminant estimation is discussed in Chapter 3 along with experimental results and discussion. In Chapter 4, we discuss our approach to global image correction along with experimental comparison with other methods. Finally, Chapter 5 concludes the thesis with a short discussion on possible future research directions.

Chapter 2

Background and Related Work

This chapter starts by providing a background on the white-balancing problem and its position in the camera image processing pipeline. Then, we will discuss the available existing approaches for single-illuminant and multiple-illuminant estimation and image correction.

2.1 Camera Image Processing Pipeline

Digital color cameras are tristimulus color systems, inspired by the human visual system. To simulate the effect of the human vision system, *e.g.*, outputting images which can be perceived by humans, it has an on-board process pipeline as shown in Figure 2.1. This pipeline is generic and may be adapted differently by various camera manufacturers. These various stages in the pipeline affect the final output image to different extent. However, the first and third stages, RAW image sensor response and white balancing, are the key to the topic of this thesis and we are going to explore them in more details.

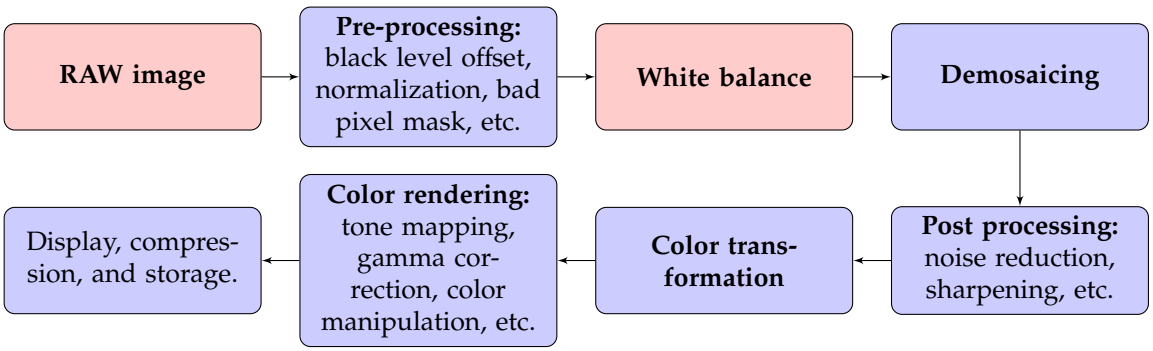


Figure 2.1: The generic steps applied onboard a camera, adapted from [Ramanath et al. 2005; Karaimer and Brown 2016]. Different camera manufacturer implementations can vary, however, most of these steps will be included and in a similar processing order.

2.1.1 RAW image formation

The light reflected on or emitted from the scene (*i.e.* scene radiance) goes through the camera lens, followed by the color filters and hits the cameras photosensors, causing *RAW sensor responses*. Generally, these color filters above the photosensors are composed from three different color filters: red, green and blue, thus resulting in an *RGB* tristimulus camera RAW responses. These color filters are generally arranged according to a particular pattern, named the *Bayer pattern*, where 50% of the filters are green filters, 25% are red and the other 25% are blue. Due to the presence of these color filters, only the response value of one color channel is recorded for each pixel. Therefore, a process called *demosaicing* must be applied to interpolate the other missing two values of each pixel from the neighboring pixels to generate a full color image. Without considering the effect of demosaicing, the physical formulation of RAW responses is similar to the tristimulus image

formation of the human retina [Wyszecki and Stiles 1982]:

$$\rho_i = \int_{\lambda \in \Omega} l(\lambda)\, s_i(\lambda)\, d\lambda, \qquad i \in \{R, G, B\}, \tag{2.1}$$

where $[\rho_R, \rho_G, \rho_B]^T$ are the camera RAW responses, $l(\lambda)$ is the spectral distribution of the light incident arriving on the photosensor, and $s_i(\lambda)$ is the effective sensitivity of the camera photosensors under the i^{th} type of color filter at wavelength λ.

2.1.2 Scene Illumination

Focusing on the RAW image sensor response, we see that there are two important physical factors that contribute to the image formation; (1) the varying surface reflectance of the objects existing in the scene, and (2) the illumination condition under which the scene is viewed. The term $l(\lambda)$ in Equation 2.1 is the result of the illuminant signal $e(\lambda)$ interacting with the surface being viewed. Ideally, it is a linear function of the incident light and the reflectance of the surface, as well as the direction of the illumination and the direction of the camera, which is expressed as the *bidirectional reflectance distribution function* (BRDF). However, the BRDF is a function of four geometric parameters, measuring the BRDF for even one surface is very tedious. It is clear that we need simpler models.

The simplest possible form of the BRDF is *a constant*. This corresponds to a perfectly diffuse reflection, also referred to as *Lambertian reflection*. A Lambertian reflector appears equally bright, regardless of the viewing direction. As a result, the interaction of surface, light and sensor can be elucidated as

$$\rho_i(\mathbf{x}) = \int_{\lambda \in \Omega} e(\lambda)\, r(\lambda, \mathbf{x})\, s_i(\lambda)\, d\lambda, \qquad i \in \{R, G, B\}, \tag{2.2}$$

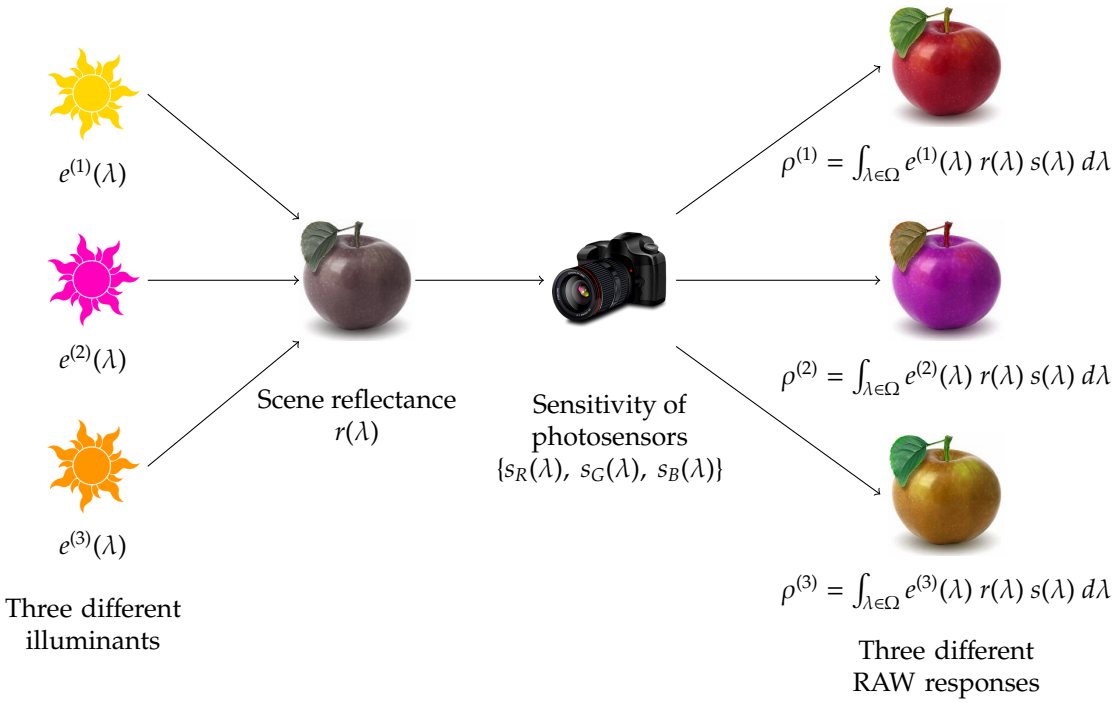


Figure 2.2: Illustration of Lambertian image formation model, adapted from [Cheng 2015].

where each RAW response ($R, G,$ *and* B) at pixel location $\mathbf{x}$ is an integrated signal resulting from the sensitivity of photosensors $s(\lambda)$, the scene reflectance $r(\lambda, \mathbf{x})$ and the scene illumination $e(\lambda)$ over the visible spectrum Ω. Figure 2.2 shows an illustration of this simple Lamertian image formation model. The Lambertian model here assumes that the scene is illuminated by one single light source uniformly as $e(\lambda)$ is constant for different pixel locations $\mathbf{x}$. The observed *color of the uniform illuminant* $\boldsymbol{\rho}^e$ depends on the spectral power distribution of the light source $e(\lambda)$ as well as the effective sensitivity of the camera photosensors $s(\lambda)$:

$$\rho_i^e = \int_{\lambda\in\Omega} e(\lambda)\ s_i(\lambda)\ d\lambda, \qquad i \in \{R, G, B\}, \tag{2.3}$$

2.2 White-Balance

As discussed in the previous sections, the camera's RAW sensor responses directly depend on the scene illumination. Achieving color constancy is significantly important to many computer vision applications as well as photo reproduction. Therefore, the goal of computational color constancy, especially white-balance, is to diminish the effect of the illumination to obtain data which more precisely reflects the physical content of the scene.

The aim of white balancing process is to correct images such that they appear as if taken under a canonical lighting condition, usually a D65 illuminant. This is a two-step process, (1) illuminant estimation, and (2) image correction, which we discuss in the following subsections.

2.2.1 Illuminant Estimation

The step of illuminant estimation is to infer the prevailing illumination on the imaged scene. It is the key to computational color constancy, as the next step, image correction, is considered to be straightforward.

If we consider the simplest case, where the illuminant is uniform and its color is depending only on the spectral power distribution of the light source and the camera sensitivity (Equation 2.3), it is still hard to solve the illuminant estimation problem. Suppose an image with N pixels is captured under a uniform illuminant, there will be $3N + 3$ unknowns (N surfaces at every pixel location and 1 global illuminant, each with 3 RGB channels), but only 3N RGB measurements are known. Even if we consider it unnecessary to recover the brightness of the light (magnitude of the RGB tristimulus), thus the number of unknowns is reduced to $3N+2$ and this

is still less than the number of known quantities. As such, illuminant estimation is an ill-posed problem that remains a challenge to solve. We dedicate the next section 2.3 for the review of illuminant estimation methods.

2.2.2 Image Correction

One common simple model of image correction for different illuminants is a single *linear transformation*, where each pixel value of the image taken under the unknown illuminant, $\boldsymbol{\rho}^U = [\rho_R^U, \rho_G^U, \rho_B^U]^T$, is mapped to the corresponding color as if the image was taken under the canonical illuminant, $\boldsymbol{\rho}^C = [\rho_R^C, \rho_G^C, \rho_B^C]^T$, by

$$\boldsymbol{\rho}^C = \boldsymbol{M}\boldsymbol{\rho}^U \tag{2.4}$$

where $\boldsymbol{M}$ is a single 3×3 matrix used for all pixels. It is clearly that this linear transformation is only an approximation as some information has been lost during the mathematical projection/integration from high dimensional space of spectral power distribution to a much lower dimensional RGB space as in Equation 2.2.

The $\boldsymbol{M}$ in Equation 2.4 can be further restricted to be a diagonal matrix. This approach is attributed to von-Kries [von Kries 1878] as a model for human eye adaptation and is thus often referred to as the *von-Kries diagonal model*, or *diagonal model* for short. This diagonal model maps the image taken under one illuminant to another by simply scaling each channel independently:

$$\rho^C(\mathbf{x}) = diag(\frac{\boldsymbol{\rho}^C}{\boldsymbol{\rho}^U})\, \rho^U(\mathbf{x}) \tag{2.5}$$

where $diag(\cdot)$ indicates the operator of creating a diagonal matrix from a vector.

This model has been used for most computational color constancy algorithms where the neutral colors should remain achromatic in the camera's color space. However, the ability of this diagonal matrix to correct non-neutral colors is ignored. Other models that address the correction of non-neutral colors are beyond the scope of this thesis, the reader may refer to [Cheng 2015] for further study. In the next section 2.3 we will review prior single-illuminant and multiple-illuminant estimation methods.

2.3 Illuminant Estimation Methods

In this section we focus our review on the key illuminant estimation methods in the literature. For the purpose of our work, we categorize them into two main categories: single-illuminant and multiple-illuminant estimation methods. Also, we make a quick discussion on the drawbacks of the multiple-illuminant estimation methods which we are going to overcome by our approach in Chapter 3. For more details and further study on illuminant estimation literature, the reader may refer to [Cheng 2015; Gijsenij et al. 2011].

2.3.1 Single-Illuminant Estimation Methods

Perhaps the simplest general approach to illuminant estimation is to compute a single statistic of the image, such as the average or mean color, and this led to the **Grey-world** assumption [Buchsbaum 1980]. In physical terms, the assumption is that the average reflectance in a scene under a neutral light source is achromatic, therefore any deviation from achromaticity in the average scene color is caused by the illuminant. This implies that the color of the light source ρ^e can be estimated

by computing the average color in the image. This method is extremely simple, however, it is very sensitive to large uniformly colored surfaces, which often leads to scenes where this assumption obviously fails.

To overcome the sensitivity to large uniformly colored surfaces, the average color could be computed among regions as opposed to pixels [Gershon et al. 1987] where the image is segmented before computing the scene average color among the segmented regions.

An important early work in color constancy is the **Retinex theory** [Land and McCann 1971]. The Retinex theory presumes that slowly spatially varying frequency in an image is related to the scene illumination. If the illumination is assumed to be uniform, then the Retinex theory amounts to the **White-patch** assumption, where the maximum response in the RGB-channels is caused by a perfect reflectance on a white surface. A surface with perfect reflection will reflect the full range of light that it captures. Consequently, the color of this perfect reflectance is exactly the color of the light source. In practice, the assumption of white-patch is alleviated by finding the maximum values of the color channels separately, especially after applying a smoothing step on the image [Shi and Funt 2012].

In [Finlayson and Trezzi 2004], the white-patch and the grey-world algorithms are shown to be special instances of a more general *Minkowski* framework:

$$\left(\int \rho_i^p(\mathbf{x})\, d(\mathbf{x}) \right)^{1/p} \qquad i \in \{R, G, B\} \tag{2.6}$$

where substituting $p = 1$ gives the grey-world assumption (average color), and when $p = \infty$ results in the white-patch assumption (maximum color). To obtain better performance, the value of p is tuned for the data set to reach the optimal

value. This is referred to as the **shades-of-gray**.

Instead of using color distribution, [Van De Weijer and Gevers 2005] incorporated high-order image spatial information based on the **Grey-edge** assumption; that the average of the reflectance differences in a scene (*i.e. image gradient*) is achromatic. A general computing framework can be formulated as:

$$\left(\int \left| \frac{\partial^n \rho_{i,\sigma}(\mathbf{x})}{\partial \mathbf{x}^n} \right|^p d(\mathbf{x}) \right)^{1/p} \qquad i \in \{R, G, B\} \tag{2.7}$$

where $|\cdot|$ indicates the Frobenius norm, p is the Minkowski norm, and derivatives of the image are defined as convolving the images with Gaussian derivative filters with scale parameter σ. Later, weighting schemes were applied on different types of edges, resulting in the **weighted grey-edge** method [Gijsenij et al. 2012a].

The **gamut-mapping** algorithm was introduced in [Forsyth 1990] based on an assumption that in real world natural images and for a particular illuminant, only a limited number of colors can be observed, called the *observable colors*. The set of observable colors under the canonical illuminant is called the *canonical gamut*. As a result, any variations in the observed colors of an image are caused by the presenting illuminant and this unknown illuminant can be found by mapping the sensor responses to the *canonical gamut*.

The discussed statistics-based methods have the clear advantage that they are *simple and fast*, but often they do not perform well. That is why recent state-of-the-art methods employ *learning-based* techniques to get better performance. Several methods [Gijsenij and Gevers 2007; Bianco et al. 2008; Bianco et al. 2010; Gijsenij and Gevers 2011], imitating the human vision system, have proposed to adopt learning of **semantic information** into the estimation of the illuminant, examples of such

information are indoor-outdoor classification, image categorization, or complex image features from decision trees. These methods tend to combine multiple illuminant estimation algorithms and use different strategies to fuse the algorithms or select the most suitable one. Other methods use high-level visual information as priors, such as colors of specific object categories [Van de Weijer et al. 2007; Rahtu et al. 2009] or, more reliably, colors of human faces [Bianco and Schettini 2012].

With large amounts of accessible images, artificial **neural networks** played a big rule in illuminant estimation [Funt et al. 1996; Cardei et al. 1998; Cardei et al. 2002], especially with kernel regression [Agarwal et al. 2006] and thin-plate spline information [Shi et al. 2011]. In [Joze and Drew 2012; Joze and Drew 2014], training data was used from surface regions segmented from the images and a K-nearest neighbor approach was used to estimate the final illuminant from multiple candidates. In [Finlayson 2013], the final illuminant estimate is proposed to be a direct linear mapping from image **statistical moments**.

Most of the discussed learning-based methods rely on complex features and have long evaluation and training times. Trying to come up with simplified and efficient algorithms for white balancing, Cheng et al. [Cheng et al. 2014] developed an illumination estimation method that chooses bright and dark pixels using a projection distance in the color distribution and then applies principal component analysis (PCA) to estimate the illumination direction only from these pixels. Recently, Cheng et al. [Cheng et al. 2015] presented a learning-based method based on four **simple color features** and use these features with an ensemble of **regression trees** to estimate the illumination.

2.3.2 Multiple-Illuminant Estimation Methods

One of the first methods to consider non-uniform or multiple illuminations is the **Retinex theory** [Land et al. 1977] that assumed illumination smoothly varies across a scene and abrupt changes in an image's content are caused by changes in scene reflectance properties. This assumption was used by [Ebner 2009] to propose a method that computes the **local average color** as the local scene illumination by convolving the image with a Gaussian or exponential kernel. This method can be interpreted as applying Grey-world [Buchsbaum 1980] locally at every pixel. While simple, this approach established a *common framework* adopted by many later methods: namely, divide an image into local patches or regions, apply single illuminant estimation methods locally, and post process the local results to obtain an *illumination map*.

Bleier et al. [Bleier et al. 2011] proposed to segment an image into **super-pixels** and then applied multiple single-illuminant estimation algorithms for each super-pixel. These per super-pixel estimations were fused to obtain the final local estimates. In a similar approach [Riess et al. 2011], an improved version of the physics-based illuminant estimation method [Tan et al. 2004] was applied on images segmented into **homogeneous regions**.

A general framework was proposed by [Gijsenij et al. 2012a] where they propose to use local image patches selected by different sampling methods (grid-based, keypoint-based, or segmentation-based sampling). After sampling each of the patches, single-illuminant estimation techniques are applied to obtain local illuminant estimates. These initial estimates are clustered into two groups and spatial filters are applied to smooth the illuminant distributions. Similarly, [Beigpour et al.

2014] formulated a multi-illuminant estimation within a *conditional random field* framework over a set of local illuminant estimates from single-illuminant estimation algorithms.

Although many works adapt this general framework, the results are often not satisfactory as they are bounded by the quality of the single-illuminant estimation methods being used. Such single-illumination methods tend to perform poorly on local regions. Another drawback is that these local methods are computationally intense. As a result, the spatial resolution used by these methods are lowered to reduce the computational time, leading to the result of illumination maps that are too coarse to be practical, *e.g.* approximately 30 super-pixels in [Bleier et al. 2011] and an illumination map of size 15×20 in [Beigpour et al. 2014]. In addition, most of these methods do not demonstrate how to use the estimated illumination map for image correction. The ability to perform good spatially varying illumination correction is still unclear.

There have been a number of *bottom-up* single-illuminant estimation methods that have been adopted to handle multi-illuminant images. The approaches by [Bianco and Schettini 2014] and [Joze and Drew 2014] respectively extended the face-based and exemplar-based color constancy algorithms to deal with a *known number* of multiple illuminants. Yang et al. [Yang et al. 2015] proposed to identify grey pixels to estimate single and multiple illuminants. For these methods, the type of image (single or multiple illuminant) must be given explicitly.

There are works that focus only on correcting scenes with multiple illumination with *user assistance*. Hsu et al. [Hsu et al. 2008] proposed treating two-illumination image correction as a mixture estimation problem using background-foreground matting where examples of illumination in the scene were provided

by **user markup**. Boyadzhiev et al. [Boyadzhiev et al. 2012] extended this matting approach to handle more illuminants with the addition of more user markup used to indicate neutral color, correct color, and homogenous scene regions.

As discussed in Chapter 1, the approach in this thesis takes a departure from all the reviewed methods in this chapter, where all of these methods are just variances of attempting to estimate multiple local-illuminants. This departure decision is made for a number of reasons. First, the nature of the aforementioned local approaches makes the algorithms computationally intensive for practical purposes, especially for smartphone cameras. More importantly, the resulting illumination estimations have not been shown to be sufficiently dense to support high-quality spatially varying illumination correction. As a result, it is felt that focusing on a computationally efficient method that can reveal only one or two illuminations in the scene, even without useful spatial information, is desirable as users likely have a preference of which illuminant they would prefer to be corrected, as will be shown in Chapters 3 and 4.

2.4 Summary

In this chapter we provided some background on the camera image processing pipeline and the white-balancing problem. And we reviewed key important single and multiple illuminant estimation methods, discussing the drawbacks of the latter category which we will address by our approach to solving the white-balancing problem in Chapters 3 and 4 .

Chapter 3

Two Illuminant Estimation

In this chapter we present our two illuminant estimation approach to address the problem of white-balancing. Departing from multiple illuminant estimation paradigm, we advocate a different strategy, we show that applying a single-illuminant estimation method on a relatively small number of large sub-images of the input image can not only detect if two distinct illuminants are present, but also provide accurate estimations for these illuminants. Then, we describe a new image data set, extracted from existing illumination and image processing data sets, which we used to evaluate our approach. Experimental comparisons with prior works and a discussion will be also presented in this chapter.

3.1 Two Illuminant Estimation Method

The main idea is to use a single illuminant estimation method on a number of large sub-images of an image to obtain several candidate illuminant estimations. If these candidate estimations show *little variation*, it is assumed the image contains a single

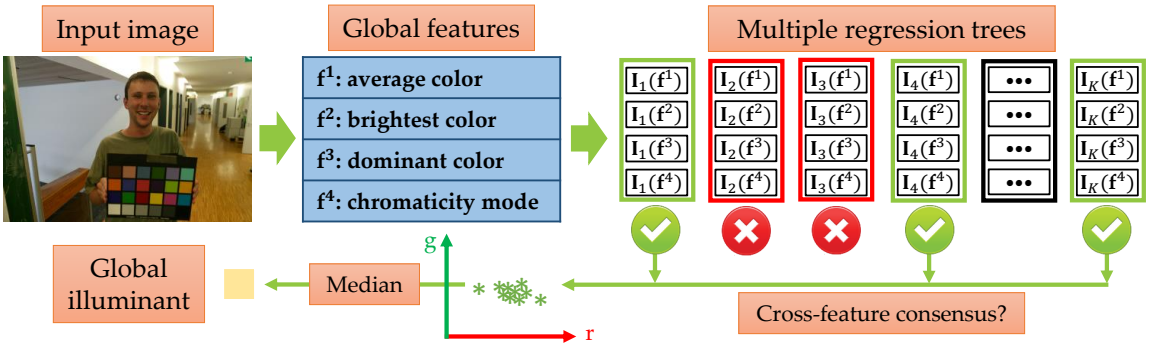


Figure 3.1: This figure provides an illustration of the regression trees method proposed by Cheng et al. [Cheng et al. 2015]. This method produces a set of reliable candidate estimates in the 2*D* rg-chromaticity space. The median of the candidates is used as the final estimate.

illuminant. Conversely, if the candidate estimations show *large variation*, it is likely there are two distinct illuminants among the candidates that can be extracted.

The accuracy of this strategy depends highly on which single illumination method is used. When we were determining the most suitable method, it was desirable to have a method that is not only *fast and accurate*, but also provides the ability to determine if a candidate estimation for a sub-image was *reliable* or not. To this end, we decided to use the recent work by [Cheng et al. 2015]. As will be discussed, not only is this method fast, but its design based on multiple classifiers provides a suitable mechanism to determine if a candidate estimation is reliable or not. We first provide a brief overview of the method by [Cheng et al. 2015] and then we describe the full procedure next.

Single Illuminant Estimation using Simple Features. Figure 3.1 illustrates the method proposed in [Cheng et al. 2015] that used simple color features and regression trees. Given a RAW image, four features from the camera-specific RGB color distribution are extracted: $\mathbf{f}^1$: average color; $\mathbf{f}^2$: brightest color; $\mathbf{f}^3$: dominant color

and $\mathbf{f^4}$: chromaticity mode of the color palette. These four features are supplied to a bank of K regression trees ($K = 30$) to get illuminant candidate estimates. Each regression tree is indicated by $\mathbf{I_i}(\mathbf{f^j})$, where $i \in \{1, ..., 30\}$ indicates the index of the tree and $j \in \{1, ..., 4\}$ indicates the feature.

These regression trees are trained using labeled images with known illuminations from a single camera. Given a new input image, the four features are computed on the input and the features are evaluated on the 30 regression trees to produce illuminant estimation candidates in the $2D$ rg-chromaticity space[1]. Note that each tree produces four candidate estimations, one for each feature. A *cross-feature consensus* is used to identify potential candidates per tree. In particular, when any three out of four results for a particular tree are sufficiently similar, these results are kept, otherwise they are rejected. The final estimate for the entire method is the *median* of all kept estimates from the 30 trees.

As noted in [Cheng et al. 2015], there are cases when all of the estimates are rejected. There are also cases when the results that were kept have a great deal of discrepancy. In these cases, [Cheng et al. 2015] uses the median of all the 30 trees as the final output. However, in our case, we can use these scenarios to reject this result due to not being reliable for the current sub-image.

Two Illuminant Estimation Method. The overall framework of our method is illustrated in Figure 3.2. The image is divided into sub-images (*e.g.* 4×6, the effect of different sized sub-images is discussed later). For each sub-image, the multiple regression trees [Cheng et al. 2015] method just described is applied. Cross-feature

[1]The $2D$ rg-chromaticity space is a simple projection from the normalized $3D$ *RGB* space obtained by dividing the three values of R, G, and B components by the sum of their values, then using only the resulting r and g components to represent the illuminant on a $2D$ plane. Note that the b value can be easily restored as $(1 - r - g)$ [Martinkauppi and Pietikäinen 2005].

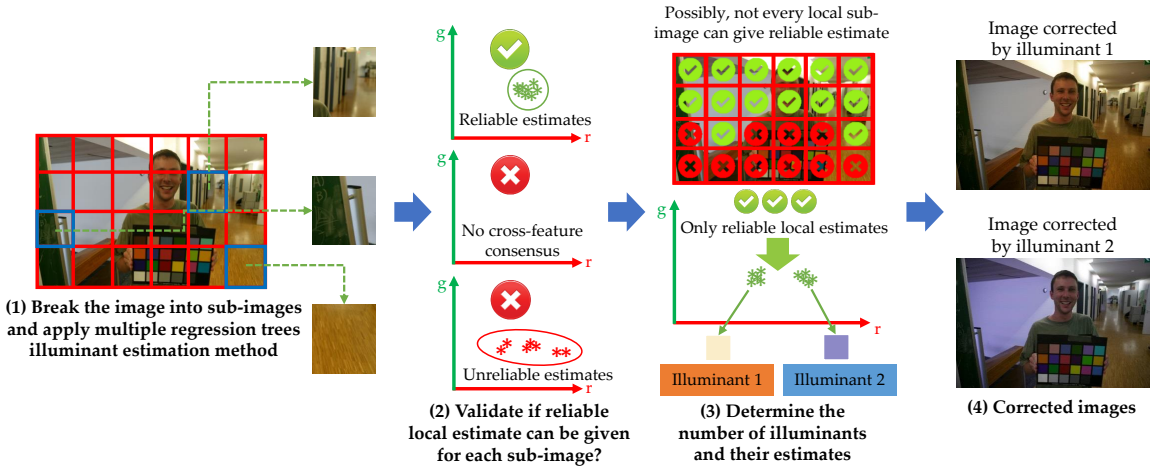


Figure 3.2: An overview of our two-illuminant estimation method. The image is divided into sub-images. A single-illuminant estimation method ([Cheng et al. 2015]) is applied on each sub-image. If the illuminant estimate candidates obtained per sub-image are similar, the estimate result is kept (denoted with a ✓, otherwise they are rejected denoted with an ×). The final set of reliable estimations (*i.e.* those kept) are examined to see if they form one or two clusters, which are used as the final illuminant estimations.

consensus is examined on these initial candidates and only candidates in agreement are kept. If the regression tree approach does not obtain a consensus or the collective candidates from the trees have too high variance (set to 0.0001 in rg-chromaticity space in our approach), the results for this sub-image are ignored, otherwise the median of the results is kept as the estimate for that sub-image. Figure 3.2 shows an example, where rejected sub-images are marked with an × and those that have passed are marked with a ✓.

After the sub-images have been processed, we are left with a set of $2D$ illuminant estimates in the rg-chromaticity space of the input image. We then compute the pair-wise distance of all candidate estimates. If the *average pair-wise distance* is less than 0.025, it is assumed there is only a single illumination in the scene and the median of all the candidates is reported as the illuminant estimation. Otherwise,

the image is classified as having two illuminations, and k-means ($k = 2$) clustering is applied and the centroids of the two clusters are taken as the estimates of the two illuminants.

Efficiency. The efficiency of the proposed approach is highly dependent on the underlying single-illuminant estimation algorithm being used. Most of processing time is consumed in applying the single-illuminant estimation algorithm to the sub-images. Then, an almost fixed amount of time is consumed in determining whether we have a single or double illuminants, which is a simple k-means clustering step. Hence, we used the single-illuminant estimation algorithm by [Cheng et al. 2015] which has been evaluated against many state-of-the-art methods and proved to be the most efficient. This supports our claim that our proposed approach is highly efficient.

3.2 Two-Illuminant Data Set

In this section, we describe how we obtained the images with two illuminations. Interestingly, we found a large number of such images in the **Gehler-Shi** data set [Gehler et al. 2008; Shi and Funt 2010], which is a data set intended and often used for single illuminant estimation. It has been noted by others (*e.g.* [Joze and Drew 2014]) that many of the images in fact contain two illuminations. We identified 66 of the 568 images from the Gehler-Shi data set as having two illuminants. Almost all of these images contain distinct illuminations of indoor and outdoor light. The original ground truth was measured by the neutral patches on the color checker chart, and it is typically positioned such that it measures the indoor illuminant. For the ground truth of the other illuminant, we manually marked it from

the image by finding neutral objects in the scene. While our manual marking is arguably not as accurate as having a color checker chart, we believe it provides a sufficiently accurate ground truth for studying this problem.

To enlarge our two-illuminant evaluation data set, we included some images from the **RAISE** data set [Dang-Nguyen et al. 2015]. This data set contains a large number of RAW images from various cameras that are used for image forensics. We examined this data set and found 34 images that are clearly containing two illuminants, mainly indoor and outdoor. The RAISE data set is not intended for illuminant estimation evaluation, so its images do not contain a color chart to get the ground truth illuminant. So, for this set of images, we estimated the two illuminants by manually selecting a small patch from each image that contains neutral material under the different illuminants.

Figures 3.3 and Figure 3.4 show some examples from the two-illuminant images from the Gehler-Shi data set and the RAISE data set, respectively. It is worth noting that the double-illuminant images targeted by the proposed method are quite common, an immediate indication to this is that the Gehler-Shi data set which is intended to contain only single-illuminant images and used for evaluating single-illuminant estimation algorithms, surprisingly, it contains 66 double-illuminant images out of 568 (about 12%) which were oversighted by many researchers. The following sections in this chapter provide details on our experimental results and findings.

Figure 3.3: Example images from our collection of two illumination images from Gehler-Shi data set [Gehler et al. 2008; Shi and Funt 2010].

Figure 3.4: Example images from our collection of two illumination images from RAISE data set [Dang-Nguyen et al. 2015].

3.3 Experimental Results

In this section, we evaluate the performance of the proposed illuminant estimation method. First, we describe the evaluation setup. Then, we describe the assessment criteria. Then, we discuss some parameter setting and method adaptation details. Finally, we discuss the performance results.

Evaluation Setup. We evaluate our method using the regression trees method by [Cheng et al. 2015] and with alternative designs using different single illuminant

estimation methods, namely: Grey-world [Buchsbaum 1980] and the learning-based Corrected-Moment method [Finlayson 2013]. We denote these two methods as *"Locally applied Grey-world"* and *"Locally applied Corrected Moment"*, respectively. We also modify an existing multi-illuminant method by Gijsenij et al. [Gijsenij et al. 2012b] to fit our framework for comparison, denoting it as *"Adapted Gijsenij et al."*. Our evaluation is performed on images containing both two illuminations and those with single illumination.

We note that Cheng et al. [Cheng et al. 2015] and the Corrected-Moment method [Finlayson 2013] require training. Since our proposed framework uses sub-images, we train these methods on sub-images instead of full images. For each method, we train using images from the Gehler-Shi data set that only contain a single illumination. This gives us the ground truth illuminant for every sub-image. For each training image, we randomly sample 40 sub-images from the original image for training. To evaluate the whole data set, we follow the standard three-fold cross validation procedure.

Assessment Criteria. Given an image to detect if it contains a single or double illuminants and to measure them, there are four possible outcomes:

- Case 1: If the input image is a single-illumination and is detected correctly as a single-illumination image (denoted as *single-single* or *SS*), only one estimate is given and there is only one angular error with respect to the ground truth.
- Case 2: For an image containing two illuminations, if it is correctly detected as having two illuminants (denoted as *double-double* or *DD*), we sort the two illuminant estimates according to their temperature and compare with the ground truth respectively: illuminant 1 represents the outdoor and illuminant

2 represents the indoor illuminant.

- Case 3 and 4: For images detected incorrectly; a single-illuminant image detected as two-illuminant (denoted as *single-double* or *SD*) or a two-illuminant image detected as single-illuminant (denoted as *double-single* or *DS*), we test to see if the method computed one illuminant estimate correctly, that is why we report the minimal angular error and maximal angular error.

We use N with a subscript of these four cases to represent the number of each outcome, *e.g.* N_{SS} indicates the number of single-illuminant images that are that have been detected correctly as single-illuminant images. Recall (R), Precision (P), and F-Measure (F) are common metrics for classification tasks and in the context of single illumination image detection problem, they can be defined as:

$$P_S = \frac{N_{SS}}{N_{SS} + N_{DS}} \tag{3.1}$$

$$R_S = \frac{N_{SS}}{N_{SS} + N_{SD}} \tag{3.2}$$

$$F_S = 2 \cdot \frac{P_S \cdot R_S}{P_S + R_S}, \tag{3.3}$$

while in the context of the double illumination image detection, they can be defined as:

$$P_D = \frac{N_{DD}}{N_{DD} + N_{SD}} \tag{3.4}$$

$$R_D = \frac{N_{DD}}{N_{DD} + N_{DS}} \tag{3.5}$$

$$F_D = 2 \cdot \frac{P_D \cdot R_D}{P_D + R_D}. \tag{3.6}$$

Parameter Setting and Method Adaptation. We evaluate our method and other adapted methods by breaking an image into different *numbers of sub-images*: 4×6 and 8×12. We choose this ratio $(2 : 3)$ for the dimensions of the sub-images grid because all images in the data set have dimensions with the same ratio. The chosen number of sub-images is related to the reason why we advocate the global correction. As discussed earlier, existing multi-illuminant color constancy methods often only give a low-resolution illuminant map that are not practical for spatially-varying correction. Thus, we proposed to estimate the local illuminants on large sub-regions. Using larger sub-image sizes, *e.g.* 2×3, would be too coarse for the two-illuminant scenes, thus we chose to start from 4×6 in the experiment. And from the comparison we will show that 8×12 doesn't improve the performance and 4×6 is generally enough. That is why we did not go through all possible sub-image sizes, from one pixel to the whole image.

We have evaluated the classification performance of our proposed method with different values of the threshold on the *average pair-wise distance* of all sub-image candidate illuminant estimates. These values range from $0.01 - 0.1$, at intervals of 0.001. Figure 3.5 shows the change in Precision (P), Recall (R) and F-Measure (F) with the different threshold values. As it can be seen, for single illuminant images, these three metrics are quite stable, while for the double illuminant image, although the precision gets higher with a higher threshold, the recall gets lower. With the threshold of 0.02, the F-measure is at its largest which means it gives the best balanced classification between single and double illuminant images. The cross-feature candidates consensus threshold in [Cheng et al. 2015] was 0.025, which is very close to 0.02, thus we used this value (0.025) as the final parameter value to calculate our final results.

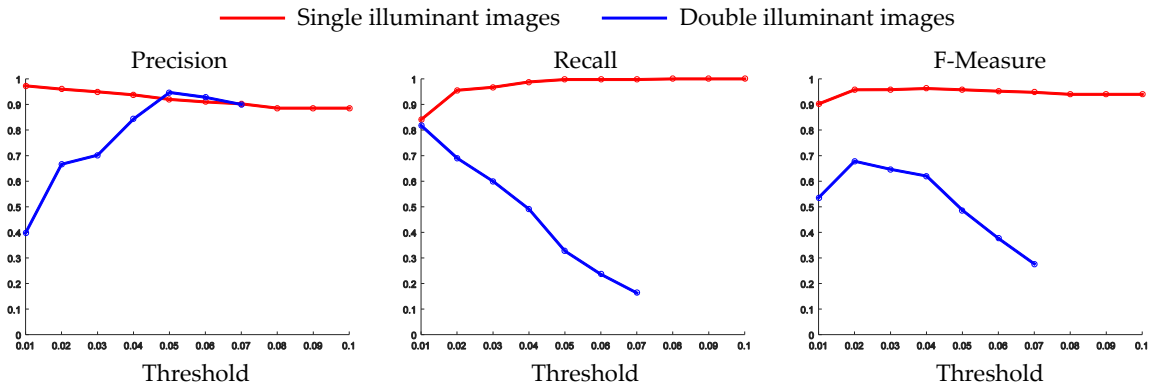


Figure 3.5: Precision, Recall, and F-Measure curves for our proposed method. Curves are shown for single and double illuminant images with varying values used as threshold on the average pair-wise distance of all candidate illuminant estimates.

It is worth noting that [Buchsbaum 1980] and [Finlayson 2013] have no mechanism to reject outliers, so the result from all sub-images are used to compute the final result. Finally, the method by [Gijsenij et al. 2012b] estimates the results on small local windows. We use all of these local window results and apply the pair-wise test, as described in Section 3.1, to determine the final illuminant estimates.

Experimental Results and Discussion To compare the general performance of our proposed method and other methods while achieving their best performance, different values of the *average pair-wise distance* of all sub-image candidates were used to determine the number of illuminants. The Precision, Recall and F-Measure curves are shown in Figure 3.6 using the values of threshold from $0.01 - 0.1$ at intervals of 0.001. Note that, for some large values of the threshold, our proposed method and the locally applied corrected moment method [Finlayson 2013] detect all images as single illuminant images, which makes the Recall and Precision both result in zero and the F-Measure undefined. Results for such threshold values are not shown in the figures.

As can be seen from the figures, the proposed method constantly achieves the best Recall performance for single illuminant images and also achieved constantly the best Precision performance for double illuminant images. In terms of a combined performance metric, the F-Measure for the proposed method is constantly better than the other methods for both single illuminant images and double illuminant images. The results show that the proposed method gives the best balanced classification for single/double illuminant images for all the threshold values.

Table 3.1 reports the performance results of our method compared to other state-of-the-art methods. All methods are reported with the threshold value that gives the best double illuminant images F-Measure (0.025 for the proposed method, 0.03 for locally applied Corrected-Moment [Finlayson 2013], 0.04 for locally applied Grey-world [Buchsbaum 1980] and 0.05 for adapted Gijsenij et al. [Gijsenij et al. 2012b]) as per Figure 3.6c.

As shown in the Table 3.1 the proposed method achieves the highest F-Measure for both single and double illumination images. When we apply our proposed method in a finer scale (2^{nd} row), however, it does not improve the performance. The chance of correctly detecting the two-illumination image may increase, but it drops quickly for single illumination images; thus the F-measure gets slightly worse, as can be seen in Figure 3.6c. Angular error of the illuminant estimates is also worse for smaller sub-images.

Compared with our proposed method using multiple regression illuminant estimation [Cheng et al. 2015], it is not surprising that the local Grey-world and Corrected Moment methods tend to mis-classify the number of illuminants, especially the single illumination images. However, we can see that the learning-based method (Corrected-Moment) gives better illuminant estimates than the statistical

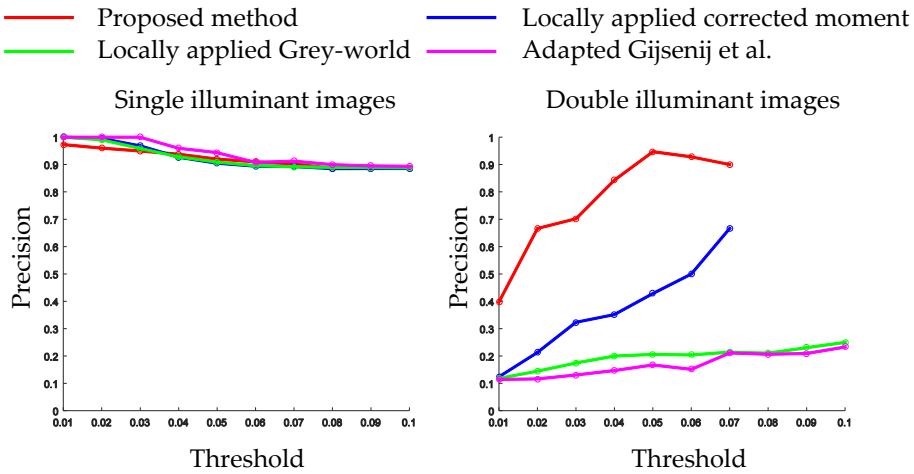
Table 3.1: Performance results of our method compared to state-of-the-art methods: Grey-world [Buchsbaum 1980], Corrected-Moment [Finlayson 2013], and Adapted Gijsenij et al. [Gijsenij et al. 2012b]. L1 and L2 are the two illuminants.

Method	Image Type	Total #	Detected #	F-Measure	Error for Correct Detections		Error for Incorrect Detections	
					L1	L2	Min Error	Max Error
Proposed (4×6)	Single	502	477	**0.9578**	**1.34**	-	**1.79**	**11.28**
	Double	66	49	**0.7000**	**1.87**	**2.05**	**2.33**	15.48
Proposed (8×12)	Single	502	464	0.9450	1.53	-	2.35	14.77
	Double	66	50	0.6494	2.32	2.33	2.51	16.48
Locally applied Grey-world	Single	502	352	0.8000	4.60	-	4.31	14.27
	Double	66	40	0.3125	8.89	6.04	4.22	13.29
Locally applied Corrected Moment	Single	502	393	0.8656	1.88	-	2.76	14.16
	Double	66	53	0.4649	3.83	7.64	2.87	**10.32**
Adapted Gijsenij et al.	Single	502	256	0.6615	4.92	-	4.57	18.03
	Double	66	50	0.2762	9.23	7.44	4.94	13.56

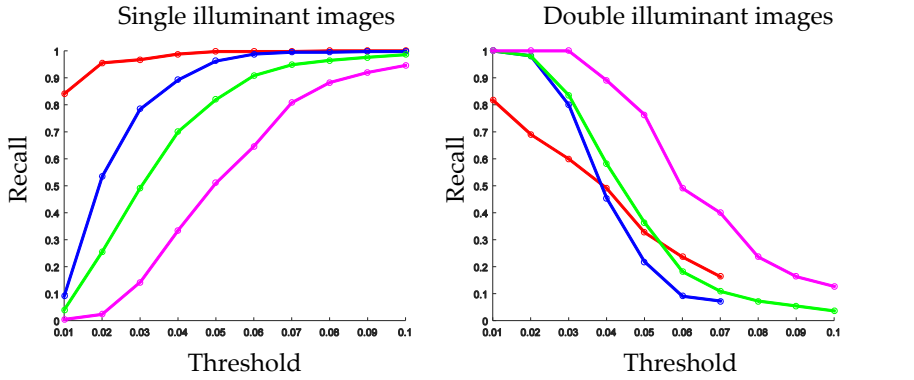
method (Grey-world). In contrast to our proposed method of estimating the illumination on big sub-images, the traditional spatially varying illuminant map in [Gijsenij et al. 2012b] obtains the worst result on almost every metric.

3.4 Summary

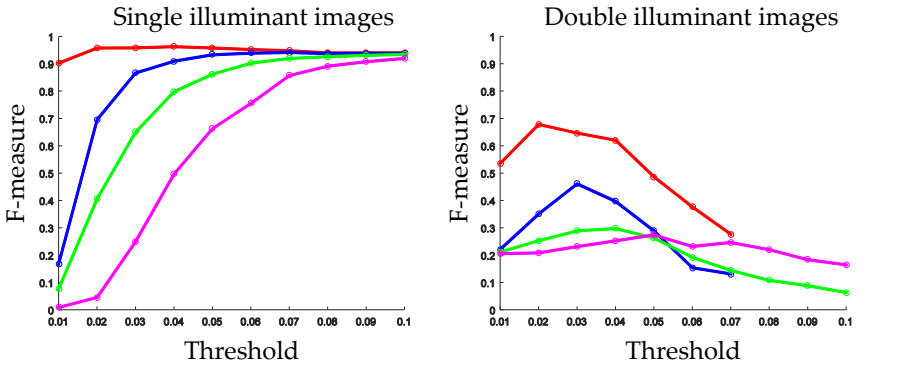
In this chapter we described our two illuminant estimation approach to address the problem of white-balancing. We showed that applying a single-illuminant estimation method on a relatively small number of large sub-images of the input image can not only detect if two distinct illuminants are present, but also provide accurate estimations for these illuminations. We showed with experimental results the better performance of our method compared to other state-of-the-art multiple illuminant estimation methods as well as other single-illuminant estimation methods adapted to fit our two-illuminant estimation framework.



(a) Precision curves



(b) Recall curves



(c) F-Measure curves

Figure 3.6: Precison, Recall, and F-Measure curves for single and double illuminant images for the four different methods evaluated with the same set of threshold parameter values.

Chapter 4

User-Preferred Image Correction

After the estimation of scene illumination in an image, the next step is to correct the image. In this chapter we will present our approach for global image correction and how we have come to it through conducting user studies. Also, we will present how to combine the two steps from this chapter and Chapter 3 for efficient white-balancing application.

As demonstrated in Chapter 3, the proposed approach is able to estimate two illuminants that are sufficiently distinct; however, there does not exist a corresponding illumination map and thus spatially varying white-balance correction is not possible. As such, we seek to determine, given two illuminants, which illuminant do users prefer to be corrected. To answer this, we carried out two user studies to elicit users' preferences. For the user studies we used images from two publicly available data sets, Gehler-Shi data set [Gehler et al. 2008; Shi and Funt 2010] and RAISE data set [Dang-Nguyen et al. 2015]. Details about our collected two illuminant images were discussed in Section 3.2. The following sections detail our experiments and findings.

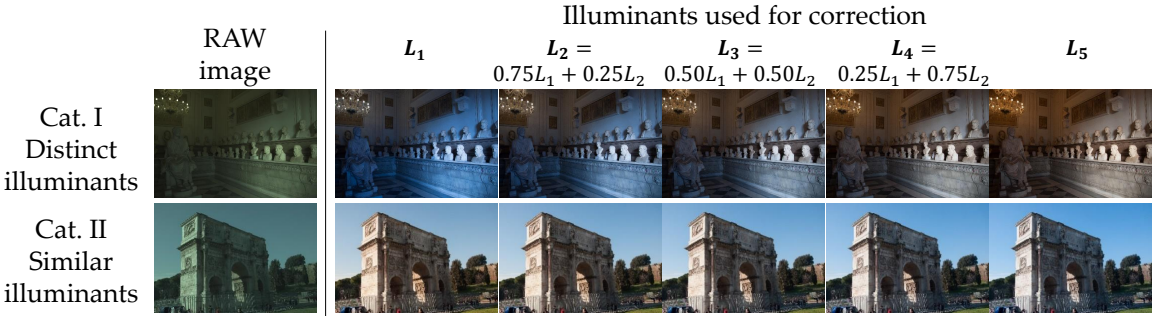
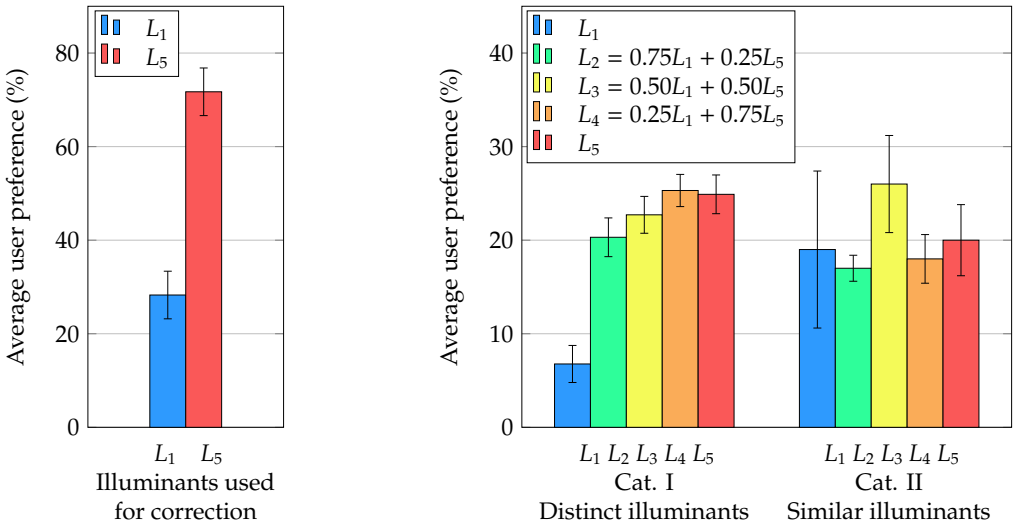


Figure 4.1: An example of image categories with 5 different illuminant corrections. The two rows represent images with two *distinct* indoor and outdoor illuminants (*Cat. I*) and images having two *similar* illuminants (*Cat. II*), such as sun and shade illuminants. The first column shows the raw image with the following columns showing the image corrected using 100%–0%, 75%–25%, 50%–50%, 25%–75%, and 0%–100% weights of the identified two illuminants, respectively.

4.1 User Study 1 (Two Choices)

For this study we used 33 images that contain two distinct illuminations (namely indoor and outdoor). The number of participants in this study was 39. We carried out the experiments in an indoor room with standard fluorescent light and calibrated monitors, that is to avoid any effects from outdoor lights, or difference in indoor lights, on the visual appearance of the viewed images. Also, the monitor calibration process is necessary because uncalibrated monitors are likely to change the visual appearance of the white-balanced images.

Procedure. For each image, the two illuminants L_1 and L_5 were estimated by manually selecting a small patch from each image that contains neutral material under the different illuminations (the second illuminant is termed L_5 in this study as we will use more in-between illuminants in the next study). Each image was corrected (white-balanced) using the two estimated illuminants, generating a pair of differently white-balanced images. In Figure 4.1, the second and last columns



(a) User study 1 (b) User study 2

Figure 4.2: User preferences for image correction by (a) two distinct illuminants (user study 1) and (b) five different illuminants (user study 2). The second user study is carried out with images containing distinct as well as similar illuminants.

show sample images corrected using the two illuminants. Each user was shown the 33 pairs of differently white-balanced images in random order. They were asked to choose the image they prefer. The images were viewed on the same screen and in the same place to avoid the effects of different lighting conditions on the visual appearance of the images.

Outcomes. The choices of the users were averaged. The results showed a higher user preference (almost 75%) of images corrected using illuminant L_5, which was the outdoor illumination. This is shown in Figure 4.2a. This means that users preferred the outdoor color casts to be corrected, which results in the indoor color casts in the image being kept. This has the effect of producing a "warm" (reddish) output image. We also performed statistical testing over the user choices to make sure they are statistically significant, which resulted in the 95% confidence interval

shown as error bars in Figure 4.2a.

4.2 User Study 2 (Five Choices)

The first user study only gave the users two choices. In that test, the results correcting one of the illuminants was strongly preferred. For the next user study, we wanted to see if the users would prefer some mixture of the results. We used the same images as user study 1, but also added some extra images that contain similar illuminants. We added these to see if we observed a similar preference trend when the images did not have distinct illuminations. This means we have two categories of images: *Cat. I*, two distinct illuminants (indoor and outdoor), and *Cat. II*, two similar illuminants, such as sun and shade illuminants.

Procedure. Images were sought to have sufficient neutral materials in the scene that we could accurately identify the two illuminations. In the end, we obtained 24 images from *Cat. I*, and 5 images from *Cat. II*. Since our main concern was two distinct illumination images (*Cat. I*), we selected more images from this category. We enlisted 34 users for the study, their average age was 22 years, with 26 males and 8 females.

For each image, two illuminants were estimated by manually selecting a small patch that contains neutral material to provide an estimation of the illumination. We label these two illuminants L_1 and L_5. We then generated mixtures of these two illuminations. Specifically, illuminant values for labels L_1-L_5 are computed using:

$$L_i = \alpha_i L_1 + (1 - \alpha_i) L_5, \tag{4.1}$$

where α_i is set to 1.0, 0.75, 0.50, 0.25, and 0.0 for L_1, L_2, L_3, L_4, and L_5 respectively. Each image I_k was corrected (white-balanced) using the 5 different illuminants. This results in five white-balanced images $\{I_k^{(L_1)}, ..., I_k^{(L_5)}\}$, for each image k, where $I_k^{(L_i)}$ means the correction of image I_k using illuminant L_i. Figure 4.1 shows some example images from both categories.

We used a two-alternative forced choice approach within a game-based strategy as recommended by [Hacker and von Ahn 2009]. A two-player game is used where both players are shown 50 randomly-selected pairs of images $\{I_k^{(L_i)}, I_k^{(L_j)}\}$ at the same time, where $i, j \in \{1, ..., 5\}$ and $i \neq j$. In other words, each pair are the same image corrected using two different illuminants picked randomly from the 5 illuminants for this image. Each pair is viewed in random order. Instead of asking each player to choose the image they prefer, each player is asked to select the image they think their partner (the other player) would prefer. This game-based strategy has been shown to be more effective in eliciting user preferences from such studies [Hacker and von Ahn 2009]. The same pair of images appear at least 4 times through the whole user study. The total number of image pair comparisons was 1700, where each of the 5 images appears for comparison at least 16 times.

As there are 5 different illuminant corrections for each image, and these corrections are shown to the users in pairs, the total number of comparisons needed to cover all 5 images in a pair-wise manner is $\binom{5}{2} = 10$ comparisons. To combine the relative user choices into an overall score that represents the user preference for each of the five corrected versions of the same image, we count the number of times each corrected version of the image $I_k^{(L_i)}$ is preferred over any other corrected version of the image $I_k^{(L_j)}$, then normalize it by the total number of pair-wise comparisons for this image. It is worth noting that in all user studies, we used color

calibrated monitors [DataColor 2016] under the same lighting conditions to avoid environmental biases.

Outcomes. The average user choices of each of the 5 illuminant corrections for each category of images are shown in Figure 4.2b along with their 95% confidence intervals represented by error bars. From this result, we see that *Cat. I* (two distinct illuminations) have a clear preference leaning towards the correction using a higher weight of the outdoor illuminant (*i.e.* the $L_4 = 0.25L_1 + 0.75L_5$ illuminant). For *Cat. II* (two similar illuminations) the preference is less pronounced and slightly favors an average result. This is consistent with the finding in [Finlayson et al. 2005] that visual difference illuminant corrections within 3° is not noticeable.

4.3 Two-Illuminant Estimation Application

Our combined findings in Chapter 3 and the previous two sections point to an approach for handling images that potentially contain two illuminations. Naming, run the algorithm in Section 3.1 to determine if two distinct illuminants exist. If so, correct the image with a 75%-25% mixture weighting the outdoor correction more. Figure 4.4 shows some examples for two-illumination images. To have a comparison, the Corrected Moment [Finlayson 2013] and weighted Grey-edge [Gijsenij et al. 2012a] methods were used to represent single illuminant estimation methods. We can see that for these images, the Corrected Moment method and the weighted Grey-edge method tend to give the indoor illuminant estimation or the mixture of indoor/outdoor. These illuminant estimates make the corrected images bluish in nature. In contrast, our correction results are close to the user preferred correction.

Failure Cases. Figure 4.3 shows some failure cases for our approach. There

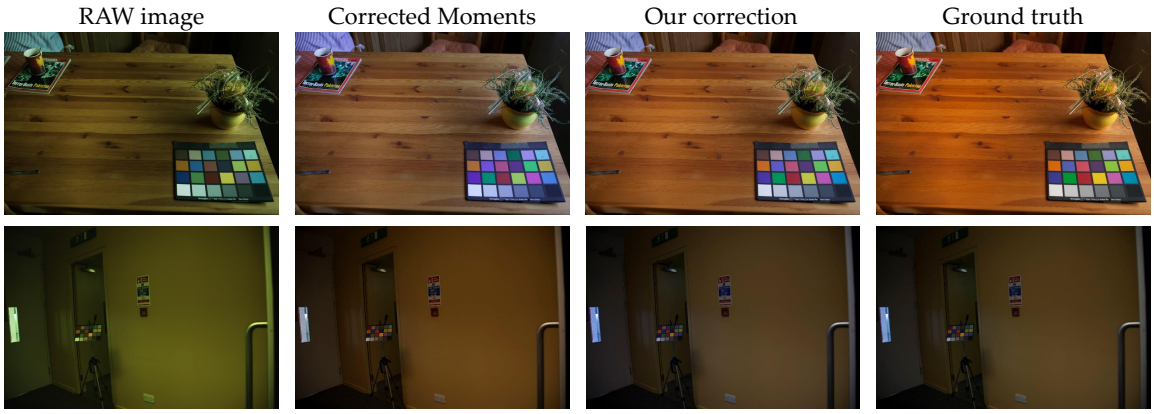


Figure 4.3: Example images that our method fails to correctly determine the number of illuminant(s). The first row shows example images that have a single illuminant, but our method estimated two illuminations. The second row shows images having two illuminants, but our proposed method can only detect one.

are two types of failure cases: single illumination image detected as multiple illumination image (first row) and the multiple illumination image detected as single illumination (second row). For the first case, this is usually because the image contains a large homogeneous region, making it hard to estimate the illuminant. For these images, the state-of-art single illuminant estimation methods also tend to fail. Although the illuminant classification is wrong, our method can still detect one of the illuminants correctly. Thus the image correction is still biased towards this illuminant. The second type of failure occurs when the image contains two illuminants, but where one illuminant is significantly more prominent. For these images, our method often estimates the dominant illumination.

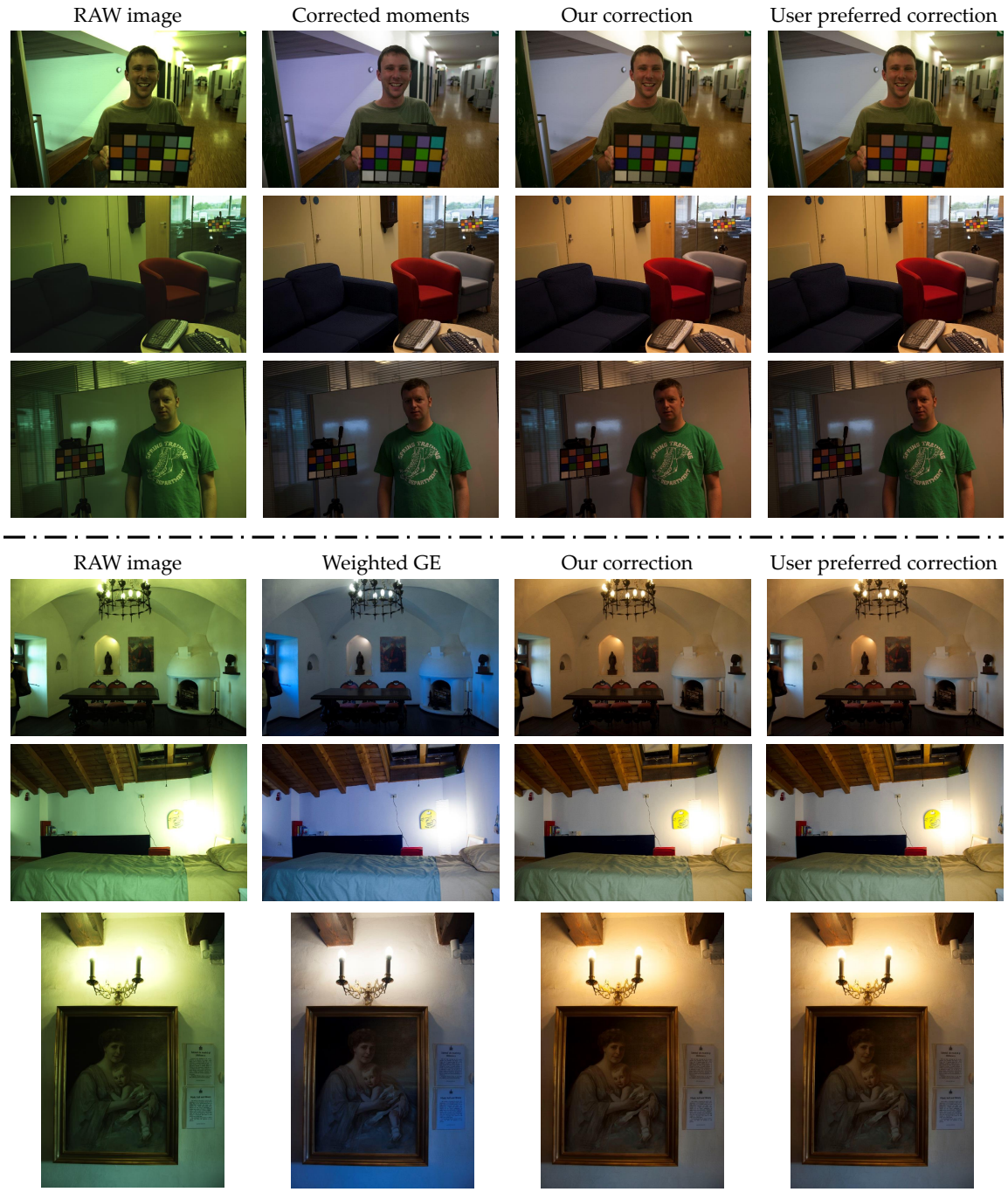


Figure 4.4: Visual comparison of image global correction. Top three images are from the Gehler-Shi data set [Gehler et al. 2008; Shi and Funt 2010] while the bottom three are from the RAISE data set [Dang-Nguyen et al. 2015]. For the images from the Gehler-Shi data set, the Corrected moment [Finlayson 2013] result is compared and for the images from the RAISE data set, the weighted Grey-edge [Gijsenij et al. 2012a] result is compared.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we proposed a simple and fast algorithm to determine if there is one or two illuminants in an imaged scene, and to estimate these illuminants. The key to our method is to use a high quality single illumination estimation algorithm on large sub-images for robust estimation instead of small patches as in previous methods. We only use the spatially varying estimates to decide if there are one or two illuminants and to globally estimate the illuminants present in the scene. Given the difficulty of local image correction, we performed the first user studies to see whether users have a preference for correcting one illuminant or the other. Indeed, our studies showed that users have a clear preference for correcting the outdoor illuminant to produce warmer images. For two illuminant detections, we perform a global correction based on this user preference. Additionally, we have provided a new evaluation data set of images containing two illuminations. This image data set is collected from other existing widely-used illumination and image processing

data sets. We believe the findings in this thesis will be beneficial in helping to develop further approaches for multi-illuminant estimation and subsequent image correction.

5.2 Future Work

Our two illuminant framework is general and works with any single illumination estimation method, ideally one that provides a confidence measure so that erroneous estimates can be discarded. The results of our two illuminant detection and correction will potentially improve in the future by using improved single illuminant estimation methods. This work focused on images where there are reasonably large areas that are illuminated mainly by each illuminant. Our method will fail when the illuminants are significantly mixed almost everywhere in an image. It will be interesting to see if our method can be improved to handle such cases and to see if the user preference to correct towards warmer images remains.

Up to the time of writing this thesis, we have no theoretical justification for the mixture ratio of the two illuminants that we used for image correction (the 75% – 25% for outdoor and indoor illuminants). In the case of only two choices, it was a 75% preference for outdoor correction. The 5-choice experiment was carried out to test if this preference still holds when other options are provided (*e.g.* do users simply prefer a 50% – 50% blend?). However, we see that the preference still remains clearly in favor of the outdoor illumination, but not at the extreme. Also, we believe scene content may play a role in this preference. Thus, it is worth further exploring this user preference for image correction in the future.

Bibliography

Agarwal, V., Gribok, A. V., Koschan, A., and Abidi, M. A. 2006. Estimating illumination chromaticity via kernel regression. In *IEEE International Conference on Image Processing (ICIP)*, IEEE, 981–984. 16

Beigpour, S., Riess, C., van de Weijer, J., and Angelopoulou, E. 2014. Multi-illuminant estimation with conditional random fields. *IEEE Transactions on Image Processing (TIP) 23*, 1, 83–96. 3, 17, 18

Bianco, S., and Schettini, R. 2012. Color constancy using faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 65–72. 16

Bianco, S., and Schettini, R. 2014. Adaptive color constancy using faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 36*, 8, 1505–1518. 3, 18

Bianco, S., Ciocca, G., Cusano, C., and Schettini, R. 2008. Improving color constancy using indoor–outdoor image classification. *IEEE Transactions on Image Processing (TIP) 17*, 12, 2381–2392. 15

Bianco, S., Ciocca, G., Cusano, C., and Schettini, R. 2010. Automatic color constancy algorithm selection and combination. *Pattern Recognition 43*, 3, 695–705. 15

Bleier, M., Riess, C., Beigpour, S., Eibenberger, E., Angelopoulou, E., Tröger, T., and Kaup, A. 2011. Color constancy and non-uniform illumination: Can existing algorithms work? In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 3, 17, 18

Boyadzhiev, I., Bala, K., Paris, S., and Durand, F. 2012. User-guided white balance for mixed lighting conditions. *ACM Transactions on Graphics (TOG) 31*, 6, 200. 3, 18

Buchsbaum, G. 1980. A spatial processor model for object colour perception. *Journal of the Franklin institute 310*, 1, 1–26. 3, 13, 17, 27, 30, 32

Cardei, V. C., Funt, B., and Barnard, K. 1998. Adaptive illuminant estimation using neural networks. In *ICANN 98*. Springer, 749–754. 16

Cardei, V. C., Funt, B., and Barnard, K. 2002. Estimating the scene illumination chromaticity by using a neural network. *JOSA A 19*, 12, 2374–2386. 16

Chakrabarti, A., Hirakawa, K., and Zickler, T. 2012. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34*, 8, 1509–1519. 3

Cheng, D., Prasad, D. K., and Brown, M. S. 2014. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A 31*, 5, 1049–1058. 3, 16

Cheng, D., Price, B., Cohen, S., and Brown, M. S. 2015. Effective learning-based illuminant estimation using simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3, 16, 22, 23, 24, 25, 27, 28, 30, 33

Cheng, D. 2015. *A STUDY OF ILLUMINANT ESTIMATION AND GROUND TRUTH COLORS FOR COLOR CONSTANCY*. PhD thesis, National University of Singapore. 10, 13

Dang-Nguyen, D.-T., Pasquini, C., Conotter, V., and Boato, G. 2015. Raise: A raw images dataset for digital image forensics. In *ACM Multimedia Systems (MMSys)*. 26, 27, 35, 42

DataColor, 2016. Spyder Display Calibration. http://spyder.datacolor.com/display-calibration//. [last accessed November 2016]. 40

Ebner, M. 2009. Color constancy based on local space average color. *Machine Vision and Applications 20*, 5, 283–301. 3, 16

Finlayson, G. D., and Trezzi, E. 2004. Shades of gray and colour constancy. In *Color and Imaging Conference*, vol. 2004, Society for Imaging Science and Technology, 37–41. 3, 14

Finlayson, G. D., Hordley, S. D., and Hubel, P. M. 2001. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 23*, 11, 1209–1221. 3

Bibliography

Finlayson, G. D., Hordley, S. D., and Morovic, P. 2005. Colour constancy using the chromagenic constraint. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 40

Finlayson, G. D. 2013. Corrected-moment illuminant estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1904–1911. 3, 16, 27, 28, 30, 31, 32, 40, 42

Forsyth, D. A. 1990. A novel algorithm for color constancy. *International Journal of Computer Vision (IJCV) 5*, 1, 5–35. 3, 15

Funt, B., Cardei, V., and Barnard, K. 1996. Learning color constancy. In *Color and Imaging Conference*, vol. 1996, Society for Imaging Science and Technology, 58–60. 16

Gehler, P. V., Rother, C., Blake, A., Minka, T., and Sharp, T. 2008. Bayesian color constancy revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 1–8. 25, 27, 35, 42

Gershon, R., Jepson, A. D., and Tsotsos, J. K. 1987. From [r, g, b] to surface reflectance: Computing color constant descriptors in images. In *IJCAI*, Citeseer, 755–758. 14

Gijsenij, A., and Gevers, T. 2007. Color constancy using natural image statistics. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 1–8. 15

Gijsenij, A., and Gevers, T. 2011. Color constancy using natural image statistics and scene semantics. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 33*, 4, 687–698. 15

Gijsenij, A., Gevers, T., and Van De Weijer, J. 2011. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing (TIP) 20*, 9, 2475–2489. 13

Gijsenij, A., Gevers, T., and Van De Weijer, J. 2012. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34*, 5, 918–929. 15, 17, 40, 42

Gijsenij, A., Lu, R., and Gevers, T. 2012. Color constancy for multiple light sources. *IEEE Transactions on Image Processing (TIP) 21*, 2, 697–707. 3, 27, 30, 32, 33

Hacker, S., and von Ahn, L. 2009. Matchin: Eliciting user preferences with an online game. In *SIGCHI Conference on Human Factors in Computing Systems*. 39

Hsu, E., Mertens, T., Paris, S., Avidan, S., and Durand, F. 2008. Light mixture estimation for spatially varying white balance. *ACM Transactions on Graphics (TOG) 27*, 3, 70. 3, 18

Joze, H. R. V., and Drew, M. S. 2012. Exemplar-based colour constancy. In *Proceedings of the British Machine Vision Conference (BMVC)*, 1–12. 16

Joze, H. R. V., and Drew, M. S. 2014. Exemplar-based color constancy and multiple illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 36*, 5, 860–873. 3, 16, 18, 25

KARAIMER, H. C., AND BROWN, M. S. 2016. A software platform for manipulating the camera imaging pipeline. In *Proceedings of the European Conference on Computer Vision (EECV)*, Springer, 429–444. 8

LAND, E. H., AND MCCANN, J. J. 1971. Lightness and retinex theory. *JOSA 61*, 1, 1–11. 3, 14

LAND, E. H., ET AL. 1977. *The retinex theory of color vision*. Scientific America. 3, 16

MARTINKAUPPI, J. B., AND PIETIKÄINEN, M. 2005. Facial skin color modeling. In *Handbook of face recognition*. Springer, 113–135. 23

MCCANN, J. J., MCKEE, S. P., AND TAYLOR, T. H. 1976. Quantitative studies in retinex theory a comparison between theoretical predictions and observer responses to the color mondrian experiments. *Vision Research 16*, 5, 445IN2–458IN3. 1

RAHTU, E., NIKKANEN, J., KANNALA, J., LEPISTÖ, L., AND HEIKKILÄ, J. 2009. Applying visual object categorization and memory colors for automatic color constancy. In *International Conference on Image Analysis and Processing*, Springer, 873–882. 16

RAMANATH, R., SNYDER, W. E., YOO, Y., AND DREW, M. S. 2005. Color image processing pipeline. *IEEE Signal Processing Magazine 22*, 1, 34–43. 8

RIESS, C., EIBENBERGER, E., AND ANGELOPOULOU, E. 2011. Illuminant color estimation for real-world mixed-illuminant scenes. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 3, 17

SHI, L., AND FUNT, B. 2010. Re-processed version of the gehler color constancy dataset of 568 images. *Simon Fraser University 1*, 2, 3. 25, 27, 35, 42

Shi, L., and Funt, B. 2012. Maxrgb reconsidered. *Journal of Imaging Science and Technology 56*, 2, 20501–1. 14

Shi, L., Xiong, W., and Funt, B. 2011. Illumination estimation via thin-plate spline interpolation. *JOSA A 28*, 5, 940–948. 16

Tan, R. T., Nishino, K., and Ikeuchi, K. 2004. Color constancy through inverse-intensity chromaticity space. *JOSA A 21*, 3, 321–334. 17

Van De Weijer, J., and Gevers, T. 2005. Color constancy based on the grey-edge hypothesis. In *IEEE International Conference on Image Processing (ICIP)*, vol. 2, IEEE, II–722. 3, 15

Van de Weijer, J., Schmid, C., and Verbeek, J. 2007. Using high-level visual information for color constancy. In *IEEE International Conference on Computer Vision (ICCV)*, IEEE, 1–8. 3, 16

von Kries, J. 1878. Beitrag zur physiologie der gesichtsempfindung. *Arch. Anat. Physiol 2*, 505-524, 2. 12

Wyszecki, G., and Stiles, W. S. 1982. *Color science*, vol. 8. Wiley New York. 8

Yang, K.-F., Gao, S.-B., Li, Y.-J., and Li, Y. 2015. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3, 18