Illumination and Object Reflectance Modeling

A dissertation submitted by **Shida Beigpour** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Director	**Dr. Joost van de Weijer** Centre de Visió per Computador
Tutor	**Dr. Ernest Valveny** Dep. Ciències de la Computació Universitat Autònoma de Barcelona
Thesis committee	**Dr. Peter Gehler** Max Planck Institute for Intelligent Systems **Dr. Damien Muselet** Jean Monnet University **Dr. Ramon Baldrich** Universitat Autònoma de Barcelona

This document was typeset by the author using LaTeX 2_ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

ISBN:

Printed by

. .

To Sven for his unconditional love
and everlasting support,
and to my sister Shima for being my idol
and introducing me to science.

. .

If you can t explain it simply, you don t understand it well enough.

.

- Albert Einstein

Abstract

Surface reflectance modeling is an important key to scene understanding. An accurate reflectance model which is based on the laws of physics allows us to achieve realistic and physically plausible results. Using such model, a more profound knowledge about the interaction of light with objects surfaces can be established which proves crucial to variety of computer vision application. Due to high complexity of the reflectance model, the vast majority of the existing computer vision applications base their methods on simplifying assumptions such as Lambertian reflectance or uniform illumination to be able to solve their problem.

However, in real world scenes, objects tend to exhibit more complex reflections (diffuse and specular) and are furthermore affected by the characteristics and chromaticity of the illuminants. In this thesis, we incorporate a more realistic reflection model in computer vision applications.

To address such complex physical phenomenon, we extend the state-of-the-art object reflectance models by introducing a Multi-Illuminant Dichromatic Reflection model (MIDR). Using MIDR we are able to model and decompose the reflectance of an object with complex specularities under multiple illuminants presenting shadows and interreflections. We show that this permits us to perform realistic re-coloring of objects lit by colored lights, and multiple illuminants.

Furthermore, we propose a "local" illuminant estimation method in order to model the scenes with non-uniform illumination (e.g., an outdoor scene with a blue sky and a yellow sun, a scene with indoor lighting combined with outdoor lighting through a window, or any other case in which two or more lights with distinct colors illuminating different parts of the scene). The proposed method takes advantage of a probabilistic and graph-based model and solves the problem by re-defining the estimation problem as an energy minimization. This method provides us with local illuminant estimations which improve greatly over state-of-the-art color constancy methods.

Moreover, we captured our own multi-illuminant dataset which consists of complex scenes and illumination conditions both outdoor and in laboratory conditions. We show improvement achieved using our method over state-of-the-art methods for local illuminant estimation.

We demonstrate that having a more realistic and accurate model of the scene illumination and object reflectance greatly improves the quality of many computer vision and computer graphics tasks. We show examples of improved automatic white balance, scene relighting, and object re-coloring. The proposed theory can be employed in order to improve color naming, object detection, recognition, and segmentation which are among the most popular computer vision trends.

Resumen

El modelado de la reflectancia de las superficies es una clave importante para la comprensión de escenas. Un modelo de reflectancia preciso, basado en las leyes de la f´sica, nos permite alcanzar resultados real´sticos y f´sicamente plausibles. Además, el uso de tal modelo nos permite establecer un conocimiento más profundo acerca de la interacción de la luz con las superficies de los objetos, y resulta crucial para una variedad de aplicaciones de visión por computador. Debido a la alta complejidad de los modelos de reflectancia, la gran mayor´a de las aplicaciones existentes de visión por computador basan sus métodos en suposiciones simplificadoras, tales como la reflectancia lambertiana o la iluminación uniforme para ser capaz de resolver sus problemas.

Sin embargo, en escenas del mundo real, los objetos tienden a exhibir reflexiones más complejas (difusas y especulares), y además se ven afectados por las caracter´sticas y la cromaticidad de los iluminantes. En esta tesis, se incorpora un modelo de reflexión más realista para aplicaciones de visión por computador.

Para abordar tal fenómeno f´sico complejo, extendemos los modelos de reflectancia de los objetos del estado-del-arte mediante la introducción de un Modelo de Reflexión Dicromático Multi-Iluminante (MIDR). Usando MIDR somos capaces de modelar y descomponer la reflectancia de un objeto con especularidades complejas bajo múltiples iluminantes que presentan sombras e interreflexiones. Se demuestra que este modelo nos permite realizar una recolorización realista de los objetos iluminados por luces de colores y múltiples iluminantes.

Además se propone un método "local" de estimación del iluminante para modelar las escenas con iluminación no uniforme (por ejemplo, una escena al aire libre con un cielo azul y un sol amarillo, una escena interior con iluminación combinada con la iluminación al aire libre a través de una ventana, o cualquier otro caso en el que dos o más luces con diferentes colores iluminan diferentes partes de la escena). El método propuesto aprovecha un modelo probabil´stico basado en grafos y resuelve el problema redefiniendo la estimación como un problema de minimización de energ´a. Este método nos proporciona estimaciones locales del iluminante que mejoran en gran medida a los métodos del estado-del-arte en constancia de color.

Por otra parte, hemos capturado nuestro propia base de datos multi-iluminante, que consiste de escenas complejas y condiciones de iluminación al aire libre o de laboratorio. Con ésta se demuestra la mejora lograda usando nuestro método con respecto a los métodos del estado-del-arte para la estimación automática del iluminante local.

Se demuestra que tener un modelo más realista y preciso de la iluminación de la escena y la reflectancia de los objetos, mejora en gran medida la calidad en muchas tareas de visión por ordenador y gráficos por computador. Mostramos ejemplos de mejora en el

balance automático de blanco, reiluminación de escenas y en la recolorización de objetos. La teor´a propuesta se puede emplear también para mejorar la denominación automática de colores, la detección de objetos, el reconocimiento y la segmentación, que están entre las tendencias más populares de la visión por computador.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many Computer Vision applications deal with the object reflectance. Conventional methods often consider the Lambertian model in order to analyze the reflectance. In real world applications, objects tend to exhibit more complex reflections due to shadows, highlights, and are furthermore affected by the characteristics and chromaticity of their illuminant. To address such a complex physical phenomenon, we make use of state-of-the-art object reflection models, which being based on the laws of physics, allow us to achieve realistic and physically plausible results. Using these models, a more profound knowledge of light interaction with objects surfaces can be established which proves crucial to a variety of computer vision application.

Considering advanced imaging technologies which result in high quality images and videos, and in order to address the increasing need for precise and realistic analysis of the images, many computer vision scientists developed more physically plausible reflectance methods. While the classical Lambertian model ignores object specularities and highlights, the Dichromatic Reflection Model (DRM) proposed by Shafer [87] tries to achieve a more realistic model by embedding a specular term in the reflection formula. State-of-the-art approaches often need to simplify the model by making assumptions (e.g. known illuminant, use of image sequence, controlled environment) in order to solve the model.

Furthermore, real-world objects often exhibit body and surface reflection under more than just one illuminant. An example of multi-illuminant scenario is an outdoor scene with a blue sky and a yellow sun, or a scene with indoor lighting combined with outdoor lighting through a window. Conventional methods often ignore secondary illuminants present in the scene to simplify the modeling.

The *Automatic White Balance* (AWB) embedded in the digital cameras is a very popular application in digital photography. Most digital cameras use statistical methods to correct for the scene illumination. While these methods appear to perform well on many images, they mostly fail in the case of complex scenes. As an example, the dominant presence of one color in the scene, like a close-up of a person with a red shirt, leads many state-of-the-art cameras to wrongfully apply a very bluish filter to compensate for what they assume to be sunset effects. Another example is the case of having multiple illu-

minants with strong effects in an image which confuses the automatic white balancing method that is trying to estimate one illuminant for the whole image.

In the next chapter, we briefly review the existing work and trends regarding illumination and object reflectance modeling.

In Chapter 3 we attempt to solve the model with minimal assumptions and to form a realistic decomposition of the object reflectance. Our aim is to solve the case of an arbitrary uncalibrated scene with multiple specular and non-specular (diffuse) colored objects under multiple illuminants presenting shadows and inter-reflections. The realistic decomposition achieved will then enable us to perform highly challenging computer vision tasks e.g., Color Naming, Object Detection, Recognition and Segmentation which are among the most popular computer vision trends.

In Chapter 4 we propose to apply a *local* illuminant estimation method in order to improve the estimations in the case of scenes illuminated by differently colored light sources at the same time. We do so, by expressing the illuminant estimation problem as an energy minimization which enable us to easily combine the state-of-the-art methods on global illuminant estimation into a mathematically sound formulation. This also further simplifies and improves the implementation and results. We propose to take advantage of *Conditional Random Fields* to achieve global consistency of the illuminant estimations. Finally, as a useful application of our framework, we demonstrate that using this local modeling, we will be able to develop a much more intelligent automatic white balancing method.

Further details on data acquisition and ground truth computation for our multi-illuminant scene datasets will be presented in Chapter 5. The first part of this chapter focuses on the data captured using camera from real scenes and followed by the details on our synthetic image dataset constructed using graphical but physically accurate 3D modeling and rendering softwares. We conclude this Chapter by comparing the two different methods presented for constructing and collecting our datasets.

The final chapter summarizes the current work and concludes the thesis as well as presenting ideas for the future directions of this research.

Chapter 2

Research Context

The aim of this chapter is to present the general context and overall background of the topics which are covered during the next chapters.

2.1 Reflectance Models

Object reflectance estimation from an image is an active subject in color-vision whose application ranges from color constancy to segmentation and classification. The main idea is: if we would be able to build a realistic model of the light interaction with the object surface, we could extract crucial knowledge about the object surface geometry as well as the illuminant light. Such knowledge would then be used in order to remove the effect of a non-white illumination (*color constancy*), locate and remove the areas of shadows and highlights (essential in *object segmentation*), and obtain the geometrical model of the object (improving *object classi cation*).

2.1.1 Lambertian Reflection

One of the simplest and most commonly used models is the *Lambertian* reflectance. This model is based on the simple assumption that the intensity of the light reflected from the surface is independent of the viewing angle and the surface luminance is considered to be *isotropic*. Some examples of Lambertian (matte) materials are chalk, soil and paper [1]. The reflected energy from the surface E is given by:

$$E(\lambda\ x) = m_b(\mathbf{x})s(\lambda\ \mathbf{x})e(\lambda\ \mathbf{x}) \tag{2.1}$$

where $e(\lambda\ \mathbf{x})$ is the illumination in the scene, λ is the wavelength and $\mathbf{x}$ the spacial coordinates of the pixel in the scene. From here on we use bold face to denote vectors.

[1] Here we ignore the translucency characteristic of many commonly used types of paper which is caused by *optical brightening* for commercial reasons.

Often in the literature it is assumed that the spectral distribution of the light source is spatially uniform across the scene and it does not depend on the position; therefore the $\mathbf{x}$ notation can be omitted. Here m^b is the geometric part of the reflectance which depends on the α angle between the direction of the incident light and the surface normal ($m_b = \cos(\alpha)$).

Now the value measured by the camera with spectral sensitivity $\rho^c(\lambda)\ c = R\ G\ B$ at position $\mathbf{x}$ is modeled as $f^c(\mathbf{x})$ by integrating over the visible spectrum ,

$$f^c(\mathbf{x}) = m_b(\mathbf{x}) \quad b(\lambda\ \mathbf{x})e(\lambda)\rho^c(\lambda)d\lambda \tag{2.2}$$

We define the *body reflectance* as

$$c_b{}^c(\mathbf{x}) = \quad b(\lambda\ \mathbf{x})e(\lambda)\rho^c(\lambda)d\lambda \tag{2.3}$$

Therefore,

$$\mathbf{f}(\mathbf{x}) = m_b(\mathbf{x})\mathbf{c_b}(\mathbf{x}) \tag{2.4}$$

In spite of it is inaccuracy for describing real-world scenes, still the majority of computer vision methods are based on the Lambertian assumption.

2.1.2 Dichromatic Reflection Model (DRM)

A more realistic reflectance model is the *dichromatic reflection model* (DRM) proposed by Shafer [86]. The model focusses on the color aspects of light reflection and has only limited usage for geometry recovery of scenes. It separates reflectance into surface body reflectance and interface reflectance. The model is valid for the class of inhomogeneous materials, which covers a wide range of materials such as wood, paints, papers and plastics (but excludes homogeneous materials such as metals). It predicts that values of a single colored object lie on a parallelogram in color space, defined by the body reflectance and the illuminant color.

For multiple light sources we assume that the combination can be approximated as a single light source for the local feature.

$$f^c\mathbf{x} = m_b(\mathbf{x}) \quad b(\lambda\ \mathbf{x})\, e(\lambda)\rho^c(\lambda)\, d\lambda + m_s(\mathbf{x}) \quad i(\lambda)\, e(\lambda)\rho^c(\lambda)\, d\lambda \tag{2.5}$$

where b is the surface albedo. We assume neutral interface reflection, meaning that the Fresnel reflectance i is independent of λ. Accordingly, we will omit i in further equations. The geometric dependence of the reflectance is described by the terms m_b and m_s which

depend on the viewing angle, light source direction and surface orientation. $\mathbf{x}$ denotes the spatial coordinates, and bold face is used to indicate vectors. In vector notation we can now write:

$$\mathbf{f}(\mathbf{x}) = m_b(\mathbf{x})\,\mathbf{c_b}(\mathbf{x}) + m_s(\mathbf{x})\,\mathbf{c_s}(\mathbf{x}) \tag{2.6}$$

The reflection of the light consist of two parts: 1. a body reflection part $m_b(\mathbf{x})\,\mathbf{c}_b$, which describes the light which is reflected after interaction with the surface albedo, and 2. the interface reflection $m_s(\mathbf{x})\,\mathbf{c}_s$ which describes the part of the light that is immediately reflected at the surface, causing specularities. Both parts consist of a geometrical part dependent on the location in the scene, and a spectral part dependent on the spectral wavelength.

Several methods have been developed to approximate the dichromatic model of an object. Kravtchenko and Little have introduced a spatial-based approach in their segmentation method in which they approximate the two dichromatic planes for specular and body reflectance considering the lighter and darker pixels separately [58]. Shen and Xin have solved the model with the assumption of a known illuminant [88].

The original application to which the DRM was applied, was the separation of shading from specularities [86]. The specularities, being dependent on scene incidental events such as viewpoint and surface normal, could be removed to simplify color image understanding. The removal of specularities allowed for improved segmentation algorithms [55, 71]. Furthermore, the estimation of the specularities also provides an illuminant estimation, thereby allowing for color constancy. A second application field which has benefited from the DRM is photometric invariant feature computation [42, 97].

2.1.3 Multi-illuminant

In cases that the assumptions made by the original DRM are not met, more complex reflectance models are required. One such case is ambient light, i.e. light coming from all directions. Ambient light occurs in outdoor scenes where next to the dominant illuminant, i.e. the sun, there is diffuse light coming from the sky. Similarly, it occurs in indoor situations where diffuse light is caused by reflectances from walls and ceilings. Shafer [86] models the diffuse light, a, by a third term

$$\mathbf{c}(\mathbf{x}) = m_b(\mathbf{x})\,\mathbf{c}_b(\mathbf{x}) + m_s(\mathbf{x})\,\mathbf{c}_s(\mathbf{x}) + a^C \tag{2.7}$$

Later work improved the modeling [70, 80] and showed that the ambient term results in an object color dependent offset which could perform crucial in handling the case of colored shadows. Furthermore, in [70] a photometric invariant with respect to ambient light is proposed.

Another case is the presents of multiple illuminants in the scene (a more generalized case of ambient light). A typical example of a ”multi-illuminant” is the interreflections occurring between objects in the complex scenes.

Figure 2.1 is a physics-based illustration of the case of an object observed under multiple-illuminant. A main focus of this thesis in the next chapters is to present novel approaches to model complex multi-illuminant scenarios in order to accurately estimate and modify the illuminant chromaticity.

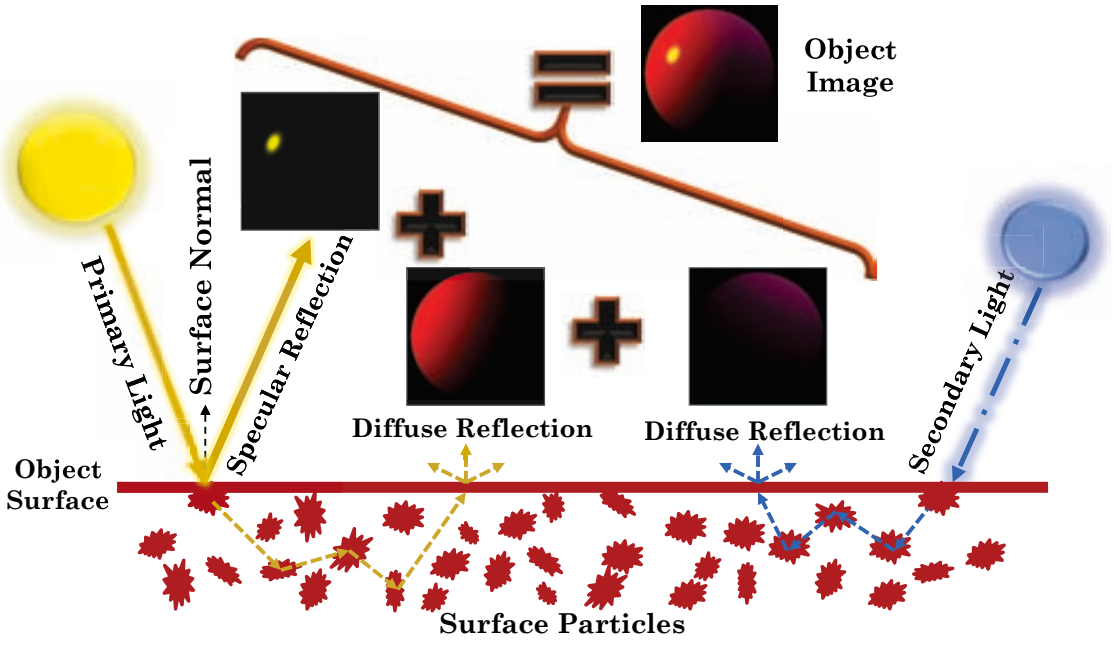


Figure 2.1: Real-world objects often exhibit body and surface reflection under more than just one illuminant (e.g., outdoor scene with blue sky and yellow sun).

2.1.4 Inverse-Intensity Chromaticity Space

Using the Dichromatic Reflection Model, Tan *et al.* [91] formulates the camera response $I_c(x)$ for each color filter c as:

$$I_c(\mathbf{x}) = m_b(\mathbf{x})\Lambda_c(\mathbf{x}) + m_s(\mathbf{x})\Gamma_c \ , \tag{2.8}$$

where $\Lambda_c(\mathbf{x}) = B_c(\mathbf{x}) / \sum_i B_i(\mathbf{x})$ and $\Gamma_c = G_c / \sum_i G_i, i \in \{R, G, B\}$ are diffuse and specular chromaticities respectively. $B_c(\mathbf{x})$ and G_c are the respective camera responses defined as below:

$$B_c(\mathbf{x}) = \int_\omega b\left(\lambda, \mathbf{x}\right) e\left(\lambda\right) \rho^c\left(\lambda\right) d\lambda$$

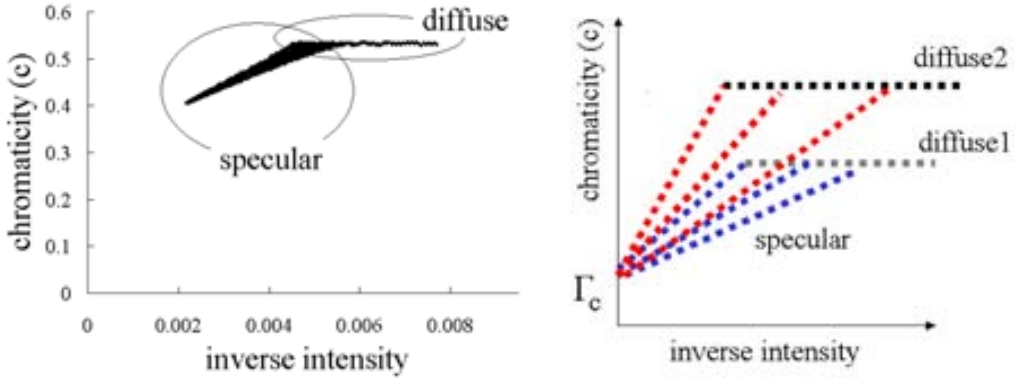
$$G_c = \int_\omega i\left(\lambda\right) e\left(\lambda\right) \rho^c\left(\lambda\right) d\lambda$$

Here we assume *neutral interface reflection* and that the color of the illumination over the input image is uniform so that the spectral distribution of the illuminant becomes independent of the image coordinate $\mathbf{x}$.

The image chromaticity is similarly defined as $\sigma_c(\mathbf{x}) = I_c(\mathbf{x}) \quad {}_i I_i(\mathbf{x})$. Tan *et al.* define the correlation between image chromaticity and illumination chromaticity as (we will omit the spatial arguments):

$$\sigma_c = p\frac{1}{I_i} + \Gamma_c \tag{2.9}$$

where $p = m_d(\Lambda_c - \Gamma_c)$. This demonstrates that we are able to determine the chromaticity of the illuminant, Γ_c, using p since the image chromaticity, Λ_c, and total image intensity, I_i can be obtained using the input image. Tan *et al.* finally conclude that in the *Inverse-Intensity Chromaticity* (IIC) space (in which the chromaticity σ_c and inverse intensity I_i are the vertical and horizontal axis respectively) the diffuse pixels form a horizontal line and the specular pixels form a diagonal line which intersect the chromaticity axis at the illuminant chroma (Γ_c). Figure 2.2 presents example of the IIC space.



(a) (b)

Figure 2.2: Here's an illustration for the intensity chromaticity space: (a) Sketch of specular points of two surface colors in inverse-intensity chromaticity space; (a) Diffuse and specular points of an image. Images are taken from Tan *et al.* [91].

2.2 Intrinsic Images

As mentioned in the previous section, the human vision has the ability to perceive characteristics intrinsic to the scene, such as color, size, shape, distance, orientation and etc. In 1978 the term *intrinsic images* has been first coined by Barrow and Tenenbaum [11] referring to a family of images each of which contains the value of one of the intrinsic characteristic at each point corresponding to the input image and additionally the explicit indications of boundaries due to discontinuity in value or gradient.

Using the idea that the main variations in an image sequence of a fairly static outdoor scene should be the illumination changes, a method for object color decomposition has been developed in which the object surface reflectance model has been extracted, assuming the camera response to be linear [102]. The author has used the assumption that when

derivative filters are applied to natural images, the filter output tend to be sparse. Then a maximum likelihood estimator has been used for surface reflectance recovery.

Recently an effort by Grosse *et al.* [50] to introduce a quantitative measure for benchmarking and evaluating the different intrinsic estimation methods has resulted in the widely used MIT dataset which has greatly encouraged more works on this topic. Since then many methods have been developed and tested using this dataset. Gehler *et al.* [39] developed a novel approach for intrinsic image recovering on single images by assuming that the reflectance values are drawn from a sparse set of basis colors. Using their probabilistic approach they managed to achieve quality results on MIT dataset, while they demonstrate that adding more cues such as edge information to their model has led to state-of-the-art results. Another method which uses probabilistic models for this task is the work of Serra *et al.* [85] who also show competitive results obtained on the MIT dataset. In this work, *ridges* have been used as extra cues to improve the results.

In chapter 5 we further approach this problem and introduce our own dataset for intrinsic image benchmarking.

2.3 Illumination Estimation and Color Constancy

Estimating the illuminant and reflectance are highly related. Having a good estimation of the illuminant could result in much more accurate reflectance model and intrinsic image decomposition as shown by many existing methods. So far, there have been many methods presented in different fields of computer vision and image processing which assume the illuminant to be white or known. But in many cases the scene is much more complex regarding the illumination. Here we start with a brief introduction on perceptual color constancy in human vision followed with a summary of computational color constancy methods.

2.3.1 Perceptual Color Constancy

A commonly accepted definition of perceptual constancy is the *"relative stability of the apparent value of object properties (size, shape, orientation, movement, etc.) when the representation at the eye (retinal image) is variant with change in observer position, posture, and movement."* [20]

Some examples of perceptual constancy in the literature are: *size constancy* when we look at the object from far away or close up; *lightness constancy* when we do not see the large differences in lightness even when objects are illuminated by intensities that differ in several orders of magnitude; *shape constancy* referring to objects being perceived similar despite the distance and the viewing angle; identification of a musical instrument as constant under changing timbre or conditions of changing pitch and loudness, in different environments and with different players; in speech the vowels or consonants are

perceived as constant categories even if acoustically, they vary greatly due to phonetic environment, tempo, speaker's age, gender, or dialect. One of the interesting forms of perceptual constancy is *color constancy* which could play an important role in crucial tasks such as finding and classifying the objects. Color of an object is a valuable cue in determining whether a fruit is ripe or an animal is poisonous.

Ebner in his book "color constancy" [24] defines this phenomenon as the human observers' ability to recognize the color of an object irrespective of the light used to illuminate them. He further explains that since a digital camera uses a sensor to measure the light reflected from the object surface, this measurement at pixel level varies according to the color of the illuminant. This could result in the color of the pixels being different from the ones perceived by the human observer. Color constancy mechanism also exists in various animals like, honeybees, goldfish, and monkeys (the last are believed to poses the vision which is most similar to humans).

Regarding the reasons for which color constancy could occur in human vision, there are various studies in the literature.These studies converge in number of possibilities. For example Goldstein in "sensation and perception" [48] points out that color perception can be changed by *chromatic adaptation*. For example, prolonged exposure to red light, bleaches the long-wavelength cone pigment in one's eye which decreases the sensitivity to red light and causes the perception of the red or reddish colors to be less saturated. Digital cameras on the other hand use sensor with fixed responsivity to wavelength and that causes the difference in between the measured pixel colors and the human perception of the objects in the scene.

Also the colors in the *surrounding* enhances color constancy which has been for many years an important cue to deal with the problem of computational color constancy. This phenomenon is best noticed when the object is surrounded by objects of many different colors. While many white balancing function in commercial cameras work this way, it is possible to trick this system to mistake a color for another because of its surrounding. Figure 2.3 demonstrates some illusions created using this matter in the literature. *Color memory* is also another reason often considered while explaining the color constancy phenomena.

2.3.2 Computational Color Constancy

Since in many applications which deal with photos and videos consistency with the human perception is desired, achieving computational color constancy plays an important role. From designing better photo filters and white balancing in the art of photography, to fundamental applications in computer vision like object classification and segmentation, estimating and modeling the illuminant could serve as a crucial pre-processing step in order to achieve good results.

Unlike perceptual color constancy, computational color constancy deals with the optics and underlying *physical* laws regarding the light's interaction with object surface

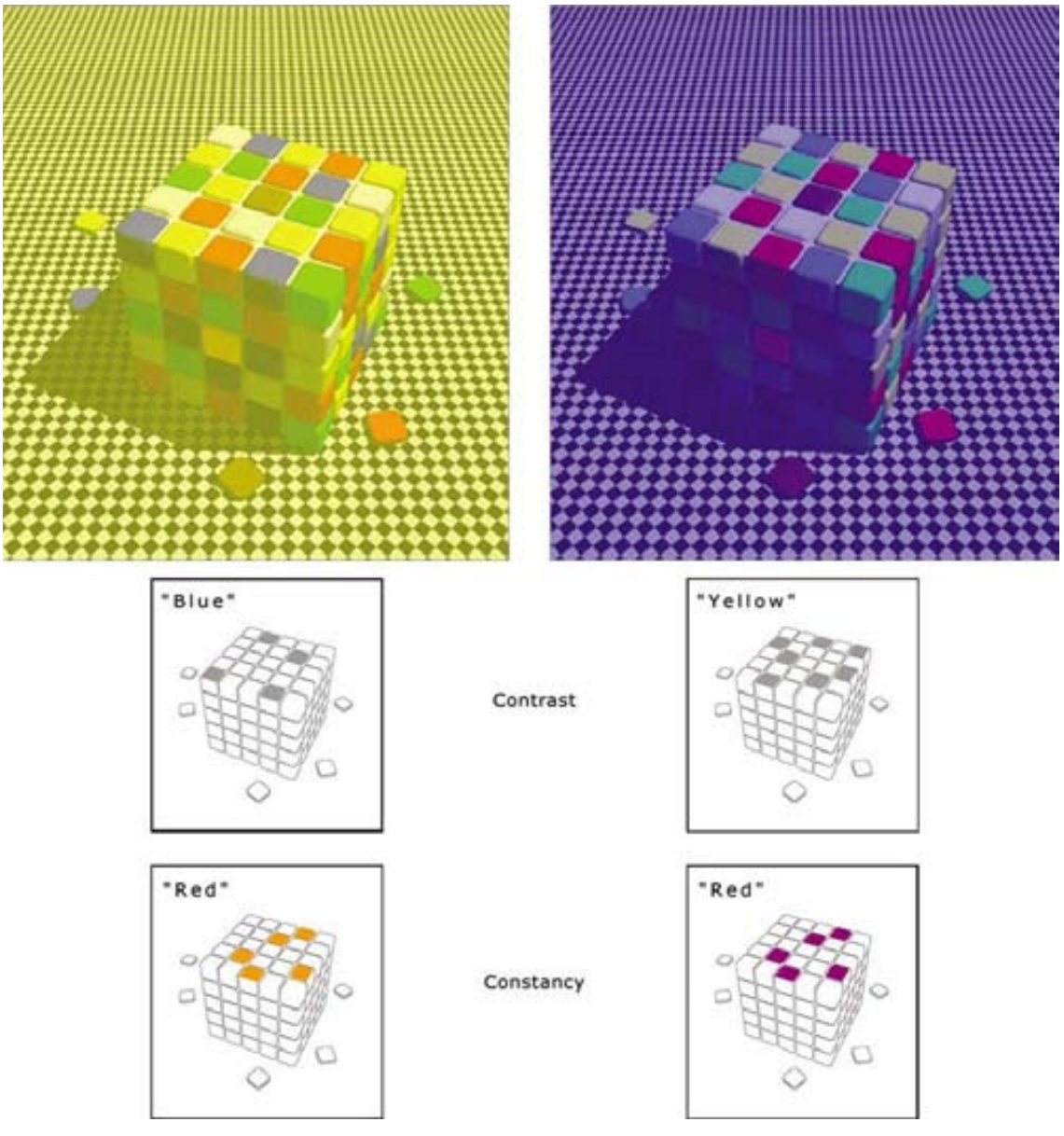


Figure 2.3: This figure demonstrates the effect of color constancy on the perception of colors. The first row is an example of such effect using two different surrounding colors. Images on the second row show that the bluish tiles on the image on the top of the cube on the left are identical to the yellowish tiles on the top of the cube in the right, and all are in fact gray. Also the red tiles on the top of the both cubes, even though appear identical, are in fact different colors as demonstrated in the third row. Figures are taken from Lotto *et al.* [67].

rather than the reasons for which a human subject would perceive a scene in a certain way. That is to say, the main goal in the computational color constancy is to estimate the chromaticity of the lights illuminating the scene and transforming the input image to the *canonical image* (i.e., the image of the scene taken under a neutral or white light source) by removing the effects of the illumination color.

Chromatic Adaptation

Using the *chromatic adaptation* transform, the appearance of the colors in the image is changed in order that they would appear as if being captured under a white light source. The simplest and most commonly used transformation of this kind is the von Kries [101] transform which independently scales each of the cone responses or RGB channels:

$$\begin{matrix} R_c \\ G_c \\ B_c \end{matrix} = \begin{matrix} d_R & 0 & 0 \\ 0 & d_G & 0 \\ 0 & 0 & d_B \end{matrix} \times \begin{matrix} R_e \\ G_e \\ B_e \end{matrix} \tag{2.10}$$

where $d_i = \frac{e_i}{3(e^2{}_R+e^2{}_G+e^2{}_B)}$ i $R\ G\ B$. In spite of being merely an approximation of illumination change, it is the commonly accepted model in the literature due to its simplicity.

Since the von Kries model might not accurately be able to model photometric changes due to disturbing effects such as highlights and interreflections, there are more accurate models proposed for chromatic adaptation. For example, *sharpening* the cone responses before transformation [18] or using and offset as in *diagonal-offset* model [30] (which ideally becomes zero for von Kries model):

$$\begin{matrix} R_c \\ G_c \\ B_c \end{matrix} = \begin{matrix} d_R & 0 & 0 \\ 0 & d_G & 0 \\ 0 & 0 & d_B \end{matrix} \times \begin{matrix} R^u \\ G^u \\ B^u \end{matrix} + \begin{matrix} o_1 \\ o_2 \\ o_3 \end{matrix} \tag{2.11}$$

Illuminant Estimation

The recovery of the illumination color from a single image is an under-constrained problem. Every observed image pixel represents an unknown combination of surface reflectance and illumination. Many color constancy algorithms try to make this problem tractable by imposing different assumptions on the observed scene (e.g. a derivative of the pixels sums up to 0 under canonical illumination as in gray edge algorithms [95] or that the convex hull of the pixels in a suitably chosen color space encompasses most illumination changes as in gamut mapping [35]). Furthermore, most illuminant color estimators typically assume globally uniform illumination. This prerequisite is essential for collecting a sufficiently large number of samples from the whole image and thus increasing the accuracy and robustness of the methodology.

Existing Illuminant estimation methods are categorized in three main groups: static and physics-based methods, gamut-based methods, and learning-based methods. The rest of this section presents a brief description regarding each of these categories.

2.3.3 Static Methods Using Low-level statistics

Here we discuss a set of illuminant estimation methods which rely on low level statistics from the image pixels. These methods are categorized as static because they use fixed parameter setting. Popularity of these methods is mainly due to the simplicity of their implementation and their high speed. They could obtain accurate results if used with adequate parameters, while their accuracy and quality of their results would drop otherwise.

The most popular and commonly used approach in this category relies on the *gray-world* assumption, that is the average reflectance in a scene under neutral light is achromatic [17]. Or more accurately, the average reflectance in a scene is equal for every wave-length. Various extensions have been proposed to improve this method, e.g. computing the scene average over image segments to reduce the effect of large uniformly colored surfaces.

Similarly, *white-patch* [60] method assumes that the maximum response in the color channels is caused by a perfect reflectance (reflecting the full range of light). In this method, the perfect reflectance represents the color of the illuminant. Various works in the literature demonstrate that smoothing the image before performing the illuminant estimation could improve the results.

Using higher order statistics, the *gray-edge* method uses the average reflectance derivatives in a scene to estimate the illuminant color [95]. The gray-edge hypothesis is based on the observation that the distribution of the color derivatives in the image has a relatively regular (ellipsoid) shape of which the long axis is in the direction of the light source.

Using the Minkovski norm, a generalized formula is introduced in the literature which can incorporate all the above mentioned methods:

$$\left(\int \left| \frac{\partial^n (\mathbf{f}^c)_\sigma(x)}{\mathbf{x}^n} \right|^m dx \right)^{\frac{1}{m}} = k(e^c)^{n\ m\ \sigma} \tag{2.12}$$

where n is the differentiation power, and m is the Minkowski norm. σ denotes the standard deviation of a Gaussian smoothing operator that is applied to the image prior to the differentiation. e stand for the the illuminant and k is a constant. Using the generalized formula has the advantage that the choice of the method can be reduced down to the choice of the parameters.

The parameter $n > 0$ produces a higher-order color constancy method (e.g., $n = 1$ gray-edge and $n = 2$ second order gray-edge). Using higher Minkowski norm will emphasize larger measurements in the image, while lover values equally distributes weights among the measurements (e.g., $m = 1$ for gray-world and $m = \inf$ for Max-RGB). Also in Chapter 4 this formulation is used in order to incorporate the static methods into a novel Conditional Random Field (CRF) framework for local color constancy.

2.3.4 Physics-based Illuminant Estimation

Although many illuminant estimation method make Lambertian assumption (Sec 2.1.1), using the more accurate DRM formulation (Sec 2.1.2) there has been some advancements in this field. By assuming the neutral interface reflection, various illuminant estimation methods have been developed which extract the illuminant chromaticity from the specular highlights in the image [ref]. Also the *Planckian* illuminant assumption has led to a physics-based illuminant estimation method by Finlayson and Schaefer [27]. In Chapter 3 we further discuss this matter.

As described in (Sec 2.1.4), the inverse intensity chromaticity space can be used in order to extract the illuminant chromaticity as in the work by Tan *et al.* [91]. Although this work has presented a very elegant formulation based on the original work by Shafer [86], in practice the method relies on a specularity segmentation in order to identify the specular pixels. In Chapter 4 we further extend their approach for the purpose of local illuminant estimation.

2.3.5 Gamut-based Methods

A color gamut is a convex part of the color space which contains the complete set of colors which can be accurately represented in a given circumstance. Gamut mapping methods assume that only a limited set of colors can be observed under an specific given illuminant. The gamut of the possible colors for a reference illuminant (often white) is referred to as the canonical gamut. Canonical gamut is constructed using as many surfaces under the refrence light as possible.

There are many different gamut mapping illuminant estimation methods in the literature all of which consist of following steps: first the canonical gamut is formed using the training images; then the input image is used to construct a gamut that is considered a subset of the gamut of the illuminant to be estimated (that is because the input image only includes a very small subset of possible colors in its gamut); in the next step, the feasible set of mappings which applied to the input gamut result in a gamut that is completely within the canonical gamut is computed; using an estimator one mapping that best maps the unknown illuminant to the canonical gamut is chosen; finally the chosen mapping is applied on the input image to obtain the image of the scene under canonical illluminant.

Existing gamut mapping methods, although being based on the original work pioneered by Forsyth [35], differ from each other in their approach to each of the steps mentioned above. Several extensions of gamut mapping are aimed to simplify, improve or reduce the costs of the implementation and execution of this algorithm like using *convex programing* [32] or using a simple cube instead of the full convex hull of the pixel values as gamut [73]. Alternatively, 2D chroma space is used instead of the original 3D color space to reduce the complexity of the implementation and visualization of the problem. However this conversion is known to slightly decrease the performance of the method that is caused by the perspective distortion. Therefore, to solve this problem, the 2D feasible

set is mapped to 3D again before choosing the best mapping.

Some approaches deal with the dependency of gamut mapping on the diagonal model for chromatic adoption since a null solution could occur if the diagonal model fails [4, 7] . Some approaches tried to address this problem by enlarging the input or the canonical gamut using different heuristics [8, 28] . Others used extensions like *gamut-constrained illuminant estimation* by limiting the possible set of illuminants in which one canonical gamut is learned for every possible illuminant and then the unknown illuminant is estimated by matching each of these gamuts to the gamut of the input image, or *diagonal-offset* which allows for translation of input colors along with the linear transformation used in original method. Alternatively there are many approaches which combine these algorithms or the results of each of them to better estimate the illuminant or use derivatives [43].

Overall, gamut mapping has a very good potential for achieving high accuracy while being based on an elegant underlying theory. However, it is quite complex for implementation and requires a large amount of training data and adequate preprocessing.

2.3.6 Local Illuminant Estimation for non-uniform illumination

Most color constancy methods assume the illumination to be uniform, while in real-world there are many cases of multi-illuminant scenes for which this assumption results in drastic artifacts. A common example of a multi-illuminant scene is a fine sunny day that is illuminated by both the yellow sun, and blue sky. In this case different points of the scene are illuminated by different mixture of these colors. Also another example of multi-illuminant scene is the presents of *interreflections* and *colored shadows*.

Ebner *et al.* [25] proposed a method for solving the case of non-uniform illumination by computing the local space average color in the image. Recently Bleier *et al.* [12] have tried to examine the accuracy of a number of existing statistical illumination estimation methods on the images with non-uniform illumination. To this end they first divided the image to sub-regions, *superpixels*, and then applied each of the methods separately on each superpixels. Also using a similar approach Riess *et al.* [79] have extended the [91] to obtain estimate for local illumination. The work published by Gijsenij *et al.* [45] presents a framework and a dataset to address the multi-illuminant scenes. In Chapter 4 a novel approach and a dataset for illuminant estimation in multi-illuminant scenes using Conditional Random Field (CRF) is presented.

Chapter 3

Object Recoloring based on Intrinsic Image Estimation

Object recoloring is one of the most popular photo-editing tasks. The problem of object recoloring is highly under-constrained, and existing recoloring methods limit their application to objects lit by a white illuminant. Application of these methods to real-world scenes lit by colored illuminants, multiple illuminants, or interreflections, results in unrealistic recoloring of objects.
In this paper, we focus on the recoloring of single-colored objects presegmented from their background. The single-color constraint allows us to fit a more comprehensive physical model to the object. We show that this permits us to perform realistic recoloring of objects lit by colored lights, and multiple illuminants. Moreover, the model allows for more realistic scene *relighting*. Recoloring results on images captured by uncalibrated cameras demonstrate that the proposed framework obtains realistic recoloring for complex natural images. Furthermore we use the model to transfer color between objects and show that the results are more realistic than existing color transfer methods.

3.1 Introduction

Recoloring refers to the modification and adjustment of color appearance in images. Object recoloring methods are used in photo montage, image color correction, visual effects in movies, and also to facilitate the industrial design by visualizing the final color appearance of the object before production. In the current work we focus on recoloring of single-colored objects in images of medium quality as typically encountered on the Internet.

One of the most popular color modification applications is the recoloring of a specific

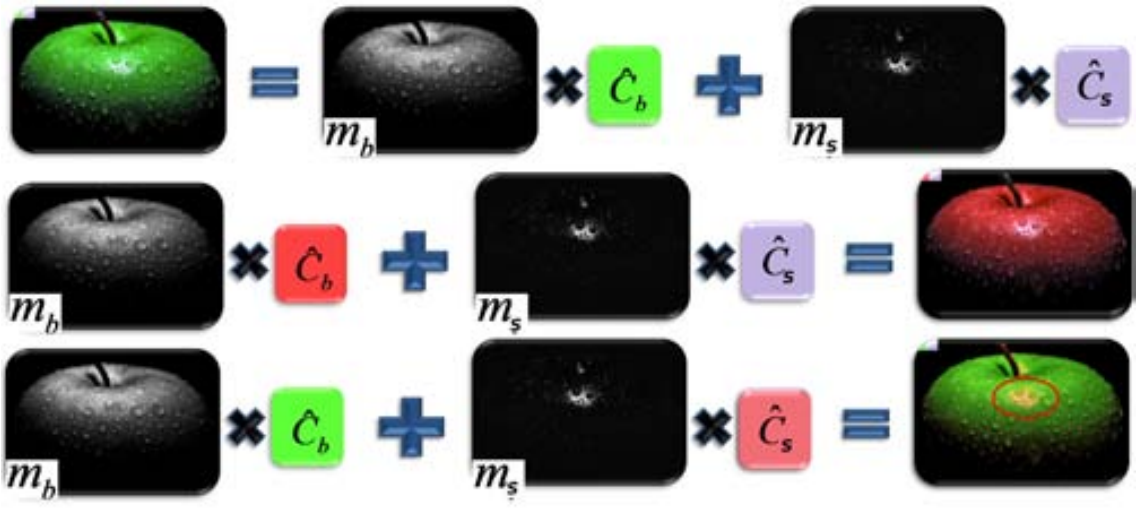


Figure 3.1: The first row is an example of the reflectance decomposition achieved by DRM [86]. Using this decomposition, *object recoloring* is performed by changing the *body reflectance* (the second row), and *illuminant recoloring* is achieved by changing the *specular reflectance* (the third row).

object with another color or under different lighting condition (e.g., warm-tone sunset or cold-tone early morning). In many circumstances, it may not be possible to create the object in the desired color or to simulate the desired lighting condition. Another case is when an impossible scenario is desired, for example a blue apple, and here the choices are to either render a 3D model of the scene or to simply photograph the object and then *recolor* it. Recoloring should result in physically plausible scenes and should require minimum user interaction.

Here our main objective is to develop a physics-based method to extract the underlying reflectance model of the object and separate the geometric characteristics from the colors of the object and the illuminant. Such physics-based model can then be used in order to generate an image of the object in the same lighting and viewing angles, varying only the object and/or illuminant colors. Fig 3.1 provides an example of reflectance decomposition as well as object and illuminant recoloring.

Images describing the underlying physical properties of the scene such as reflectance, orientation, and illumination are known as *intrinsic images* and were first introduced by Barrow and Tenenbaum [11]. Intrinsic images are more appropriate for higher-level scene analysis than the original light intensity images, because they are less prone to scene accidental events such as illuminant direction and color changes. The *Dichromatic Reflection Model*(DRM) [86] models the object reflectance using two chromatic coefficients: *body reflectance* $\mathbf{c}^b$, and *specular reflectance* $\mathbf{c}^s$:

$$\mathbf{f}(\mathbf{x}) = m_b(\mathbf{x})\mathbf{c_b} + m_s(\mathbf{x})\mathbf{c_s} \tag{3.1}$$

where, for each pixel $\mathbf{x}$, m^b and m^s are the intrinsic images describing the interaction between the light and the surface as a function of geometric parameters such as incident angle, viewing angle, and surface normal.

In this paper, we investigate the application of the *single-colored object* constraint to derive the intrinsic images of a scene. We assume a segmented mask of a single-colored object to be given as an input. A user working in a photo-editing environment has multiple segmentation tools to quickly segment objects [63, 82]. This single-colored object constraint greatly simplifies the estimation of intrinsic images. We show that this constraint allows us to further extend the DRM to model more complex scenes with multiple illuminants which proves crucial for outdoor scenes where two illuminants (e.g, the sun and a blue skylight) illuminate the object.

We propose a Multi-illuminant Dichromatic Reflection model (MIDR), and provide an algorithm for solving the case of two illuminants. This algorithm is then embedded in a framework which is capable of recoloring complex objects in the presence of shadows and specularities formed by two *unknown* illuminants (e.g, colored-shadows and interreflections) and achieving physically plausible results for uncalibrated natural scene images. As an additional application we show that our framework applied to color transfer, handles complex objects with specularity and under multiple illuminant better than existing methods.

3.2 Related work

Intrinsic images. Several methods have been proposed to compute the intrinsic images of Eq. 3.1 based on various constraints. A common constraint is to assume Lambertian reflectance ($m_s = 0$). For this case, Weiss [102] shows that for an image sequence assuming $\mathbf{c_b}$ to be constant over time, and using the prior that illumination images give rise to sparse filter outputs, estimation of the intrinsic images is achievable. Tappen et al. [92] show that by assuming that shading and reflectance boundaries do not occur at the same location the intrinsic images can be derived from a single image.

Fewer works have concentrated on solving the case where $m_s = 0$. Klinker et al. [55] propose a method where segmentation and intrinsic image estimation are iteratively alternated. Within each segment a single DRM is estimated. Hypotheses of possible illuminant and object colors are verified for the segments and neighboring segments. This method is further extended to include multicolored objects in [71, 72]. The main drawback of these approaches is that they face a chicken-and-egg problem: for a good segmentation you need approximately correct DRM parameters, and vice versa. Furthermore, these methods are only evaluated on high-quality images taken in controlled environments, typically without complex backgrounds, which greatly enlarges the hypothesis space to be checked, and limits the probability of correct convergence.

Several highlight/specularity removal methods have been proposed using the assumption of a *known* illuminant $\mathbf{c_s}$ and that the specular pixels have the same diffuse value as their neighboring diffuse pixels. For example, Robbie Tan et al. [90] proposed an iterative method for reflectance decomposition of textured surfaces. Tan at al. [89] improve the previous methods by adding spacial distribution and texture constraints when available.

Mallick et al. [68] uses partial differential equation that iteratively erodes the specular component at each pixel.

Object recoloring. Many *colorization* methods have also been used for recoloring. They mainly consist of partial hand-coloring of regions in an image or video and propagating the colored points (known as *color markers* or *hot-spots*) to the rest of the image using an optimization algorithm [22, 57]. Since these algorithms are based on the luminance image they lack the additional color information which allows to separate the Lambertian reflectance and specular reflectance, causing them to fail in the presence of specularities.

Color transfer methods extract the color characteristics from a source image and apply it to a target image. Many color transfer methods are based on pixels color distribution [77,78]. Local color transfer [37,103] and user-interactive methods [1] try to improve the results by providing more cues. The main issue of the color transfer is that it requires a target scene, while here we solve the case for which no information about the target distribution is given. Furthermore, these methods are generally applied to matte surfaces and do not consider the presence of specularities.

The recoloring embedded in professional photo-editing applications performs by calculating an offset in the hue and saturation between the source and target colors. The source image is adjusted to produce the desired color [49]. This method is fast and capable of producing realistic results. However, as it ignores the underlying physical reflectance, it fails in the case of colored or multiple illuminant.

Omer et al. [74] present an *image speci c* color representation robust to color distortion and demonstrated a recoloring example for a Lambertian surface. A more physics-based approach, the closest method to our own, is a DRM based color transfer method [88] in which the object (body) color is estimated and transfered between images. And realistic results on lab conditioned high quality images of objects under single *known* illuminant are presented.

Hsu et al. [52] proposed a novel method to estimate the light mixture in a single image illuminated by two lights specified by user while the reflectance is modeled as solely diffuse. The method achieves good results on white balance and light color change.

3.3 Object Reflectance Modeling

In this section, we describe a physics-based reflectance model for object pixels to achieve a high quality recolored image. We begin with an overview of the DRM and then we extend it for the Multi-illuminant case.

3.3.1 Dichromatic Reflection Model (DRM)

According to Shafer, pixel values for a set of points on a single colored surface must lie within a parallelogram in the RGB space, bounded by body reflectance $\mathbf{c_b}$ and the specular reflectance $\mathbf{c_s}$ [86]. Validity of the DRM has been proven for a variety of inhomogeneous dielectric materials commonly observed in natural scenes [93]. In this paper, we assume that color changes can be modelled by a diagonal model, or Von Kries model, which has been proven a sufficient approximation [29]. We indicate the illuminant color by $\mathbf{l}$, and $\mathbf{L} = diag(\mathbf{l})$ is its diagonal matrix representation. In this case the DRM can be written as

$$\mathbf{f} = m_b \mathbf{c}_b + m_s \mathbf{c}_s = m_b \mathbf{c}\mathbf{L} + m_s \mathbf{l} \tag{3.2}$$

where $\mathbf{f}$ is the RGB triple defining the color of every pixel in the object surface, m_b and m_s are the intrinsic images denoting the magnitude of the body and specular reflectance respectively (Fig 3.1). The body reflectance is a multiplication of the *material reflectance* $\mathbf{c}$ and the illuminant according to $\mathbf{c}_b = \mathbf{c}\mathbf{L}$. We assume *neutral interface reflectance*, causing the specular reflectance to have the same chromaticity as the illuminant color $\mathbf{c}_s = \mathbf{l}$. This equation can be divided into intrinsic images and the chromaticity of the object and illuminant in matrix notation according to

$$\mathbf{f} = [m_b(\mathbf{x}) \;\; m_s(\mathbf{x})] \, [\mathbf{L}\,\mathbf{c} \quad \mathbf{l}]^{\mathrm{T}} = \mathbf{M}\,\mathbf{C}^{\mathrm{T}} \tag{3.3}$$

where $\mathbf{x}$ is a vector of $n \times 2$ coordinates, $\mathbf{f}$ is the $n \times 3$ matrix of pixels RGB values, and the intrinsic image matrix $\mathbf{M} = [m_b(\mathbf{x}) \;\; m_s(\mathbf{x})]$ is $n \times 2$ matrix containing intrinsic images. The color characteristics matrix $\mathbf{C} = [\mathbf{L}\,\mathbf{c} \quad \mathbf{l}]$ contains the relevant parameters for scene recoloring. In Section 3.4 we purpose methods to estimate the model parameters.

3.3.2 Multi-illuminant Dichromatic Reflection (MIDR) model

Real-world objects often exhibit body and surface reflection under more than just one illuminant. An example of multi-illuminant scenario is an outdoor scene with blue sky and yellow sun, or a scene with indoor lighting combined with outdoor lighting through a window. Conventional methods often ignore the secondary illuminants present in the scene to simplify the modelling. Here we extend the reflectance model to the Multi-illuminant Dichromatic Reflection model(MIDR) to account for the secondary illuminants. The MIDR for n illuminants is given by

$$\mathbf{f} = [\mathbf{M}^1 \quad \mathbf{M}^n] \left[\mathbf{C}^1 \quad \mathbf{C}^n\right]^{\mathrm{T}} = \mathbb{M}\mathbb{C}^{\mathrm{T}} \tag{3.4}$$

where $\mathbf{M}^n$ contains the intrinsic images regarding the n^{th} illuminant and $\mathbf{C}^n$ is the corresponding color charactristics matrix. Note that the material reflectance $\mathbf{c}$ remains constant for all intrinsic color matrices. Due to the high complexity of the model, in Section 3.5 we solve for a simplified case of the MIDR model.

The dichromatic reflection model has also been extended to include *ambient lighting*. Originally Shafer [86] modelled ambient light as a constant offset over the scene. Later

work improved the modelling [70] and showed that the ambient term results in an object color dependent offset. For the matter of simplification, in this work we assume the ambient illuminant to be negligible.

3.4 Dichromatic Reflection Model estimation

Since the introduction of the DRM multiple approaches to solve this model have been proposed [55, 68, 72, 90]. In this paper, we are interested in solving the DRM for the application of recoloring single colored objects. Users interested in object recoloring work within a photo-editing environment, allowing them to quickly segment the object of interest. This *single-colored object* constraint allows us to fit a more realistic illumination model, allowing the object to be lit by multiple illuminants.

A successful object recoloring algorithm has to face several challenges:

- Uncalibrated images: Photo-editing software users typically work with uncalibrated, compressed images of medium quality and unknown settings. Most previous methods experiment on high quality calibrated images taken in lab conditions [55, 72], and known illumination [68,90]. To handle these lower quality images we propose a *robust estimator*(Section3.4.1).

- Complex color distribution: several existing approaches estimate the illuminant by fitting L and T-shapes to the color distribution [55, 72]. These methods are based on the hidden assumption that the m_b is assumed constant while m_s is changing. In real-world images we often face much more complex distribution which rather form a plane.To tackle this problem we use the *illuminant estimation* described in Section 3.4.2.

- Complex lighting conditions: the objects in real-world images are often lit by multiple illuminants, colored shadows, and interreflections. Ignoring these lighting conditions would make the resulting object recoloring look artificial. Therefore, in Section 3.5, we propose an iterative algorithm to solve for two illuminants.

3.4.1 Robust Body Reflectance Estimation (RBRE).

For the task of body reflectance color ($\mathbf{c_b}$) estimation on medium quality images we propose the Robust Body Reflectance Estimation (RBRE). Since object pixel values of the non-specular part ($m_s = 0$) form a line passing through the origin, fitting a line through these pixels allows us to compute $\mathbf{c_b} = \mathbf{cL}$. The fitting error of an object pixel $\mathbf{x}$ to a line given by the normalized vector $\mathbf{c_b}$ is

$$e(\mathbf{x}) = \left\| \mathbf{f}(\mathbf{x}) - \left((\mathbf{f}(\mathbf{x}))^T \hat{\mathbf{c}}_\mathbf{b} \right) \hat{\mathbf{c}}_\mathbf{b} \right\| \tag{3.5}$$

Although the least squares (LS) orientation estimation would perform well in the case that all pixels belong to the same orientation, in our case in which there are two main orientations ($\mathbf{c_b}$ and $\mathbf{l}$), the LS estimation will mix the two orientations and give a wrong result. In order to avoid that, a *robust estimator* [98] is constructed:

$$e = \int_{\Omega} \rho(e(\mathbf{x}))dx \tag{3.6}$$

In the current work we apply the *Gaussian error norm*:

$$\rho^m(e) = 1 - \exp\left(-\frac{e^2}{2m^2}\right) \tag{3.7}$$

In a robust estimator, large deviations from the model are considered as outliers, and therefore, they are not taken into account very heavily. While LS estimation is very sensitive to outliers. In our application large deviations from the model are mainly due to the mixing of two different directions, $\mathbf{c_b}\mathbf{L}$ and $\mathbf{l}$. The error, Equation 3.6, can now be rewritten as (we will omit the spatial arguments):

$$e = \int_{\Omega} \rho^m\left(\sqrt{\mathbf{f^T f} - \hat{\mathbf{c}}_\mathbf{b}^\mathbf{T}(\mathbf{f}\mathbf{f^T})\hat{\mathbf{c}}_\mathbf{b}}\right)dx \tag{3.8}$$

A Lagrange multiplier is then used for minimization subject to the constraint $\hat{\mathbf{c_b}}^\mathbf{T}\hat{\mathbf{c_b}} = 1$,

$$\frac{d}{d\hat{\mathbf{c_b}}}\left(\lambda\left(1 - \hat{\mathbf{c_b}}^\mathbf{T}\hat{\mathbf{c_b}}\right) + e\right) = 0 \tag{3.9}$$

Using Equation 3.7 as the error function leads to

$$\eta(\hat{\mathbf{c_b}})\hat{\mathbf{c_b}} = \lambda\hat{\mathbf{c_b}} \tag{3.10}$$

where η is defined according to

$$\eta(\hat{\mathbf{c_b}}) = \int_{\Omega} \mathbf{f}\mathbf{f^T} G^m\left(\sqrt{\mathbf{f^T f} - \hat{\mathbf{c_b}}^\mathbf{T}(\mathbf{f}\mathbf{f^T})\hat{\mathbf{c_b}}}\right)dx \tag{3.11}$$

The main difference with the ordinary LS estimator is that here the matrix η is dependent on $\mathbf{c_b}$. Eq 3.10 can be solved by a *xed point* iteration scheme. We start iteration with the initial estimate $\hat{\mathbf{c_b}}^\mathbf{0}$ given by the LS. Let $\hat{\mathbf{c_b}}^\mathbf{i}$ be the orientation vector estimate after i iterations. The estimate is updated as the eigenvector $\hat{\mathbf{c_b}}^{\mathbf{i+1}}$ of the matrix $\eta(\hat{\mathbf{c_b}}^\mathbf{i})$ corresponding to the largest eigenvalue, i.e. we solve

$$\eta(\hat{\mathbf{c_b}}^\mathbf{i})\hat{\mathbf{c_b}}^{\mathbf{i+1}} = \lambda\hat{\mathbf{c_b}}^{\mathbf{i+1}} \tag{3.12}$$

Again, points far away from the line direction $\hat{\mathbf{c_b}}$ are considered outliers, and therefore, do not corrupt the estimation. Iterative application of Equation 3.12 yields the estimate of the body reflection, $\hat{\mathbf{c}}_b$. The original estimation made by ordinary LS is refined at each iteration by changing the weights leading the method to converge to a robust, and in this case a much better, estimation of the $\hat{\mathbf{c}}_\mathbf{b}$.

Figure 3.2: An example of intrinsic images recovered for an object. (a) Original image; Intrinsic images: (b) Body reflectance and (C) Specular reflectance; (d) An example recoloring result.

3.4.2 Con ned illuminants estimation (CIE)

Having the body reflectance color, there exists a set of possible illuminants which could generate the color distribution of the object. Many of these illuminants are unrealistic. It is shown that the chromaticity of common light sources closely follows the Planckian locus of black-body radiators [33]. We propose to use this constraint to estimate the illuminant.

We sample Planckian colors ($T \subset 1000 \sim 40000$) which vary from orange to yellow to white to blue, resulting in a set of illuminants $\mathbf{l}_1 \quad \mathbf{l}_m$ .We define the reconstruction error of the intrinsic images $\mathbb{M}$ and intrinsic color characteristics $\mathbb{C}$ by

$$E_r\left(\mathbf{f}\ \mathbb{M}\ \mathbb{C}\right) = \left(\mathbf{f} - \mathbb{M}\mathbb{C}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{f} - \mathbb{M}\mathbb{C}^{\mathrm{T}}\right) \tag{3.13}$$

Then, we perform an exhaustive search to find the best matching Planckian light. In other words, we solve Equation 3.14 by choosing the Planckian light, which minimizes the reconstruction error.

$$\hat{\mathbf{l}} = \underset{\mathbf{l}\in\{l_1 \quad l_m\}}{\arg\min}\ E_r\left(\mathbf{f}\ \mathbb{M}\ [\mathbf{cL}\ \mathbf{l}]\right) \tag{3.14}$$

In the next section we will outline the computation of the intrinsic images $\mathbb{M}$ given $\mathbb{C}$, which are needed for the computation of the reconstruction error.

3.4.3 Intrinsic images

The estimation of the intrinsic images, given an estimation of $\mathbb{C}$, is based on the convex optimization problem:

$$\begin{array}{l} \underset{M}{\text{minimize}}\ E_r\left(\mathbf{f}\ \mathbb{M}\ \mathbb{C}\right) \\ \text{subject to } m_b\left(\mathbf{x}\right) \geq 0\ \ m_s\left(\mathbf{x}\right) \geq 0 \end{array} \tag{3.15}$$

Fig 3.2 demonstrates an example of intrinsic images recovered for an object. Note that the specular reflectance is correctly separated from the body reflectance.

3.5 Two-illuminant MIDR model estimation

Many real-world objects are lit by multiple illuminants. Here we propose an algorithm to estimate the case of two illuminants. Since the problem is highly underconstraint, we need further assumptions: Firstly, we assume one illuminant to be Planckian and demonstrate specularities; Secondly, specularities of the secondary illuminant to be negligible. We use this as an additional constraint ($m_s^2(\mathbf{x}) = 0$). Note that we make no assumption on the chromaticity of the secondary illuminant. Hence the model is given by

$$\mathbf{f} = m_b^1 \mathbf{c}\ \mathbf{L^1} + m_s^1 \mathbf{l^1} + m_b^2 \mathbf{c}\ \mathbf{L^2} \tag{3.16}$$

An iterative algorithm to solve this MIDR model is given in Algorithm 1. First we will assume pixels to be illuminated by only one of the two illuminants $m_b^1(\mathbf{x})m_b^2(\mathbf{x}) = 0$ and $m_s^1(\mathbf{x})m_b^2(\mathbf{x}) = 0$. In the final Step we remove this restriction to allow for pixels being lit by both illuminants at the same time. Here, we also use the $diag$-function to convert vectors to diagonal matrices and vice versa. First an initial estimation is made based on all pixels on the object (Steps 1-4) which gives us the initial values for the dominant illuminant and object color. Based on this model pixels which could be described by this model with affordable error are separated from the rest (Step 5) which are indicated by the *Mask*. At each iteration the estimations and seperation mask are refined. We estimate a Lambertian reflectance model for the pixels outside the Mask (Steps 10 and 11). Iteratively the illuminant color estimations are refined until convergence (Step 13). The final model estimation is then given by the object material reflectance color $\mathbf{c}$, the two illuminant colors $\mathbf{l_1}$ and $\mathbf{l_2}$, and the corresponding intrinsic images m_b^1, m_b^2, and m_s.

Although the algorithm gives good estimates for $\mathbf{c}$, l^1 and l^2, the constraint that pixels can only be illuminated by a single illuminant results in artificial edges in the m_b^1 and m_b^2 estimates. In reality there are regions where both lights illuminate the object. To solve this, Step 14 finalizes the algorithm by keeping $\mathbf{c}$, l^1 and l^2 and m_s^1 constant in Eq 3.16, and estimates m_b^1 and m_b^2 constraining them to be positive.

In Fig 3.3 we show the results of the algorithm on an outdoor car image. The car is illuminated by a white outside lighting as well as a greenish light caused by the light coming from the grass field. The mask is given for several iterations of the algorithm. The algorithm correctly separates the two illuminants. In the last row the intrinsic images show the estimates of the body and specular reflection.

3.6 Experimental results

In the experimental section we analyze our proposed algorithm for MIDR estimation on synthetic images. Additionally we show some results on challenging real-world images. Here we assume images are in sRGB format; and do *gamma correction*. Further applications of the model are discussed in the end of the section.

Algorithm 1 Two-illuminant MIDR model estimation

1: Consider the whole object segment as $Mask$
2: Estimate $\mathbf{c_b}$ using RBRE for the pixels $\mathbf{x} \quad Mask$
3: Estimate the Planckian illuminant $\mathbf{l_1}$ using CIE method
4: $\mathbf{c} \quad diag\left(\mathbf{c_b}\ {\mathbf{L_1}}^{-1}\right)$
5: Iniciate $Mask$ to only include the pixels $\mathbf{x}$ for which $E_r(\mathbf{f}(\mathbf{x})\ \mathbf{M}^1\ \mathbf{C}^1) < Threshold$
6: **repeat**
7: Estimate $\mathbf{c_b^1}$ using RBRE for the pixels $\mathbf{x} \quad Mask$
8: Estimate the Planckian illuminant $\mathbf{l_1}$ using CIE method
9: $\mathbf{c} \quad diag\left(\mathbf{c_b^1}\ {\mathbf{L_1}}^{-1}\right)$
10: Estimate $\mathbf{c_b^2}$ using RBRE for the object pixels $f(\mathbf{x}) \quad Mask$
11: $\mathbf{L_2} \quad diag(\mathbf{c}_b^2)\ diag(\mathbf{c})$ (using the $\mathbf{c}$ from Step 9).
12: Update $Mask$ to only include the pixels $\mathbf{x}$ for which $E_r(\mathbf{f}(\mathbf{x})\ \mathbf{M}^1\ \mathbf{C}^1) < E_r(\mathbf{f}(\mathbf{x})\ \mathbf{M}^2\ \mathbf{C}^2)$
13: **until** $\mathbf{L_1}$ and $\mathbf{L_2}$ estimates converge
14: Recalculate the ${m_b}^1$ and ${m_b}^2$ using the previous estimates for $\mathbf{c}\ \mathbf{l_1}\ \mathbf{l_2}$ and ${m_s}^1$.

Please refer to the supplementary video for more examples of the results in real-world images.

3.6.1 Synthetic Images

Here we test our algorithm on synthetic images which satisfy the assumptions, namely they are lit by two lights, one of which is Planckian. The groundtruth intrinsic images m_b^1, m_b^2 and m_s^1 are given (Fig 3.4). With these we generate a set of 60 test images by varying the illuminants and the object color. Some examples are given in Fig 3.4. The soundness of our algorithm has been verified on synthetic test images on which the intrinsic image estimation performs with an error close to zero even though a large part of the object is lit by both lights simultaneously.

Since we want to apply our method to standard Internet images, we further investigated its robustness to both Gaussian noise and JPEG compression (Fig 3.6). The comparison is made using the *Angular Error* (E_a) in radians between the ground-truth ($\hat{\boldsymbol{c}}_{\boldsymbol{gt}}$) and estimated ($\hat{\boldsymbol{c}}_{\boldsymbol{est}}$) colors as defined below,

$$E_a = \arccos(\hat{\boldsymbol{c}}_{\boldsymbol{gt}} \cdot \hat{\boldsymbol{c}}_{\boldsymbol{est}}) \tag{3.17}$$

As can be seen the algorithm is sensitive to Gaussian noise but relatively robust to JPEG compression (angular error of all estimations for 60% compression is under 0.07 radian). To better interpret the results in the graphs we also provide the reconstruction results on one synthetic object for several noise and JPEG compression settings in Fig 3.5.

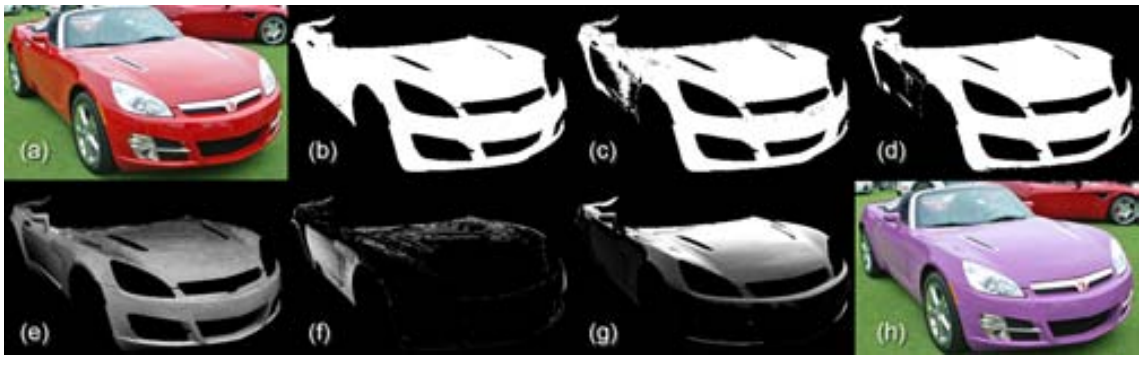


Figure 3.3: An example of the MIDR algorithm performance: (a) Original image; (b) Object mask as the initial *Mask* for the illuminant l_1; (c) The Mask after 1st iteration; (d) The Mask at 3rd (final) iteration; (e) estimated m_b^1; (f) estimated m_b^2 (the interreflection area,l_2, has been correctly detected); (g) estimated m_s^1; (h) An example recoloring (the interreflection is preserved).

Figure 3.4: The first four images are examples of the synthetic images. The last three images are the m_b^1, m_b^2,and m_s^1 ground truth.

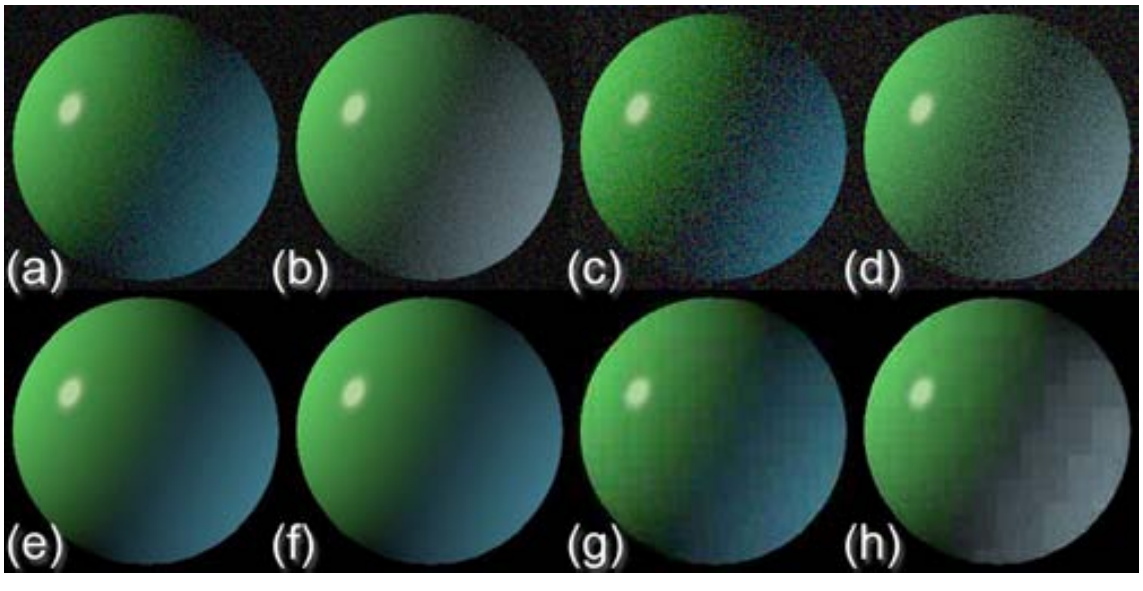


Figure 3.5: Effect of noise and JPEG compression: (a) and (c) examples of applying noise by sigma 4.0 and 9.0; (b) and (d) their corresponding reconstructions; (e) and (g) examples of applying JPEG compressions of 20% and 80%; (f) and (h) their corresponding reconstructions.

3.6.2 Real-world Images

Fig 3.8 compares MIDR-based recoloring with the one done by hue-saturation shift method. The secondary illuminant (greenish interreflection) is correctly preserved by MIDR while

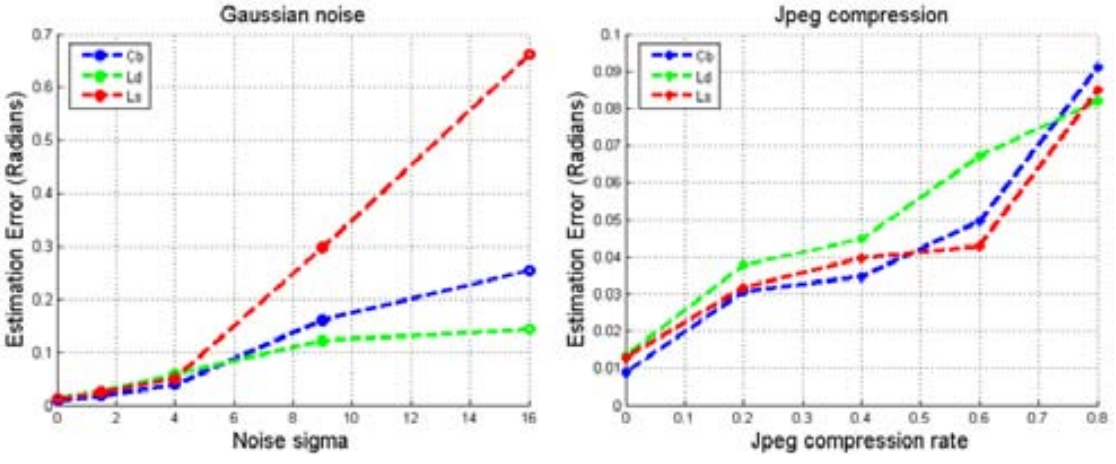


Figure 3.6: Median angular error (in radian) as a function of: Gaussian noise sigma (left) and JPEG compression (right) for c_b, l_1 and l_2 estimates.

wrongfully changed to blue by the professional photo-editor. In Fig 3.9 the MIDR and DRM has been compared for the accuracy of their recoloring results. The secondary illuminant (bluish shadow) is well preserved by MIDR while lost in the case of DRM. Note that here we only modeled two illuminants and therefore the third illuminant (the small brownish interreflection on the back of the car) is lost.

Note that theoretically the method fails to correctly make the intrinsic image decomposition in the case object and illuminant colors are collinear. Also having no Planckian illuminant confuses the CIE estimator. The latter is shown in the example of Fig 3.7.

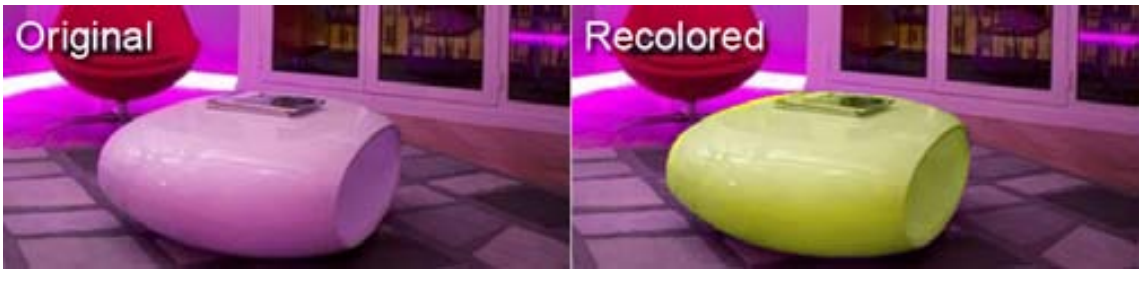


Figure 3.7: An example failure case: Here the planckian light assumption is violated by having a purple light. Since purple is not Planckian, the method wrongfully picked white as the illuminant and purple as the object color. The recoloring shows that even though the object itself looks realistic it does not match the scene illumination.

3.6.3 Other Applications of MIDR

Here we show two other interesting applications for the proposed framework, namely *Physics-based Color Transfer* and *Photo Fusion*.

Physics-based Color Transfer. A popular photo-editing task is transferring the color characteristics of an image to another. Even though color transfer methods are often

Figure 3.8: Comparing the MIDR method performance with a professional photo-editor: (a) Original image (containing complex interreflection); (b) Recoloring result by MIDR (the secondary illuminant,green interreflection, has been preserved); (c) Recoloring result using the hue-saturation shift method (the green interreflection is wrongfully changed to blue).

Figure 3.9: Comparing the methods based on MIDR and DRM: (a) Original image; (b) Recoloring result by MIDR (zoomed area: blue shadows have been preserved); (c) Recoloring result using DRM (missed the colored-shadows).

successful in transferring the atmosphere of one image onto the other, they make unrealistic assumptions (e.g, Gaussian distribution, Lambertian objects). These shortcomings become apparent when applied to the object color transfer. Fig 3.10 compares the physics-based color transfer performed using MIDR and DRM models with the methods from [77, 78]. We apply the color transfer only to the presegmented objects. After inferring the object color and two illuminants, MIDR successfully transfers the object color. Note that the methods of [77, 78] mixing the illuminants and object colors resulted in unrealistic images. Furthermore, the resulting objects exhibit different colors than the target objects.

Figure 3.10: Comparing the *Color Transfer* results by DRM, and [77, 78]. (a) and (f) Original images;(b) and (g) MIDR results; (c) and (h) DRM results; (d) and (i) results by [78]; (e) and (j) results by [77] . Note that the secondary illuminants (interreflections) on the side of the car and the plane wing are lost in (c) and (h), wrongfully transformed in (d), (e),(i), and (j), while being preserved in (b) and (g).

Figure 3.11: An example of photo montage: The interreflection of the green grass (zoomed area) in the original image is re-lighted by the red color of the carpet to match the target scene.

Photo Fusion. Fig 3.11 is an interesting example made possible by our method. The car in Fig 3.3 is copied into another scene. Here the object is recolored using the estimated intrinsic images. But to match the target scene, the interreflection caused by the grass is re-illuminated using the color of the carpet simply by changing the second illuminant color to the red of the carpet resulting in a more realistic scene where the red carpet is reflected in the side of the car.

3.7 Conclusion and future work

We have presented a method for recoloring single-colored objects based on intrinsic image estimation. The single-color constraint allows us to fit more complex reflectance models which better describe real-world images. Whereas most existing recoloring methods assume a white illuminant, we presented a method to recolor objects taken under colored illuminants, and the more complex case of multiple illuminants. Results on synthetic images demonstrate that our algorithm correctly estimates the intrinsic parameters of the scenes. Further we show that the proposed method is able to achieve physically realistic recoloring results in challenging real-world images. In addition we present how our method improves other photo-editing applications like *Color Transfer* and *Photo Fusion*.

As future research, we will investigate further extensions of the dichromatic reflection model, such as the bi-illuminant reflection model recently proposed by Maxwell [70]. This model allows for the modeling of ambient light which we believe could improve the quality of the recoloring for the low luminance regions of the image.

Chapter 4

Multi-Illuminant Estimation with Conditional Random Fields

Most existing color constancy algorithms assume uniform illumination. However, in real-world scenes, this is not often the case. Thus, we propose a novel framework for estimating the colors of multiple illuminants and their spatial distribution in the scene. We formulate this problem as an energy minimization task within a Conditional Random Field over a grid of local illuminant estimates. To quantitatively evaluate the proposed method, we created a novel dataset of two-dominant-illuminants images comprised of laboratory, indoor and outdoor scenes. Unlike prior work, our database includes accurate pixelwise ground truth illuminant information. The performance of the method is evaluated on multiple datasets. Experimental results show that our framework clearly outperforms single illuminant estimators, as well as a recently proposed multi-illuminant estimation approach.

4.1 Introduction

The vast majority of existing color constancy algorithms are based on the assumption that there exists a single illuminant in the scene. Many images, however, exhibit a mixture of illuminants with distinct chromaticities. Consider for example indoor scenes which are lit by both indoor light sources and outdoor light coming through the windows. Or an outdoor scene, where parts of the image are in direct sunlight, while others are in shadow which is illuminanted by the blue skylight. Another example where single illuminant white balancing is known to give unsatisfactory results, is in pictures which are taken using a camera-flash. Illuminant estimation methods that assume uniform illumination cannot accurately recover the illuminant chromaticity and its variations across such scenes. Examples of multi-illuminant pictures, and the color-coded pixel-wise influence of each illuminant, can be seen in Fig. 4.1.

Figure 4.1: Sample images from our database. The bottom row shows the relative influence of the two incident illuminants color-coded in blue and red.

Extending existing color constancy methods to successfully compute multi-illuminant estimates is a challenging problem. Consider two of the most popular branches of existing color constancy approaches: statistics-based methods and physics-based ones. The success of statistics-based techniques [16, 31, 47, 96] depends on the size of the statistical sample. Applying these methods to small image regions introduces inaccuracies [13] and is unlikely to yield stable results. Physics based methods either assume purely diffuse scenes, e.g. [15, 41] which is not often applicable in real scenes, or exploit the presence of specularities in an image, e.g [91], which occur very sparsely in an image. As a result, a direct extension of global (image-wide) color constancy methods to region-based ones is likely insufficient. Spatial constraints between the estimates will be required to obtain acceptable results.

We propose a multiple illuminant estimation method which first extracts local estimates. We overcome the inherent instability of local measurements by globally solving the illuminant labelling problem by means of a Conditional Random Field (CRF). We prove that several existing approaches, namely statistics- and physics-based methods, can be written in the form of a CRF. The CRF formulation provides a natural way to: a) combine various approaches into a single multi-illuminant estimate and b) incorporate spatial information about the illuminant distribution. We show that representing these methods by such a model allows us to robustly extend them to multi-illuminant estimation. Furthermore, we created a new database for multi-illuminant color constancy with highly accurate, computationally extracted (instead of manually annotated) pixelwise ground truth. Our database contains: a) laboratory images, for evaluation under close-to-ideal conditions and b) real-world multi-illuminant scenes, which more closely approximate real-world scenarios.

In summary, the main contributions of this paper are:

- The formulation of multi-illuminant estimation as a CRF model.
- The expression of existing bottom-up approaches to color constancy as an energy minimization problem.
- The creation of a new dataset for multi-illuminant estimation.
- An extensive experimental evaluation which shows that the proposed method addresses the intrinsic challenges in multi-illuminant scenes, i.e. the estimation of the illuminant colors and their spatial distribution, with superior accuracy compared to prior work.

The paper is organized as follows. In Sec. 4.2, we present related work in color constancy, and in particular methods that have been used as a foundation for the proposed approach. The theoretical foundation of the proposed framework is introduced in Sec. 4.3. Unary and pairwise potentials for the CRF are derived in Sec. 4.4 and Sec. 4.5, respectively. In Sec. 4.6, we outline the overall algorithm for color constancy under non-uniform illumination. Section 5.4 contains a description of the new multi-illuminant dataset, and a derivation of the ground truth extraction. Experimental results are presented and discussed in Sec. 4.8. We summarize the findings of this work in Sec. 4.9. Additional mathematical details are presented in the appendix.

4.2 Related Work

4.2.1 Single-illuminant Estimation

Stastistics based color constancy methods derive the estimate of the illuminant color from assumptions on the statistics of reflectances in the world. The grey-world algorithm [16] is the most well-known method of this family, and computes the illuminant of a scene by assuming that the average scene reflectance is grey. Another popular method is the MAX-RGB algorithm which computes the illuminant in a scene from the maximum responses in the RGB channels [59]. It was noted by Gershon *et al.* [40] that it is often beneficial to assume that the average of a scene is equal to the average reflectance of a database. Finlayson and Trezzi [31] showed that both the grey-world and the MAX-RGB algorithms are instantiations of the more general shades-of-grey method which estimates the illuminant of images by computing the Minkowski norm of the pixels in a scene. Van de Weijer *et al.* [96] further extended this theory to also include image derivatives. Finally, Gijsenij *et al.* [47] showed that weighting edges according to their physical cause (shadow, specularity, or material transition) can further improve results.

In comparison physics-based methods exploit the interaction between light and material to infer the illuminant color in an image. Some methods e.g. [15, 41] assume the

scene is entirely composed of diffuse surfaces, while others e.g. [61, 91] exploit the presence of specular highlights. These latter methods are based on the dichromatic reflection model [87] which models the reflected light as a combination of diffuse and specular reflectance. Based on the assumption of neutral interface reflection, the color of the specular reflectance is the same as the illuminant color and therefore an important cue for color constancy (see, e.g., [87, 91]).

Gamut based methods exploit the fact that only a limited set of RGB values can be observed under a known canonical illuminant. This set of RGB values can be represented by a canonical convex hull in RGB space [34]. Thus, feasible illuminants can be estimated by computing all possible mappings from a single image's convex hull to the canonical convex hull. The scene illuminant is heuristically selected from the feasible illuminants. This method was further extended by Finlayson *et al.* [26] by constraining the possible illuminants to be on the Planckian locus. Gijsenij *et al.* [44] extended this theory to higher-order derivative structures of the images.

For a more complete overview of color constancy, see e.g. the recent overview articles [36, 46, 64].

4.2.2 Multi-illuminant Estimation

There are illuminant estimation methods explicitly designed to handle varying illumination. In 1997, Barnard *et al.* [5] were the first ones to develop a methodology that automatically detects non-uniform illumination. They then proceeded with removing the illumination variation, at which point they could apply any gamut-based color constancy method. Though this method was pioneering at that time, its smooth illumination assumption restricts its applicability on real-world images. Ebner [23] followed a different approach of applying a diffusion-based technique on pixel intensities. However, he too assumes a smoothly varying illumination, which together with his underlying theory of regional grey-world can result in inaccuracies, especially in colorful scenes [51]. More recently, Kawakami *et al.* [54] proposed a physics-based method specifically designed to handle illumination variations between shadowed and non-shadowed regions in outdoor scenes. Due to its explicit assumption of hard shadows and sky-light/sunlight combination (or even more general Planckian illuminants), this method does not generalize well on arbitrary images. Gijsenij *et al.* [45] recently proposed an algorithm for scenes with two light sources . The reported experimental results are promising. However, it is not clear how to extend this methodology for non-local illuminant cues. When the chromaticity of the two incident illuminants is known, Hsu *et al.* [52] proposed an algorithm for high quality white-balanced images. However, their assumption of two known illuminants limits the applicability of the method to close-to laboratory conditions. Thus, by construction, none of the existing multi-illuminant estimation methods can handle arbitrary images and as such, none of them has been extensively tested on a large variety of real-world images.

4.3 Methodology

As discussed in the introduction, the single illuminant assumption, which is the basis of many existing color constancy algorithms, is often unrealistic. Quite frequently multiple illuminants are present in a scene. In such cases, the spatial distribution of the illumination conditions becomes very important. We propose to solve the multiple illuminant estimation problem by using a Conditional Random Field (CRF) framework. The nodes in the graph represent patches, the labels correspond to illuminant colors, and the edges connect neighboring patches. In such a representation local illuminant estimation becomes equivalent to finding the maximum a posteriori (MAP) labelling of the CRF. Such a framework facilitates both the local computation of illuminant color, as well as the incorporation of spatial information about the distribution of illuminants.

More specifically, a conditional random field can be viewed as an undirected graph model, globally conditioned on observations. Let $\mathcal{G} = (\mathcal{V}\ \mathcal{E})$ be a graph where $\mathcal{V} =$ $1\ 2 \quad N$ is the set of nodes representing the N patches and $\mathcal{E}$ is the set of edges connecting neighboring patches. We define a discrete random field $\mathcal{X}$ over the graph $\mathcal{G}$. Each node $i \quad \mathcal{V}$ is associated with a random variable $\mathbf{X}_i \quad \mathcal{X}$, which can take on a value $\mathbf{x}_i$ from the illuminant-color label set $\mathcal{L} = \ \mathbf{l}_1\ \mathbf{l}_2 \quad \mathbf{l}_k\ .$ At each node $i \quad \mathcal{V}$ we also have a local observation $\mathcal{F}_i$, which is the set of $(R\ G\ B)$ values of all the pixels belonging to the corresponding patch together with their spatial distribution. The probability $P(\mathcal{X} =$ $\breve{\mathbf{x}}\ \mathcal{F})$ of a particular labelling $\breve{\mathbf{x}} = \ \mathbf{x}_1\ \mathbf{x}_2 \quad \mathbf{x}_N$ conditioned on the observations $\mathcal{F}$ of the entire image will be denoted as $P(\breve{\mathbf{x}}\ \mathcal{F})$. Then according to the Hammersley-Clifford theorem

$$P\left(\breve{\mathbf{x}}\ \mathcal{F}\right) \quad \exp \quad - \quad \sum_{c \in C} \xi_c(\breve{\mathbf{x}}_c\ \mathcal{F}) \tag{4.1}$$

where $\xi_c(\breve{\mathbf{x}}_c\ \mathcal{F})$ are potential functions defined over the observations $\mathcal{F}$ and the variables $\breve{\mathbf{x}}_c = \ \mathbf{x}_i\ i \quad c$ belonging to clique c. A clique c is a set of random variables $\mathbf{X}_c$ which are conditionally dependent on each other and C is the set of all cliques in $\mathcal{G}$. Finding the labelling $\breve{\mathbf{x}}^*$ with the maximum a posteriori (MAP) probability $\breve{\mathbf{x}}^*$ is then equal to

$$\breve{\mathbf{x}}^* = \arg\max_{\breve{\mathbf{x}} \in \mathbf{L}} P\left(\breve{\mathbf{x}}\ \mathcal{F}\right) = \arg\min_{\breve{\mathbf{x}} \in \mathbf{L}} E\left(\breve{\mathbf{x}}\ \mathcal{F}\right) \tag{4.2}$$

where $\mathbf{L}$ is the set of all possible labellings on $\mathcal{X}$ and $E(\breve{\mathbf{x}}\ \mathcal{F})$ is the corresponding Gibbs energy defined as

$$E\left(\breve{\mathbf{x}}\ \mathcal{F}\right) = \sum_{c \in C} \xi_c(\breve{\mathbf{x}}_c\ \mathcal{F}) \tag{4.3}$$

Hence, computing the MAP labelling is equal to finding the labelling which minimizes the energy $E(\breve{\mathbf{x}}\ \mathcal{F})$. In our case, this means that obtaining the MAP assignment of

illuminants to patches can be accomplished by finding that assignment which minimizes the corresponding Gibbs energy. Considering only up to pairwise clique potentials, the energy function becomes:

$$E\left(\breve{\mathbf{x}}\ \mathcal{F}\right) = \sum_{i \in \mathcal{V}} \lambda\left(\mathbf{x}_i\ \mathcal{F}_i\right) + \theta_p \sum_{(i\ j) \in \mathcal{E}} \left(\left(\mathbf{x}_i\ \mathbf{x}_j\right)\left(\mathcal{F}_i\ \mathcal{F}_j\right)\right) \tag{4.4}$$

where λ denotes the unary potential and the pairwise potential. The unary potentials λ penalize the discrepancy between the observations, i.e. the colors of the pixels in a patch $\mathcal{F}_i$, and the solution, i.e. the illuminant-color label assigned to the patch. The pairwise potentials provide a definition of smoothness by penalizing changes in the labels of neighboring patches. Then the constant $\theta_p > 0$ controls the balance between smoothness and data fit. In the next section we propose various unary potentials which allow us to represent several well-known illumination estimation algorithms as CRFs.

4.4 Unary Potentials

We show that by choosing a particular unary potential we can express several existing color constancy methods as an error minimization problem. When we use a pairwise potential function that enforces a single label for all patches, we obtain the same result as traditional single illuminant estimation methods. Reducing the influence of the pairwise potential results in multi-illuminant estimates for the scene.

4.4.1 Statistics-based Color Constancy

There exists a family of color constancy methods which is based on the statistics of reflectances in the world. Examples of this group of methods are grey-world, grey-edge and max-RGB algorithm [16,31,96]. We show that several of these algorithms can be written as an error minimization problem.

We denote $\mathbf{f}^j = (f_R^j\ f_G^j\ f_B^j)^{\mathrm{T}}$ to be the j-th pixel in an image. We assume that an image is segmented into a number of patches $P = p_1\ p_2 \ldots p_N$ where p_i contains the indices to the pixels in patch i. From the set of observations $\mathcal{F}_i$ in a patch we can obtain an estimate of the local illuminant color $\mathbf{i}(\mathcal{F}_i)$, which, for conciseness, we will denote as $\mathbf{i}_i$. If the estimate is computed with the grey-world algorithm, then the local illuminant color is determined by the average color in the patch, as defined by

$$\mathbf{i}_i = \frac{\sum_{j \in p_i} \mathbf{f}^j}{\left\| \sum_{j \in p_i} \mathbf{f}^j \right\|} \tag{4.5}$$

where $\|\cdot\|$ is the L_2 norm which is applied to ensure that $\mathbf{i}_i$ has unit length. Illuminant estimation methods are generally evaluated based on the angular error, which for two

normalized illuminants (typically the estimated and true illuminant) is given by

$$(\mathbf{i_1}\ \mathbf{i}_2) = \arccos\left((\mathbf{i}_1)^{\mathrm{T}}\mathbf{i}_2\right) \tag{4.6}$$

We now define the statistics-based unary potential λ^s, which defines the cost for patch i to take on illuminant $\mathbf{x}_i$ as

$$\lambda^s(\mathbf{x}_i\ \mathcal{F}_i) = w^i \rho^\theta\big(\ \ (\mathbf{i}_i\ \ \mathbf{x}_i)\big) \tag{4.7}$$

where w_i is a scalar weight per patch, and ρ is the error norm. For example, choosing $\rho(e) = e^2$ yields the least squares error. The influence of outliers on the unary potential can be reduced by choosing a robust error norm. We discuss several choices of error norms and weights below.

Choosing as an error norm of $\rho(e) = 1 - \cos(e)$ and for the weights per patch the summed intensity of its patches $w^i = \sum_{j\in p_i} \mathbf{f}^j$, we obtain the following unary potential

$$\lambda^s(\mathbf{x}_i\ \mathcal{F}_i) = \Big\| \sum_{j\in p_i} \mathbf{f}^j \Big\| \Big(1 - \cos\big(\ \ (\mathbf{i}_i\ \ \mathbf{x}_i)\big)\Big) \tag{4.8}$$

When the illuminant given by the label, $\mathbf{x}_i$, and the illuminant derived directly from the observations are equal this unary potential is zero. When they are maximally different this unary potential is equal to the summed intensity of the patch. In Appendix A we include a proof showing that this particular unary potential leads to the standard grey-world solution when we enforce a single illuminant label for all patches in Eq. 4.4.

We can also use the more general class of statistics based illuminant estimation [96], given by

$$\mathbf{i}_i^{n\,m} \approx \sqrt[m]{\sum_{j\in p_i} \left| \frac{{}^n\mathbf{f}^j_{\sigma_{\mathrm{GW}}}}{\mathbf{x}^n} \right|^m} \tag{4.9}$$

where n is the differentiation power, and m is the Minkowski norm. σ_{GW} denotes the standard deviation of a Gaussian smoothing operator that is applied to the image prior to the differentiation. Depending on the choice of parameters m and n the estimate is equal to the grey-world, shades of grey, or grey-edge algorithm. As the unary potential for the general case we propose

$$\lambda^s(\mathbf{x}_i\ \mathcal{F}_i) = \left\| \sqrt[m]{\sum_{j\in p_i} \left| \frac{{}^n\mathbf{f}^j_{\sigma_{\mathrm{GW}}}}{\mathbf{x}^n} \right|^m} \right\| \Big(1 - \cos\big(\ \ (\mathbf{x}_i\ \ \mathbf{i}_i)\big)\Big) \tag{4.10}$$

For $n = 1$ and $m = 1$, minimizing Eq. 4.4 with this unary potential results in the standard grey-edge algorithm [96].

We proceed by proposing several adaptations to the unary potential to adapt it for multi-illuminant estimation. If we increase the influence of the pairwise potential, by choosing a large θ_p in Eq. 4.4, we can enforce the whole image to have the same label,

and therefore the same estimate for the illuminant. There are several choices for error norm and patch weight which in this case lead to well-known color constancy algorithms. If we look at the other extreme where we pick $\theta_p = 0$ every patch would take on the label of the illuminant which is closest (in an angular error sense) to its local estimate. However, the local estimates of the statistical color constancy algorithms are very noisy and in general this will lead to unsatisfying results. This can be countered by choosing an intermediate θ_p (by means of cross validation), that enforces multiple neighboring patches to take on the same label, and thereby reducing the noise of the statistical estimate. We will look at two additional adaptations to the unary potential which improve robustness with respect to noisy statistical measurements.

Robust error norm: To reduce the influence of outliers on the energy, we found the usage of a robust error norm indispensable. Throughout the paper we use the following error norm

$$\rho_{\sigma_r}(e) = 1 - \exp\left(-\frac{e^2}{2\sigma_r^2}\right) \tag{4.11}$$

Its main effect is that outliers have less influence on the overall energy. Using robust error norms in a CRF formulation has been found beneficial before.

Uneven color balance: Statistical methods are known to be biased towards large segments of the same color. To counter this we propose the following adaptation:

$$\lambda^s(\mathbf{x}_i\ \mathcal{F}_i) = \left(w^i\right)^q \rho_{\sigma_r}(\quad(\mathbf{i}_i\ \mathbf{x}_i)) \tag{4.12}$$

The parameter q allows to dampen the results of uneven color balance in the image. Consider the standard grey-world assumption ($p = 1$ and $n = 0$) if we then choose $q = 0$, the unary potential is equal to

$$\lambda^s(\mathbf{x}_i\ \mathcal{F}_i) = \left(1 - \cos\left(\quad(\mathbf{x}_i\ \mathbf{i}_i)\right)\ \right) \tag{4.13}$$

which is one of the more popular implementation of grey-world where instead of each pixel, one value for each patch is chosen. This was also proposed by Barnard et al. [6] to counter the dominance of large uniformly colored regions in images on the outcome. In the results we consider $q \quad \left\{0\ \frac{1}{2}\ 1\right\}$.

4.4.2 Physics-based Color Constancy

Another family of color constancy methods is based on the physics of light and surface interactions [15, 41, 61, 91]. In this work we focus on the approach by Tan *et al.* [91] because it is very competitive performance-wise and is applicable to a wider family of surfaces that exhibit a mixture of diffuse and specular reflectance. More specifically, we follow the extension by Riess *et al.* [79] which can be applied to local regions and even in patches that are just moderately specular.

Specularity based approaches follow the neutral interface assumption which states that the color of pure specularities is the color of the illuminant. In general, these approaches

are comprised of two steps: a specularity detection step where possible regions which contain specularities are identified, and an illuminant estimation step based on the color of the detected specular regions.

The method proposed by Tan *et al.* [91] exploits the inverse-intensity chromaticity (IIC) space. IIC is a two-dimensional space where the horizontal axis represents the inverse intensity $1 \quad \mathbf{f}^j \ _1$ and the vertical axis is given by a pixel's chromaticity. Thus, a pixel $\mathbf{f}^j = (f_R^j \ f_G^j \ f_B^j)^T$ is mapped to

$$f_C^j \qquad \frac{1}{\mathbf{f}^j \ _1} \ \frac{f_C^j}{\mathbf{f}^j \ _1} \tag{4.14}$$

where $\cdot \ _1$ is the L_1-norm and $C \quad R \ G \ B$.

The advantage of IIC is that the relationship between image chromaticity and illumination chromaticity becomes linear. According to [91], one generates per color channel a scatterplot of the pixels in inverse-intensity chromaticity space. In IIC space purely diffuse pixels of the same material and albedo form a horizontal cluster. Pixels of the same material and albedo but with a specular component form a triangular-shaped cluster. The base of the triangle intersects the diffuse horizontal cluster. The tip of the triangle intersects the vertical axis. This point of intersection on the vertical axis is the corresponding illuminant chromaticity component i_C^j.

Tan *et al.* [91] identify potential specular regions by thresholding on brightness and saturation values, an approach that was originally proposed by Lehmann and Palm [62]. The two thresholding parameters for this method, t_b and t_s, were set to $0 \ 2$ and $0 \ 8$, respectively. We then average the intensities of a specular pixel. A patch is considered specular if the sum of specular intensities s_{sp} exceeds a threshold t_{sp}. The specularity-based illuminant estimate is only employed if a sufficiently large percentage of pixels within a patch are detected as specular. Thus, the detection of specular regions is independent of the local patches over which the illuminant color is estimated.

The actual estimation is conducted in two steps [79]. First, a set of noisy estimates is obtained from rectangular subregions within one patch. The pixels of this region are projected in one IIC space per color channel. If the pixels do not satisfy two straightforward shape criteria (i.e., do not exhibit an elongated shape towards the y-axis), no estimate is obtained from this grid cell [79]. The intercept between the y-axis and the eigenvector that is associated to the largest eigenvalue of the covariance ellipse determines the illuminant color estimate i_j^c for the j-th subregion in the c-th color channel.

In the second step, all estimates per patch are collected in per-color channel histograms $\mathcal{H}_{\mathrm{IIC}}^R$, $\mathcal{H}_{\mathrm{IIC}}^G$, $\mathcal{H}_{\mathrm{IIC}}^B$. and the final estimate is determined as

$$\mathbf{i}_p = \underset{i^c}{\operatorname{argmax}} \, \mathcal{H}_{\mathrm{IIC}}^c(i^c) \qquad c \qquad R \ G \ B \tag{4.15}$$

where $\mathcal{H}_{\mathrm{IIC}}^c(i^c)$ denotes the count for i^c in H_{IIC}^c. For further details, please refer to [79].

Our physics-based unary potential is then defined as:

$$\lambda^p\left(\mathbf{x}_i\ \mathcal{F}_i\right) = w^i \rho_{\sigma_r}\left(\quad\left(\mathbf{i}_i\ \ \mathbf{x}_i\right)\right) \tag{4.16}$$

where $\rho_{\sigma_r}(\cdot)$ is the robust error norm introduced in Eq. 4.11. The weight w^i is a binary weight, defined on the specularity threshold t_{sp} as

$$w = \begin{cases} 1 & \text{if } s_{\mathrm{sp}} \geq t_{\mathrm{sp}} \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

4.4.3 Combining Statistical and Physics-based Illuminant Estimation

Both the statistical and physics-based illuminant estimation can be incorporated in a CRF framework using different unary potentials. An advantage of defining each method as an energy minimization problem is that there is a natural way for combining them into a single color constancy method by defining the local potential as

$$\lambda\left(\mathbf{x}_i\ \mathcal{F}_i\right) = (1-\lambda_p)\lambda^s\left(\mathbf{x}_i\ \mathcal{F}_i\right) + \lambda_p\lambda^p\left(\mathbf{x}_i\ \mathcal{F}_i\right) \tag{4.18}$$

where λ_p is weighting the importance of the physics-based unary potential versus the statistical-based unary potential. Minimizing this energy will combine information from statistical cues as well as specularities into the final local illuminant estimate.

4.4.4 Constraint Illuminant Estimation

Constraint illuminant estimation methods have been popular because they allow to incorporate prior knowledge about the illuminants. Several methods have been proposed which constrain the illuminant set to be on the Planckian locus [26]. Incorporating such constraints is straightforward in our framework. The constraints can be enforced on the illuminant label set $\mathcal{L}$. In this paper, we use a simple constraint where we exclude illuminants which are too saturated, i.e., $\quad i \quad \left(\mathbf{l}_i\ \ \mathbf{i}_w\right) < \lambda_d \quad$ where $\mathbf{i}_w = \frac{1}{\sqrt{3}}\left(1\ \ 1\ \ 1\right)^{\mathrm{T}}$ is the white illuminant.

As a second constraint on the illuminants, we use the fact that in the majority of the multi-illuminant scenes only two illuminants are present. Given a pair of labels $\mathbf{l}_i$ and $\mathbf{l}_j$ the optimal labeling $\breve{\mathbf{x}}^*(i\ \ j)$ for the observation $\mathcal{F}$ is determined with:

$$\breve{\mathbf{x}}^*(i\ \ j) = \arg\min_{\breve{\mathbf{x}} \in \mathbf{L}^{\mathbf{ij}}} E\left(\breve{\mathbf{x}}\ \mathcal{F}\right) \tag{4.19}$$

where $\mathbf{L}^{\mathbf{ij}}$ is the set of all possible labellings on $\mathcal{X}$ restricting the illuminants to $\mathbf{l}_i$ and $\mathbf{l}_j$. The two illuminant constrain is enforced by finding those two illuminants which minimize the energy function. Thus, the selected illuminants are computed with:

$$\hat{\mathbf{L}} = \arg\min_{(l_i\ l_j) \in \mathcal{L}^2} \left(E(\breve{\mathbf{x}}^*(i\ \ j)\ \mathcal{F})\right) \tag{4.20}$$

Note that this also allows for single illuminant estimation in the case that $i = j$.

4.5 Pairwise Potential

The purpose of the pairwise potential function, $((\mathbf{x}_i\ \mathbf{x}_j)\ (\mathcal{F}_i\ \mathcal{F}_j))$ is to ensure, when appropriate, the smooth transition of labels in neighboring vertices. Similar to Boykov *et al.* [14] we consider pairwise potentials that resemble a well. In MRFs, especially as described in [14], $(\mathbf{x}_i\ \mathbf{x}_j) = u(1-\delta_{ij})$, where u is the well "depth" and the function $(1-\delta_{ij})$ controls the shape of the well. In [14], u is defined as a constant and the unit impulse function, $\delta_{ij} = \delta(\mathbf{x_i} - \mathbf{x_j})$, determines the well-shape.

In a CRF (see also [56]) the "depth" depends on the observations $h(\mathcal{F}_i\ \mathcal{F}_j)$. Thus, our pairwise potential function has the form:

$$((\mathbf{x}_i\ \mathbf{x}_j)\ (\mathcal{F}_i\ \mathcal{F}_j)) = h(\mathcal{F}_i\ \mathcal{F}_j)(1-\delta_{ij}) \tag{4.21}$$

We also propose the use of a smoother well function which permits small deviations in illuminant colors between neighboring patches. Thus, our well is defined as:

$$(1-\delta_{ij}) = (1-\cos^{\eta}(\quad(\mathbf{x}_i\ \mathbf{x}_j))) \tag{4.22}$$

where η controls the sharpness of the impulse-like function.

If two neighboring labels are distinct, then there are two possibilities. It can be that the two patches, though spatially close, are illuminated by distinct illuminants, in which case, we should allow for a transition in labels and not significantly penalize the difference in their values. It may, however, be the case that an erroneous label was assigned and the two patches are illuminated by the same illuminant. The depth function $h(\mathcal{F}_i\ \mathcal{F}_j)$ attempts to distinguish between these two cases.

In this work, we use the insight of Logvinenko *et al.* [66] that the shape of an edge (curvature, fuzziness and closedness) conveys discriminatory information about illuminant versus material edges. Influenced by this idea, we use the length of the border between two adjacent patches as an indicator of whether the patches should be sharing incident illumination:

$$h(\mathcal{F}_i\ \mathcal{F}_j) = \text{length}\left(\text{boundary}\left(\mathcal{F}_i\ \mathcal{F}_j\right)\right) \tag{4.23}$$

Longer boundaries imply that the distinct color of the patches is due to differences in material and, hence, the illuminant labels of the adjacent patches should be similar.

However, the proposed framework is general and allows the incorporation and/or combination of multiple methods that can provide information on the discontinuity of illuminants in the scene. For example, one could employ the Retinex [59] heuristic that illumination is expected to vary slowly, thus large changes in surface reflectance are due to differences in material. A Retinex-inspired depth function could then be $h(\mathcal{F}_i\ \mathcal{F}_j) = \exp\left(-\beta_R\ \bar{\mathcal{F}}_i - \bar{\mathcal{F}}_j\ ^2\right)$ where $\bar{\mathcal{F}}_i$ is the average $(R\ G\ B)$ value in patch p_i.

Yet another option is to employ photometric quasi-invariants [94] which help distinguish between shading edges and material edges.

4.6 MIRF: Overall Algorithm

Our algorithm which leads to multi-illuminant estimation is presented in this section. We call it Multi-Illuminant Random Field (MIRF). In the first step we divide the image into subregions or patches. There are several ways used in the literature for obtaining adequate patches. We decided against using superpixels because they are more likely to follow object boundaries rather than subtle illuminant changes. Hence, a grid provides more diverse patch content, and thus more information for the statistical estimators.

Next, we obtain a local illuminant estimate for each patch using the Eq. 4.9 and Eq. 4.15. To add more robustness, these illuminants are then clustered to K illuminants based on their chroma. Additionally, we add a single illuminant estimate I_0 to the illuminant set by applying Eq. 4.9 on the whole image. To reduce the computational cost, we reduce the number of labels by averaging the ones whose angular distance is less than half a degree. We calculate the unary potentials using equation Eq. 4.12 and Eq. 4.16.

In the next step, for every pair of labels we perform the expansion on the graph and obtain the proper labeling (the assignment of the labels to patches) along with the estimation error for the whole image (Eq. 4.19). The pair of two labels which minimizes the error is then chosen (Eq. 4.20). Finally, the label colors are assigned to the patches and the estimated illumination map M is generated. In the last step of the algorithm, a Gaussian smoothing filter with standard deviation σ_p is applied to M as a post processing step in order to reduce artifacts created by the patch boundaries. The methodology is compactly presented in Algorithm 2.

4.7 Multi-illuminant Multi-object Dataset

Several datasets are available for single illuminant estimation. The first datasets where taken under laboratory settings with controlled illumination [3, 6]. Later datasets — often much larger — consist of images of real-world scenes where the ground truth is computed from a reflectance target in the scene, which is either a grey ball [19] or a Macbeth color checker [38]. Gijsenij et al. [45] have introduced a multi-illuminant dataset. To obtain the ground truth of each pixel, the area of every light source is manually annotated. However, manual annotations are difficult to do on complex scenes, and prone to errors[1].

[1]NOTE TO REVIEWERS: From correspondence with the authors of [45], we learned that the PhD student involved in the work left with the data and is not responding to request of returning the raw data. Only compressed version of the data are currently available. Additionally, it is also not possible to recompute the ground truth in a higher quality, due to this reason.

Algorithm 2 Method

1: Apply an $m \times m$ grid on the image to divide it into a set of patches $P = p_1\ p_2\ \ \ p_N$
2: Extract the local illuminant colors for each patch $\mathbf{I} = \mathbf{i}_1\ \mathbf{i}_2\ \ \ \mathbf{i}_N$ (using Eq. 4.9 and Eq. 4.15).
3: Cluster the illuminants (using K-means) and get the K centers. Add the single estimate $\mathbf{I}_0$.
4: Reduce the number of labels by removing the ones whose angular distance is less than .5 degrees. $\mathbf{L}$ denotes the set of remaining illuminants.
5: Calculate the unary potentials (Eq. 4.12 and Eq. 4.16).
6: **for all** $\mathbf{l}_i$ and $\mathbf{l}_j$ $\mathbf{L}$ **do**
7: Calculate $\breve{\mathbf{x}}^*(i\ j)$ (Eq. 4.19)
8: **end for**
9: Find the pair of illuminants $\hat{\mathbf{L}}$ (Eq. 4.20) which produce the lowest error when assigned to the image patches.
10: Back project $\hat{\mathbf{L}}$ to create an illumination map M.
11: Post processing: Apply Gaussian smoothing on M to fade out the artificial edges of the grid.

To address these limitations, we propose two new datasets for multi-illuminant estimation: one taken in a controlled laboratory setting, and one consisting of real-world indoor and outdoor scenes. Each of the sets includes complex scenes with multiple reflectances and specularities. A variety of lighting conditions and illuminant colors are presented in the dataset. Instead of manually annotating the ground truth like [45], we exploit the linearity of the camera response to compute a pixelwise ground truth for our dataset. This way we avoid the subjective task of manually segmenting the image and obtain high-resolution ground truth. In addition, this allows us to have a weight of each illuminant at each pixels rather than a binary decision for each pixel on the dominant illuminant. As can be seen in Figure 4.3 large regions in the image are lit by both illuminants.

4.7.1 Data Acquisition

We used the Sigma SD10 single-lens reflex (SLR) digital camera which uses a sensor with the unique Foveon X3 sensor technology. We chose this camera for its Bayer-pattern-free image sensor and lossless raw 12 bit per color high quality output. Matlab code for the perceptual enhancement of the images has been made publicly available by Parraga *et al.* [75]. We captured the images in the linear RAW format, i.e. without additional gamma or JPEG compression. The original image size is 2304×1531 pixels, i.e. roughly 3.5 megapixels. Upon acceptance of the paper, the original images, together with the ground truth, will be made publicly available.

To compute the ground truth we exploit the linearity of light: we use the fact that the scene taken under both illuminants is equal to the sum of the two scenes taken under a

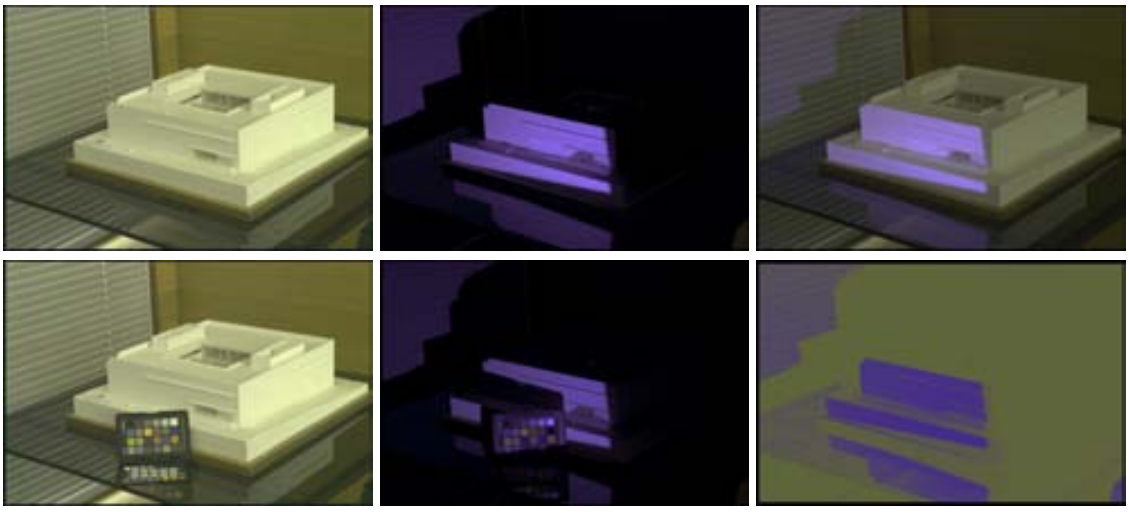


Figure 4.2: Example input images to compute the ground truth. In the top row, from left to right: scene under illuminant 1, 2, and a mixture of 1 and 2. In the bottom row: scene under separate illuminants 1 and 2 again, but with a Macbeth color chart to estimate the illuminant chromaticity. On the bottom right, the ground truth is shown, i.e. the two illuminant colors with their spatial distribution. Gamma is added to the images for visualization.

single illuminant. The basic idea is that we can use the single illuminant images to derive the relative strength of the two illuminants for each pixel in the multi-illuminant scene. In addition, we acquire two images with either the Macbeth color chart or a grey reflectance target for the two single illuminant scenes, from which we can estimate the illuminant color. The five indoor scenes taken are shown in Fig. 4.2.

Since the single illuminants sum up to the two-illuminant scene, in reality we only need two of the three scenes to derive the third one. We use this fact for the scenes where it was not possible to obtain the two single illuminant images. An example is an indoor scene with indoor illuminant and outdoor light coming through the window (which we were unable to block). Taking two images - one with both illuminants and one with only the outdoor lightning after switching off the indoor illuminant- we can compute the ground truth for this scene by using the linearity relation.

Consider the ground truth computation in more detail. We first obtain the colors of the illuminants using the Macbeth color chart, or a grey reflectance target for the lab scenes, respectively. As explained above, a pixel $\mathbf{f_{ab}}$ from a two-illuminant scene is equal to the sum of the pixels from two scenes with a single illuminant, i.e. $\mathbf{f_{ab}} = \mathbf{f_a} + \mathbf{f_b}$. We verified that this assumption holds for the Sigma SD10 single-lens reflex camera. Thus, for every pixel, we compute the relative contribution of illuminant $\mathbf{a}$. Using the von Kries assumption, the images that are only exposed to $\mathbf{a}$ and $\mathbf{b}$ are divided by their respective illuminant chromaticity to obtain scenes under white illumination. The intensity differences in these images reveal the individual influence of each illuminant. We denote a pixel of the illumination-normalized images in the green channel as $f_{\mathbf{a}\,g}$ and $f_{\mathbf{b}\,g}$, respectively.

The relative difference for illuminant $\mathbf{a}$ in the green channel is obtained for each pixel as

$$r_{\mathbf{a}\,g} = \frac{f_{\mathbf{a}\,g}}{f_{\mathbf{a}\,g} + f_{\mathbf{b}\,g}} \tag{4.24}$$

See appendix 4.9.1 for a derivation of Eq. 4.24. In principle, any color channel could be used. However, we found the green channel yielded the most stable (noise-resilient) results.

The ground truth illuminant color $\mathbf{i_{ab}}$ of a pixel $f_{\mathbf{ab}}$ of the mixed-illuminant image is then a pixelwise linear interpolation of $\mathbf{a}$ and $\mathbf{b}$,

$$\mathbf{i_{ab}} = r_{\mathbf{a}\,g} \cdot \mathbf{a} + (1 - r_{\mathbf{a}\,g})\mathbf{b} \tag{4.25}$$

4.7.2 Controlled Laboratory Scenes

The first dataset is taken in a controlled laboratory setting. The scenes vary from simple single-object scenes to more difficult multi-object (cluttered) scenes. The scene content also varies between diffuse objects, specular objects, and mixtures of diffuse and specular objects. In total, the dataset consists of 10 scenes, each under 6 distinct illumination conditions. For computing the ground truth, each scene was captured also under only one illuminant from each position. After removing images that are misaligned, we ended up with a total of 58 benchmark images. We used three differently colored lights, referred to as "blue", "white" and "red", with a chromatic difference of $5\ 9^\circ$ between blue and white, $6\ 1^\circ$ between white and red, and $11\ 4^\circ$ between blue and red. Each scene is lit from two different angles (referred to as "left" and "right" illumination) by different pairs of illuminants. To reduce the influence from ambient illumination, the data acquisition for this set has been done in a box with black diffuse walls.

The left side images in the Fig. 4.1 and Fig. 4.3 show two example scenes illuminated by a red illuminant from the left and a white illuminant on the right. The lion in Fig. 4.1 is an example of a single object scene and the toys in the Fig. 4.3 are an example of a mixture of specular and diffuse objects. The bottom row shows the influence of the both illuminants. A stronger blue component denotes stronger influence of the left illuminant, while red represent the illuminant on the right.

4.7.3 Real-world Scenes

In order to test our framework on more challenging real world images, we captured 20 indoor and outdoor scenes. Here the data is converted to sRGB to mimic a more standard user setting. The scenes contain two dominant illuminants, namely an ambient light source and a direct light. For the outdoor images, shadow regions provide ambient light. For indoor images, the room illumination is used. The direct light source is either added from a projector, the sun, or another additional light bulb. In Fig. 4.1 and Fig. 4.3, two

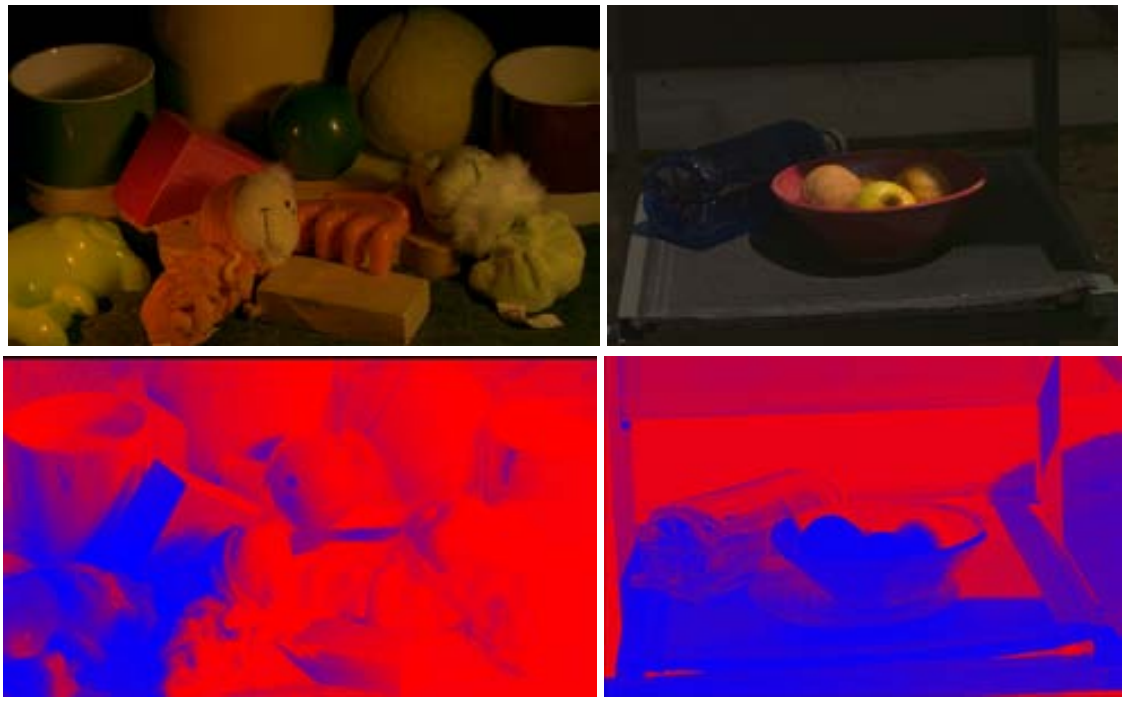


Figure 4.3: Example images from our laboratory and real-world dataset. In the bottom row, the relative influence of the two illuminants is shown, color-coded from blue to red.

example scenes are provided. One scene shows a two illuminant indoor scene, and the second scene shows a strong color shadow in an outdoor scene. The main difference between the two datasets is that in the real-world scenes, the ambient illuminant is present on almost the whole image area, while the direct illuminant covers only a part of each scene.

4.8 Experiments

In this section we compare the performance of the proposed method MIRF to several other approaches. As an error metric, we obtain an error per image by computing the mean of the pixelwise angular distance (Eq. 4.6) between the estimated illuminant color and the ground truth maps. Pixels that were too dark (i.e., for our 12 bit images, pixel intensities below 50) have been excluded from evaluation due to their relatively large noise component. Over these per-image errors, we computed the median and mean errors per dataset. The evaluation was conducted on three datasets, namely our proposed laboratory dataset, our proposed real-world dataset, and the outdoor dataset that has been used by Gijsenij *et al.* [45].

As a baseline, we computed results for a number of established algorithms that address color constancy under uniform illumination. So far, little prior work exists for estimating non-uniform illumination. We implemented the recent method by Gijsenij *et al.* [45], as it showed very competitive performance in a number of experiments.

Both, the method by Gijsenij *et al.* [45] and MIRF use as input illuminant estimates with small spatial support. Such illuminant estimates can be obtained from different estimators. We chose to use grey world ("GW"), which can be obtained from Eq. 4.9 by using the parameters $n = 0$, $m = 1$ $\sigma_{\mathrm{GW}} = 0$, white patch ("WP", with $n = 0$, $m =$, $\sigma_{\mathrm{GW}} = 0$), first order grey-edge ("GE1", with $n = 1$, $m = 1$, $\sigma_{\mathrm{GW}} = 1$) and second-order grey edge ("GE2", with $n = 2$, $m = 1$, $\sigma_{\mathrm{GW}} = 1$). Additionally, we use the physics-based estimator, as presented in Eq. 4.15, denoted as "IEbV" (derived from "illuminant estimation by voting"). We used these base estimators for comparing the performance of the three families of methods as described above. Additionally, we provide the error if the illuminant color is assumed to be already perfectly balanced to white. The "do nothing" ("DN") estimator shows these results. For the evaluation on our proposed dataset, we resampled the images to 20% of their original size to reduce the computational load.

4.8.1 Parameters

A number of parameters have been fixed for the evaluation of MIRF. As patches we used a rectangular grid with cells of 20×20 pixels for the downscaled version of our proposed dataset, and cells of 10×10 pixels for the outdoor images by Gijsenij *et al.* [45]. In both cases, this corresponds to a cell size of about 15×20 pixels. The number of cluster centers k for the k-means algorithm has been set to the square root of the number of grid cells. To obtain the physics-based estimates, we set the Lehmann and Palm parameters $t_b = 0\ 2$ and $t_s = 0\ 8$, and the overall specularity threshold $t_{\mathrm{sp}} = 10$ for pixel intensities between 0 and 1. The subgrid size for single physics-based estimates was 20×20 pixels with a step size of 10 pixels[2], as in [79]. The settings for the CRF framework were as follows: the saturation threshold λ_d for illuminant labels (see Sec. 4.4.4) is set to $15°$. The parameter σ_r in Eq. 4.11 for robust thresholding on the unary potentatials has been set to $2\ 5°$. Finally, the standard deviation σ_p for the Gaussian smoothing on the reprojected illuminant labels has been set to 10.

Besides these globally fixed parameters, we determined three parameters via two-fold cross validation on each dataset. These were the weighting between unary and pairwise potentials θ_p (see Eq. 4.4), the power q (see Eq. 4.12) for computing the unary potentials, and finally, if datacosts from different estimators are combined, λ_p (see Eq. 4.18) for the relative influence of physics-based and statistical estimators.

4.8.2 Comparing Single- and Multi-illuminant Methods

In Tab. 4.1, we present the mean and median errors on our proposed laboratory dataset. In the column "single-illuminant", these results are based on a single global illuminant estimate. The columns "Gijsenij *et al.*" and "MIRF" report results for the multi-illuminant

[2]Note that for the downscaled images from our dataset, this leads to only one estimate per patch, i.e. the voting part is effectively clamped off. However, if the method is applied on larger images (or patches, respectively), the histogram voting is used.

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	10.6°	10.5°	-	-	-	-
GW	3.2°	2.9°	6.4°	5.9°	3.1° (-3%)	2.8° (-3%)
WP	7.8°	7.6°	5.1°	4.2°	3.0°(-41%)	2.8°(-33%)
GE1	3.1°	2.8°	4.8°	4.2°	2.7°(-13%)	**2.6°** (-7%)
GE2	3.2°	2.9°	5.9°	5.7°	**2.6°**(-19%)	**2.6°**(-10%)
IEbV	8.5°	8.3°	-	-	4.5°(-47%)	3.0°(-64%)

Table 4.1: Comparative results on the proposed laboratory dataset.

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	8.8°	8.9°	-	-	-	-
GW	5.2°	4.2°	4.4°	4.3°	3.7°(-16%)	3.4°(-19%)
WP	6.8°	5.6°	4.2°	3.8°	**4.1°** (-2%)	**3.3°**(-13%)
GE1	5.3°	3.9°	9.1°	9.2°	4.0°(-25%)	3.4°(-13%)
GE2	6.0°	4.7°	12.4°	12.4°	4.9°(-18%)	4.5° (-4%)
IEbV	6.0°	4.9°	-	-	5.6° (-7%)	4.3°(-12%)

Table 4.2: Comparative results on the perceptually enhanced real-world images.

methods by Gijsenij *et al.* [45] and our proposed algorithm "Multi-Illuminant Random Field". It turns out, that some single-illuminant estimators, namely GW, GE1 and GE2, already perform relatively well on our dataset. This comes from the fact that in many cases, the ground truth illuminant colors are not very distant from each other. Thus, the overall error can be small, even if only one of the two illuminants (or a color in between both illuminants) is reported as global estimate. However, in all cases, MIRF improves over these estimates. The physics-based estimates for IEbV yield a considerably weaker performance in the mean error, which might be due to the fact that the individual patches are relatively small, such that the voting becomes ineffective. The method by Gijsenij *et al.* performed surprisingly weak, even worse than the single-illuminant estimators. We investigated this case more closely. It turned out that relatively often, weak candidate estimates are selected by the method, which penalizes the overall algorithm. MIRF avoids this particular problem, as the remaining energy from the energy minimization is used as a criterion for the quality of a solution. In Sec. 4.8.3, we excluded this source of error, to directly compare the performance for determining only the distribution of illuminants.

Table 4.2 shows a similar tendency in the results, but this time on our proposed real-world dataset. Note that the overall errors are higher, which is mainly due to the fact that the images have been perceptually enhanced, such that the overall spread of the colors in the image is larger. The largest gain is obtained using localized estimates of the physics-based estimates. This performance gain comes mostly from the robust error metric, which suppresses gross outliers in the physics-based estimates.

In Tab. 4.3, we report results on the outdoor dataset by Gijsenij *et al.* [45]. Note that the reported numbers for the method by Gijsenij *et al.* deviate from what the authors reported in their paper. When investigating their method, we noted that the evaluation

in [45] was conducted on the non-gamma-corrected images[3]. In our implementation, we performed gamma correction on the input images, as it was also originally intended by [45]. The overall errors are higher than in the previous two experiments. First, the images of this dataset are relatively small snippets, consisting mostly of two relatively homogeneous regions in sunlight and shadow. Thus, the underlying localized illuminant color estimators have to estimate on relatively uninformative input. Note that we did not evaluate on the laboratory data by Gijsenij *et al.*, as we found upon manual inspection that the ground truth for these images is not very reliable.

4.8.3 Benchmarking Separate Components of the Algorithm

Estimating multiple illuminants can be considered as two interleaved tasks, namely estimating the illuminant colors and their spatial distribution. The recovery of the spatial distribution was not required for single-illuminant estimators. Hence, we empirically investigated the capability of finding the proper spatial distribution, by providing the methods in this experiment the ground truth illuminant colors. The results on our laboratory dataset are shown in Tab. 4.4. In the left two columns, it can be seen that the performance of the method by Gijsenij *et al.* greatly improved, compared to Tab. 4.1. Thus, we conclude that the selection of the correct illuminant color is one of the major challenges in the method of Gijsenij *et al.*. In the right columns, we show the performance of the proposed method. The best performing method is first order grey edge, with a median error of 1 7°. This shows that the spatial distribution of the illuminants is well approximated by our proposed framework.

In another experiment, we investigated the relative gain of the various improvements we have introduced (see Tab. 4.5). As an example illuminant estimation algorithm, we used the grey world ("GW") estimator. If we remove the constraint of two illuminants and allow an arbitrary number of illuminants, the error increases significantly on our two datasets. Similarly, the robust error norm (see Eq. 4.11) yields an important performance gain on both our datasets. Removing the parameter q which counters uneven color balances only effects results on the Gijsenij dataset. Finally, removing the saturation constraint on the illuminants results in a performance drop on all datasets.

[3]Without gamma correction, we obtain the same numbers as reported in [45].

	Single-illuminant		Gijsenij *et al.*		MIRF	
	Mean	Median	Mean	Median	Mean	Median
DN	4 4°	3 6°	-	-	-	-
GW	15 0°	13 8°	12 2°	13 8°	10 0°(-18%)	10 1°(-27%)
WP	10 3°	11 3°	10 0°	8 4°	7 7°(-23%)	6 4°(-24%)
GE1	10 1°	10 1°	8 5°	7 6°	**7 1**°(-16%)	**4 7**°(-38%)
GE2	8 7°	8 5°	8 1°	7 4°	7 2°(-11%)	5 0°(-32%)
IEbV	10 0°	7 3°	-	-	9 3° (-7%)	7 3° (-0%)

Table 4.3: Evaluation results on the gamma corrected version of the outdoor dataset by Gijsenij *et al.* [45]

	Gijsenij		MIRF	
	Mean	Median	Mean	Median
GW	2.4°	2.3°	2.3°	2.3°
WP	2.2°	2.1°	2.0°	1.9°
GE1	2.1°	2.0°	**1.8°**	**1.7°**
GE2	2.2°	2.1°	1.9°	1.8°

Table 4.4: Performance on our laboratory data for recovering the spatial distribution. The ground truth illuminant colors are provided to the methods.

4.8.4 Combination of Statistical and Physics-based Estimates

Table 4.6 demonstrates another benefit of the framework. By defining the unary potentials as a weighted sum of the physics-based and the statistical unary potentials, we are able to combine cues from multiple methods in a natural way. To determine the parameters, we performed a full cross-validation over θ_p, q and λ_p (see Sec. 4.8.1). It turns out, that a combination of physics-based and statistical estimates can indeed further improve the results (confer Tab. 4.6 (left) and Tab. 4.1), in particular for the white patch and first order grey edge estimates. On the other hand, the performance of the combination of IEbV with GE2 slightly dropped, thus there is no guarantee that a combination of the unary potentials brings a performance gain.

The right columns of Table 4.6 show the performance on our proposed real-world dataset. It is interesting to note that the impact of combined unary potentials on the overall performance is quite different from the experiments on the laboratory data. Here, the majority of the results is slightly worse than the results reported in Tab. 4.2. This behavior, however, is not consistent. For instance, the mean error of IEbV-WP lies slightly below the reported error in Tab. 4.2, similarly the median error for IEbV-GE2. From these results, we conclude that the framework is general enough to allow the straightforward integration of multiple cues. However, whether such a combination indeed brings the desired performance gain has to be investigated on a case-by-case basis.

	Laboratory data		Real-world data		Gijsenij *et al.*	
	Mean	Median	Mean	Median	Mean	Median
MIRF	3.1°	2.8°	3.7°	3.4°	10.0°	10.1°
all lights	4.6°	4.0°	4.2°	4.0°	10.0°	10.2°
w/o Eq. 4.11	3.9°	3.7°	4.3°	4.0°	10.1°	10.1°
$q = 1$	3.0°	2.8°	3.6°	3.3°	10.7°	10.3°
w/o λ_d	3.6°	3.3°	4.6°	3.2°	11.2°	10.1°

Table 4.5: Grey-world results for different configurations of the proposed framework for each dataset.

	Laboratory data		Real-world data	
Combination variant	Mean	Median	Mean	Median
IEbV-GW	3 0°	2 8°	4 2°	4 3°
IEbV-WP	2 6°	2 5°	4 0°	3 4°
IEbV-GE1	2 6°	2 4°	4 5°	4 2°
IEbV-GE2	2 8°	2 8°	4 7°	3 9°

Table 4.6: Combination of physics-based and statistical methods on our laboratory data.

4.8.5 Automatic White Balance

Example results for automatic white balancing are shown in Fig. 4.4. All images are contrast enhanced for improved visualization. In the top row, from left to right, the input scenes "toys", "lion", "camera", and "detergents" are presented. The second row shows perfectly white balanced output using the computed ground truth. The third row shows white balancing results for a single global grey world estimator. The resulting images suffer from a color cast, as both illuminant colors in the scene are corrected with only one estimate. Using the same estimator within the framework by Gijsenij *et al.* [45] (fourth row) clearly improves over the global estimator. However, the images look more greyish and with faded colors as the local estimations were not able to fully separate the effect of illumination from the object color. Also the "lion" is more reddish on the right side. Finally, in the last row, the output of the proposed MIRF is shown. In this case, the improved performance results from the improvement in the selection of the illuminant color, thus the global color cast is removed. Some inaccuracies in the estimation of the spatial distribution of the illuminants may lead to local color casts (e.g., several bluish "blobs" overlay considerable regions of the "camera" image). However, the overall performance of MIRF is in general quite solid, as demonstrated in the "toys" and "detergents" images.

4.9 Conclusions

We proposed the algorithm "Multi-Illuminant Random Field" (MIRF) as an approach for color constancy under non-uniform illumination. In scenes that are exposed to multiple illuminants, it is required to estimate the illuminant colors and their spatial distribution. In our approach, these two tasks are jointly solved within an energy minimization framework. At the same time, the framework is general enough to a) allow the natural combination of different illuminant estimators, like statistical and physics-based approaches, and to b) allow the incorporation of additional cues if they are available, like, for instance, estimates for illuminant edges.

For quantitative evaluation, we present a highly accurate, per-pixel ground truth dataset for scenes under two illuminants. It consists of 58 laboratory images and 20 real-world images. In contrast to prior work, the spatial distribution of the illuminant colors is computed from multiple, spatially aligned input images. Evaluation results on these images and on the real-world dataset by Gijsenij *et al.* are promising. MIRF outperforms single-

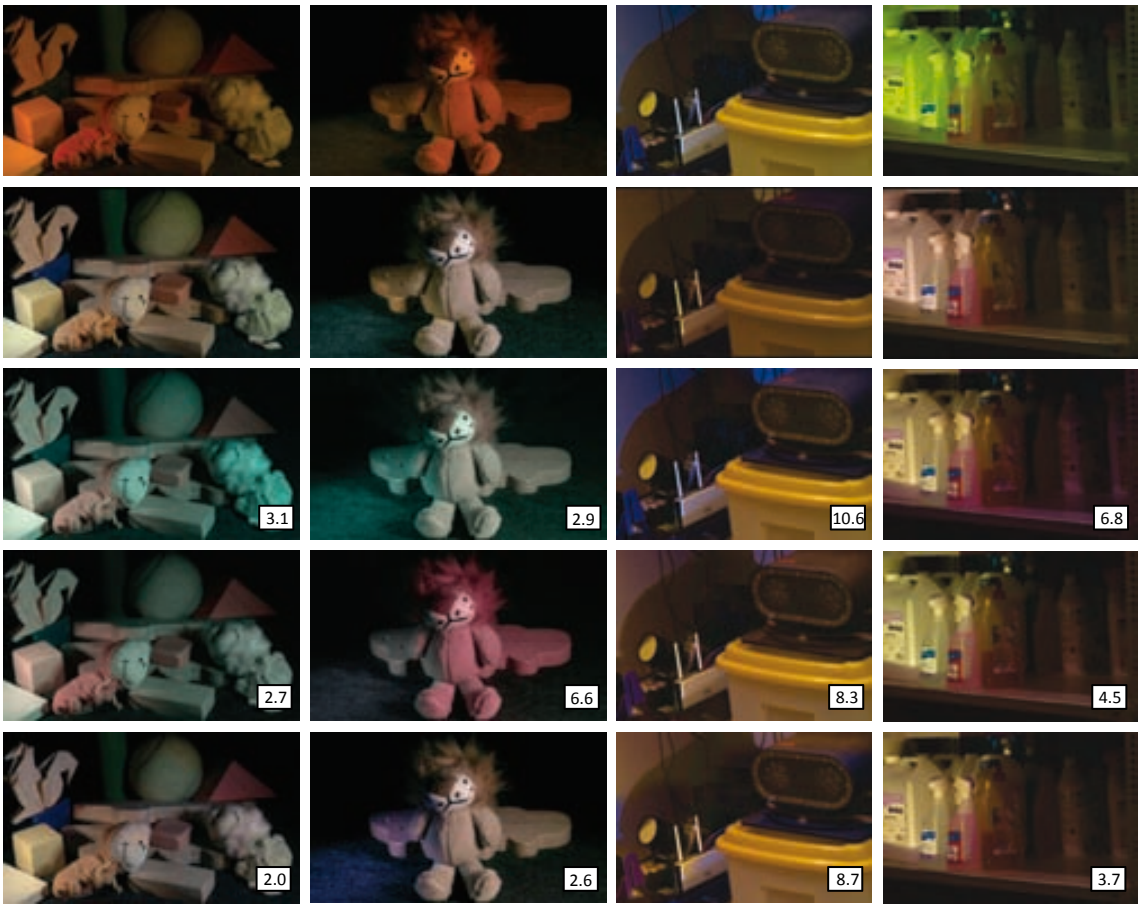


Figure 4.4: Example for the automatic white balance (WB). From top to bottom the rows present: original image from the camera, the WB images using the ground truth, global grey world, Gijsenij *et al.* [45], and MIRF. Note that the images are enhanced to sRGB for visualization. The captions on the images denote their estimation error.

illuminant estimators. Additionally, we show that MIRF's joint estimation of the illuminant color and its spatial distribution consistently outperforms the recently proposed method by Gijsenij *et al.* [45], which solves these two steps separately. In an experiment with ground-truth illuminant colors, we show that also the individual tasks of color estimation and localization perform superiorly. A combination of physics-based and statistical estimates yields competitive results.

As a future extension to this work, it is worth investigation the incorporation of top-down semantic cues into the framework [83,99]. Recognition of common materials in the scene such as grass, stone, and faces could further improve multi-illuminant estimation.

Appendix

In this appendix we prove that the actual choice for the unary potential as given by Eq. 4.8 leads to the same estimate as standard grey-world algorithm in the case of a large θ_p, which

forces the method to find a single illuminant estimate for all patches.

If the energy cost of label changes is chosen large enough (large θ_p), no label changes will be allowed between the patches. As a consequence all patches will have the illuminant estimate, which will essentially be determined by the unary potential. The solution will be that illuminant which yields, summed over all patches, the lowest energy. Consider $\mathbf{x}$ the illuminant choice for all patches (we drop the subscript on $\mathbf{x}$ since it is equal for all patches), the energy can be written as

$$E(\mathbf{x}\ \mathcal{F}) = \sum_{i \in \mathcal{V}} \lambda(\mathbf{x}\ \mathcal{F}_i) = \sum_i \left\| \sum_{j \in p_i} \mathbf{f}^j \right\| \left(1 - \mathbf{i}_i{}^{\mathrm{T}} \mathbf{x}\right) \tag{4.26}$$

where we used $\cos(\quad(\mathbf{i}_i\ \mathbf{x})) = \mathbf{i}_i^{\mathrm{T}}\mathbf{x}$. Filling in Eq.4.5 for $\mathbf{i}_i$ we find that

$$E(\mathbf{x}\ \mathcal{F}) = \sum_i \left\| \sum_{j \in p_i} \mathbf{f}^j \right\| - \sum_i \left(\sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} \tag{4.27}$$

where only the second part depends on $\mathbf{x}$. Since we want to compute $\mathrm{argmin}_{\mathbf{x}}\, E\,(\mathbf{x}\ \mathcal{F})$, this is equal to maximizing the second part of the equation

$$\underset{\mathbf{x}}{\mathrm{argmax}} \sum_i \left(\sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} = \underset{\mathbf{x}}{\mathrm{argmax}} \left(\sum_i \sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} \tag{4.28}$$

since the inner product is distributive over vector addition. From this it follows that

$$\mathbf{x}^* \sum_i \sum_{j \in p_i} \mathbf{f}^j \tag{4.29}$$

which is the solution of the grey-world for the whole image. In conclusion, we have seen that by choosing the particular unary potential of Eq. 4.8, standard grey-world can be written as a energy minimization problem. Hence, when θ_p is chosen large enough, minimizing Eq. 4.4 leads to the same result as the grey-world algorithm. It should be noted, that this is only true when the solution of the grey-world algorithm is in the illuminant label set $\mathcal{L}$. In practice this can easily obtained by choosing the solution of the grey-world as one of the labels.

A similar derivation could be given to prove that minimizing Eq. 4.4 with the unary potential of Eq. 4.10 yields the grey-edge algorithm. Enforcing exactly one label leads to the same answer as the single illuminant in case $p = 1$ in Eq. 4.9.

4.9.1 Estimation of the Two-illuminant Ground Truth

We add details on Eq. 4.24. Let analogously to Eq. 4.24 $f_{\mathbf{a}\,g}$ and $f_{\mathbf{b}\,g}$ denote aligned pixels from the green channels of two images, exposed to illuminant $\mathbf{a}$ and to illuminant $\mathbf{b}$, respectively. We seek the influence of $\mathbf{a}$ and $\mathbf{b}$ in $f_{\mathbf{ab}\,g}$ where both illuminants are additive, i.e. $f_{\mathbf{ab}\,g} = f_{\mathbf{a}\,g} + f_{\mathbf{b}\,g}$. Intuitively, if a pixel is brighter in $f_{\mathbf{a}\,g}$ than in $f_{\mathbf{b}\,g}$,

then the influence of $\mathbf{a}$ is stronger in $f_{\mathbf{ab}\,g}$. The brightness difference comes from a) the intensity of the illuminant and b) from different angles between the light source and the surface normal (for instance, the laboratory lights are located left and right of the scene). Thus, we seek per pixel a weighting factor w, such that

$$\mathbf{i_{ab}} = w \cdot \mathbf{a} + (1 - w) \cdot \mathbf{b} \tag{4.30}$$

i.e. the illumination chromaticity $\mathbf{i_{ab\,g}}$ in this pixel is a weighted sum of the chromaticities of the two illuminants.

To obtain w, we first compute illumination-normalized versions $f_{\mathbf{a}\,g}$, $f_{\mathbf{b}\,g}$ using the von Kries assumption. Thus, $f_{\mathbf{a}\,g}$ and $f_{\mathbf{b}\,g}$ are divided by the green chromaticities of $\mathbf{a}$ and $\mathbf{b}$, respectively. w is then obtained by computing the relative contribution of $f_{\mathbf{a}\,g}$ with respect to $f_{\mathbf{b}\,g}$,

$$w = \frac{f_{\mathbf{a}\,g}}{f_{\mathbf{a}\,g} + f_{\mathbf{b}\,g}} \tag{4.31}$$

Assuming Lambertian reflectance and sharpened sensors, $f_{\mathbf{a}\,g} = \ _{\mathbf{a}}k_{\mathbf{a}}\rho_g$, where $\ _{\mathbf{a}}$ and $k_{\mathbf{a}}$ denote scaling factors due to geometry and the intensity of the light source, respectively. ρ_g denotes the intensity of the pixel's surface albedo. Note that the illuminant color is omitted, as it has been neutralized. Expanding Eq. 4.31, the ratio of the pixel under both illuminant corresponds to the ratio of their scaling factors $\ _{\mathbf{a}}$ and $k_{\mathbf{a}}$,

$$w = \frac{\ _{\mathbf{a}}k_{\mathbf{a}}\rho_g}{(\ _{\mathbf{b}}k_{\mathbf{b}} + \ _{\mathbf{a}}k_{\mathbf{a}})\rho_g} = \frac{\ _{\mathbf{a}}k_{\mathbf{a}}}{\ _{\mathbf{b}}k_{\mathbf{b}} + \ _{\mathbf{a}}k_{\mathbf{a}}} \tag{4.32}$$

as albedo and neutral illuminant are identical in $f_{\mathbf{a}\,g}$ and $f_{\mathbf{b}\,g}$. This leads directly to the formulation in Eq. 4.24.

In practice, we clip the weight w if one of the illuminants is t_{B} times brighter than the other, i.e.

$$w = \begin{array}{cl} 1 & \text{if } f_{\mathbf{a}\,g}\ \ f_{\mathbf{b}\,g} > t_{\mathrm{B}} \\ 0 & \text{if } f_{\mathbf{b}\,g}\ \ f_{\mathbf{a}\,g} > t_{\mathrm{B}} \\ w f_{\mathbf{a}\,g}\ \ f_{\mathbf{b}\,g} & \text{otherwise} \end{array} \tag{4.33}$$

where w normalizes the range of values between 0 and 1. For our dataset, we empirically determined $t_{\mathrm{B}} = 40$ as a reasonable threshold.

In real-world images, the assumption of sharpened sensors and Lambertian reflectance are typically violated. We alleviate this issue with two "engineering decisions". First, we use only the green channel, as an approximation to a sharp sensor. Second, some pixels contain specular reflectance, i.e. are not fully Lambertian. In such cases, the intensity of the specularity often exceeds the clipping range, which assigns the respective pixel fully to the specular illuminant (which agrees with the neutral interface assumption [87]). Finally, note that interreflections are in general not well modeled by this approach. Despite these shortcomings, we manually investigated all scenes, and concluded that the cases that violate our assumptions are rare or do not considerably influence the result. Thus, the proposed approach is a economic, feasible way to obtain pixelwise multi-illuminant ground truth on real-world scenes.

Chapter 5

Dataset

Datasets play an important role in many computer vision and image processing applications. Often developed algorithms need to be validated not only qualitatively but also quantitatively using a relatively large number of examples to prove their performance. Various existing methods make use of prior knowledge to train their parameters in order to achieve better results. In this case the quality, content, and size of the dataset is a decisive factor to avoid over-training the parameters. Here we present our *synthetic scenes* dataset for intrinsic image decomposition.

5.1 Gray Paint Spray Technique for Ground Truth Extraction

A common technique for ground truth estimation used in developing existing shape and intrinsic image estimation datasets is using gray paint. That is, first the scene is captured under all the desired lighting conditions. Then using a matte gray paint spray, all the surfaces are covered by a diffuse gray layer and the scene is captured again under the very same lighting conditions. It is assumed that using these gray-painted scenes one does construct the *shading* image for that scene from which is possible to obtain also the shape.

Often this task is down for scenes which consist of only one single object in order to simplify the problem. One popular example of datasets made using this technique is the widely used collection of MIT by Grosse *et al.* [50] for intrinsic images. Also recently Bleier *et al.* [12] have proposed a dataset for local illuminant estimation.

One of the main advantages of these datasets is their relatively high accuracy and resolution. However, there are several drawbacks regarding this method above which is the slow and cumbersome task of constructing and capturing these images one by one. Since an exact matching between the captured scenes and the ground truth is required, each object should be placed exactly at the same position during the whole process of

capturing that scene which could in practice require a lot of care and repetition. In practice these issues does highly reduce the number of images and variations in the dataset.

Another important shortcoming of such technique is that using the matte gray paint, capturing the ground truth information is only limited to the diffuse reflectance while essential material characteristics such as specularities are omitted. Grosse *et al.* have used polarizing filters to separate specular from Lambertian reflectance and provided some specularity decomposition, while most of the objects on the MIT dataset are mainly diffuse. As acknowledge by Bleier *et al.*, gray paint reduces the captured ground truth illumination information from the scene to direct lighting, losing crucial information such as inter-reflections.

To overcome the shortcomings of the existing datasets in intrinsic image decomposition, and to encourage research also into more complex reflectance models, we propose a synthetic dataset for intrinsic image decomposition.

5.2 3D Object Modeling and Physics-based Renderers

Recent advancements in digital 3D modeling programs have enabled the users to rely on these methods for graphical use, from digital animations and visual effects in movies to computer aided industrial design.

Rendering is the process of generating a 2D image from a description of a 3D scene which is often done using computer programs by calculating the projection of the 3D scene model over the virtual image plane. Both the commercial and open-source rendering programs are moving toward achieving more realistic results and accuracy using physics-based models in optics. There are currently various softwares available which embed the known illumination and reflectance models [76].

Due to the increased accuracy with which 3D renderers visualize the world, using synthetic data to train and test complex computer vision tasks has attracted growing attention. In addition synthetic data allows for easy access to the groundtruth, making it possible to prevent the expensive manual labeling process. Marin et al. [69, 100] show that a pedestrian detector trained from virtual scenarios can obtain competitive results on real-world data. Liebelt and Schmid [65] use synthetic data to improve multi-view object class detection. Finally, Rodriguez et al. [81] generate synthetic license plates to train recognition system.

In the current work we have used Blender to model the scenes. Yafaray, is used as a renderer for its photo-realism and physically plausible rendering. Both of these applications are free and open source which makes them suitable for scientific use.

5.3 Global Lighting

In order to obtain more photo-realistic lighting results for 3D scene rendering, a group of rendering algorithms have been developed with are referred to as *global illumination*. These methods, in addition to taking into account the light which riches the object surface directly from a light source, *direct lighting*, also calculate the energy from the same light source which is reflected by other surfaces in the scene. The later is also known as *indirect lighting*. This indirect lighting is what causes the reflections, refractions, shadows, ambient lighting, and inter-reflections. Figure 5.1 presents the competitive quality and photo-realism for synthetic scenes rendered using global illumination.

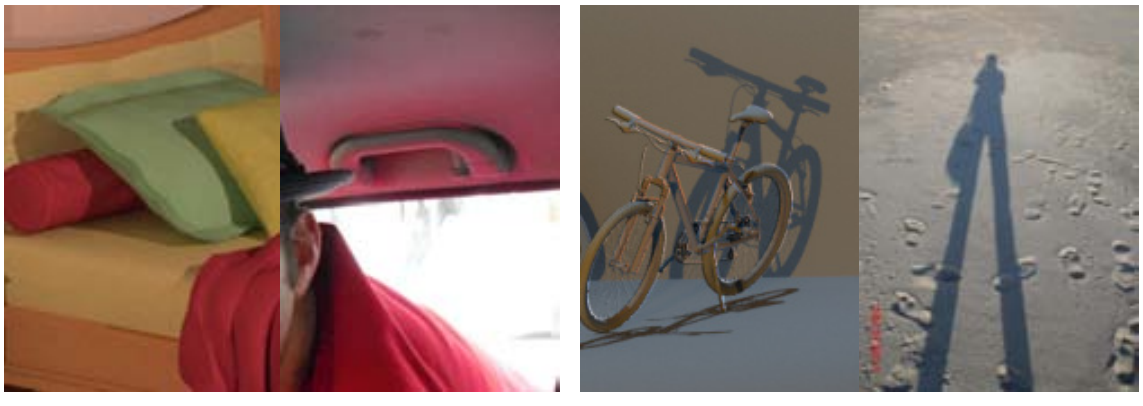


Figure 5.1: The above examples compare the real-world photographed scenes versus rendered scenes from our dataset. The first two images on the left are examples of diffuse inter-reflections, while the images on the right present colored shadows. Similar effect can be observed in the synthetic images (the first and the third from the left) as in the real-world photographs (the second and the last images from left).

There are many popular algorithms for rendering global illumination (e.g, radiosity, raytracing, and image-based lighting). One of the most popular methods of such is *photon mapping* [53] developed by Henrik Wann Jensen. To achieve physically sound results and photo-realism in our dataset we make use of the photon mapping method embedded in Yafaray. Figure 5.2 shows the importance of indirect lightning in scenes. For this purpose we compare the final renderings of our data set to the renderings which only consider direct lighting (one bounce). Here the global illumination is achieved using photon mapping which appears more realistic due to preservation of the *diffuse inter-reflection*.

5.4 Intrinsic Image Datasets

Intrinsic image algorithms and datasets can be distinguished by their assumptions on the underlying reflectance models. Consider the reflection model which models the color

Figure 5.2: Comparing the different rendering methods: *direct lighting* (left) and *photon mapping* (right) on an example scene from our dataset.

Figure 5.3: An example of a synthetic scene. From left to right: the rendered scene, reflectance component, and shading.

observation f^c with $c \quad R\ G\ B$ as:

$$f^c(\mathbf{x}) = m(\mathbf{x}) \quad s(\lambda\ \mathbf{x})\, e(\lambda\ \mathbf{x}) \rho^c(\lambda)\, d\lambda \tag{5.1}$$

where the integral is over all wavelengths λ of the visible spectrum . The material reflectance is given by $s(\lambda\ \mathbf{x})$ and m is a scalars depending on the scene geometry (viewpoint, surface normal, and illuminant direction). The camera sensitivity is given by ρ^c.

We will use this basic reflection model to demonstrate the differences between existing datasets and our dataset. In the MIT dataset [50] the illuminant is considered to be independent of $\mathbf{x}$ and white, i.e. $e(\lambda\ \mathbf{x}) = 1$. This assumption is shared by most

of the intrinsic image methods [9] [39] [84]. Recently, Barron and Malik [10] relaxed this assumption: they allow the illuminant color to vary but only consider direct illuminantion (ignoring interreflections). Their assumption on the illuminant is given by $e(\lambda\ \mathbf{x}) = e(\lambda\ n(\mathbf{x}))$, where $n(\mathbf{x})$ is the surface normal at location $\mathbf{x}$. They construct a dataset by synthetically relighting the real-world MIT dataset [10].

In this paper, we go one set further and create a synthetic dataset by using rendering techniques from the computer graphics field. This allows us to remove the restriction other dataset put on $e(\lambda\ \mathbf{x})$. The illuminant color and strength can change from location to location. This allows us to consider more complex reflection phenomena such as self-reflection and inter-reflections. To the best of our knowledge this is the first intrinsic image dataset which considers these more complex reflection models. In the next section we analyze rendering accuracy for such reflection phenomena.

Note that the above reflection model assumes that the materials are Lambertian reflectances. Even though specular materials can be accurately rendered, we exclude them from this dataset because most existing intrinsic image algorithms are not able to handle non-Lambertian materials . The MIT dataset [50] applies polarizing filters to provide both images with and without specular reflection.

5.5 Analysis Color Rendering Accuracy

For synthetic datasets to be useful to train and evaluate computer vision algorithms, they should accurately model the physical reality of the real-world. Therefore, in this section, we analyze the accuracy of color rendering based on the diagonal model as is typical in computer graphics. To prevent propagating the full multispectra data, which is computationally very expensive, rendering engines approximate Eq. 5.1 with

$$f^c = \quad s(\lambda)\rho^c(\lambda)\,d\lambda \quad e(\lambda)\rho^c(\lambda)\,d\lambda \tag{5.2}$$

Here we removed the dependence on $\mathbf{x}$, and the geometrical term m and focus on the color content of f. In vector notation we could write this as

$$\mathbf{f} = \mathbf{s} \circ \mathbf{e} \tag{5.3}$$

where we use bold to denote vectors, and $\circ$ is the Hadamard product, and we replaced $\mathbf{s} = \quad s(\lambda)\rho^c(\lambda)\,d\lambda$ and $\mathbf{e} = \quad e(\lambda)\rho^c(\lambda)\,d\lambda$. In real scenes the light which is coming from objects in the scene is not only caused by direct lightning of the illuminant but part of the light is reflected from other objects in the scene. Considering both direct and interreflection from another surface we can write:

$$\mathbf{f} = \mathbf{s^1} \circ \mathbf{e} + \mathbf{s^2} \circ \mathbf{s^1} \circ \mathbf{e} \tag{5.4}$$

where the superscript is used to distinguish the material reflectance of different objects. The accuracy of the approximations in Eq. 5.3 and Eq. 5.4 is dependent on the shape and

the number of sensors c considered. Typically rendering machines apply three sensors c $R\ G\ B$, however it is known that the rendering accuracy increases when considering more sensors [21] [2].

To test the accuracy of f^c we perform a statistical analysis. We use the 1269 Munsell color patches and we compute both f^c and f^c. For sensors ρ^c we use Gaussian shaped sensors which are equally spaced over the visible spectrum. We compare the reconstruction error

$$= \left\| \mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\mathbf{x}) \right\| \quad \mathbf{f}(\mathbf{x}) \tag{5.5}$$

for the cases of three, six and nine sensors. We consider both single bounce (Eq. 5.3) and two bounce situation (Eq. 5.4). As illuminant we have taken the standard D65 daylight illuminant. Dark patches where discarded because the reconstruction error is unstable for these cases.

In Table 5.1 the results of the experiment are provided. For single bounce the three sensor approximation which is common in graphics is acceptable and only leads to a maximum error of $2\ 88\%$. However, if we consider interreflections the maximum error reaches the unacceptable level of $22\ 7\%$. Based on these results we have chosen to use a 6 sensors system to propagate the multispectral color information, resulting in a maximum error of $7\ 7\%$. This can be conveniently achieved by running existing rendering software (build for 3 channel propagation) twice for three channels [21] [2]. The final 6-D result image is projected back to an RGB image using linear regression. In the only available intrinsic image data set for multi-illuminants [10], illuminants where introduced synthetically by using a 3 channel approximation. Since this data set only considers direct lightning, our analysis shows that this suffices. However, in the case of interreflections, synthetically relightning of real-world scenes would introduce significant errors.

	One bounce		**Two bounces**	
sensor	Mean (%)	Max (%)	Mean (%)	Max (%)
3	0.53	2.88	1.26	22.70
6	0.18	1.25	0.51	7.70
9	0.11	0.86	0.32	3.40

Table 5.1: Reconstruction error for single and two bounce reflection for 3,6, and 9 sensors.

Chapter 6

Conclusions and Future Directions

In this thesis, we aim at improving the modeling of the illumination and its interaction with object surface. The first part of this chapter summarizes the work. In the second part we discuss possible directions for future work.

6.1 Conclusions

In this thesis, we have investigated various models for illumination and object reflectance modeling. We have extended the existing reflection models to account for real-world Multi-illuminant scenes. In the first part we have presented a framework to estimate the chromaticity of the lights illuminating the scene using specular highlights and decomposition of a color image to a set of images capturing its intrinsic characteristics (e.g. diffuse and specular components). We have demonstrated results on challenging real-world images with complex illumination and reflectance in the presence of colored shadows and inter-reflections. Moreover, we showed some examples for the possible applications of our framework in order to improve automatic and semi-automatic photo-editing tasks (e.g, photo-fusion and color transfer).

In Chapter 4 we further improved over state-or-the-art illuminant estimation methods by formulating the scene's illumination as an energy minimization combining bottom-up color constancy methods for global illuminant estimation into a mathematically sound formulation which embeds both the statistical and physics-based method. Using Conditional Random Fields (CRF) we achieved global consistency of the illuminant estimations. Using an extensive experimental evaluation we demonstrated that the proposed method addresses the intrinsic challenges in multi-illuminant scenes, i.e. the estimation of the illuminant colors and their spatial distribution, with superior accuracy compared to prior work. Therefore, we show that proposed framework is able to perform high quality automatic digital white balancing in complex scenes.

As discussed in Chapter 5 We have created two main datasets which are explained in

Chapter 5. To our knowledge our multi-illuminant scene dataset captured from both laboratory controlled scenes and real-world indoor and outdoor scenes is most general and complex among the existing dataset for multi-illuminant color constancy which includes a highly accurate estimated ground-truth. This chapter further presents our synthetic scene dataset for intrinsic image decomposition using physically sound 3D modeling and rendering programs. We have used this synthetic dataset to perform a benchmarking on state-of-the-art intrinsic image estimation and decomposition methods.

6.2 Future Directions

As an extension to our multi-illuminant estimation method, we proposed to include gamut-mapping approaches as an extra cue for our estimation framework, as well as combining various statistical methods in order to improve the bottom-up local estimations. Furthermore, we acknowledge the necessity of expansion of our datasets to include more variety of the scenes and illumination conditions.

We propose as our future work, to investigate material characteristics based on their reflectance in order to perform material classification from which both science and industry could benefit. Also by removing the effect of complex illumination facilitates other Computer Vision tasks (segmentation, classification, etc.). Moreover, the illumination and reflectance modeling could be applied to *image forensics* approaches.

Bibliography

[1] Xiaobo An and Fabio Pellacini. User-controllable color transfer. *Eurographics*, 2010.

[2] R.S. Berns T. Chen B. Darling, J.A. Fewerda. Real-time multispectral rendering with complex illumination. 2010.

[3] K. Barnard, V. Cardei, and B. Funt. A Comparison of Computational Color Constancy Algorithms – Part I: Methodology and Experiments With Synthesized Data. *IEEE Transactions on Image Processing*, 11(9):972–983, September 2002.

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

[5] K. Barnard, G. Finlayson, and B. Funt. Color Constancy for Scenes with Varying Illumination. *Computer Vision and Image Understanding*, 65(2):311–321, February 1997.

[6] K. Barnard, L. Martin, A. Coath, and B. Funt. A Comparison of Computational Color Constancy Algorithms – Part II: Experiments With Image Data. *IEEE Transactions on Image Processing*, 11(9):985–996, September 2002.

[7] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms. ii. experiments with image data. *Image Processing, IEEE Transactions on*, 11(9):985 – 996, sep 2002.

[8] Kobus Barnard. Improvements to gamut mapping colour constancy algorithms. In *Proceedings of the 6th European Conference on Computer Vision-Part I*, ECCV '00, pages 390–403, London, UK, UK, 2000. Springer-Verlag.

[9] J.T. Barron and J. Malik. High-frequency shape and albedo from shading using natural image statistics. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2521–2528. IEEE, 2011.

[10] J.T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *Proc. European Conf. on Computer Vision*, 2012.

[11] H.G. Barrow and J.M. Tenenbaum. Recovering intrinsic scene characteristics from images. *In A. Hanson and E. Riseman, editors, Computer Vision Systems*, 1978.

[12] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troger, and A. Kaup. Color constancy and non-uniform illumination: Can existing algorithms work? In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 774 –781, nov. 2011.

[13] M. Bleier, C. Riess, Sh. Beigpour, E. Eibenberger, E. Angelopoulou, T. Troeger, and A. Kaup. Color Constancy and Non-Uniform Illumination: Can Existing Algorithms Work? In *IEEE Workshop on Color and Photometry in Computer Vision*, pages 774–781, 2011.

[14] Y. Boykov, O. Veksler, and R. Zabih. Markov Random Fields with Efficient Approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 648–654, 1998.

[15] D. H. Brainard and W. T. Freeman. Bayesian Color Constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

[16] G. Buchsbaum. A Spatial Processor Model for Color Perception. *Journal of the Franklin Institute*, 310(1):1–26, July 1980.

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

[18] Bryan Rigg Robert W. G. Hunt Changjun Li, M. Ronnier Luo. Cmc 2000 chromatic adaptation transform cmccat2000. *Color Research and Application*, 2002.

[19] F. Ciurea and B. Funt. A Large Image Database for Color Constancy Research. In *Color Imaging Conference*, pages 160–164, 2003.

[20] Ross Henry Day. The basis of perceptual constancy and perceptual illusion. *Investigative Ophthalmology and Visual Science*, 11(6):525–532, 1972.

[21] M.S. Drew and G.D. Finlayson. Multispectral processing without spectra. *JOSA A*, 20(7):1181–1193, 2003.

[22] M.S. Drew and G.D. Finlayson. Realistic colorization via the structure tensor. *ICIP*, 2008.

[23] M. Ebner. Color Constancy Using Local Color Shifts. In *European Conference in Computer Vision*, pages 276–287. Springer-Verlag, 2004.

[24] M. Ebner. *Color Constancy*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2007.

[25] Marc Ebner. Color constancy using local color shifts. In Tomás Pajdla and Jiri Matas, editors, *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III*, volume 3023 of *Lecture Notes in Computer Science*, pages 276–287. Springer, 2004.

[26] G. D. Finlayson, S. D. Hordley, and I. Tastl. Gamut Constrained Illuminant Estimation. *International Journal of Computer Vision*, 67(1):93–109, 2006.

[27] G. D. Finlayson and G. Schaefer. Solving for Color Constancy using a Constrained Dichromatic Reflection Model. *International Journal of Computer Vision*, 42(3):127–144, 2001.

[28] G.D. Finlayson. Color in perspective. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(10):1034 –1038, oct 1996.

[29] G.D. Finlayson, M.S. Drew, and B.V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *JOSA A.*, 11(5), 1994.

[30] G.D. Finlayson, S.D. Hordley, and R. Xu. Convex programming colour constancy with a diagonal-offset model. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III – 948–51, sept. 2005.

[31] G.D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Proc. IS&T SID s Color Imaging Conf.*, pages 37–41, 2004.

[32] G.D. Finlayson and R. Xu. Convex programming color constancy. In *The IEEE Workshop on Color and Photometric Methods in Computer Vision*, 2003.

[33] Graham D. Finlayson and Gerald Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *IJCV*, 42, 2002.

[34] D. Forsyth. A Novel Algorithm for Color Constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

[35] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, September 1990.

[36] D.H. Foster. Color constancy. *Vision Research*, 51(7):674–700, April 2011.

[37] Daniel Freedman and Pavel Kisilev. Object-to-object color transfer: Optimal flows and smsp transformations. In *CVPR*, 2010.

[38] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian Color Constancy Revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[39] P.V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Neural Info. Proc. Systems (NIPS)*, 2011.

[40] Ron Gershon, Allan D. Jepson, and John K. Tsotsos. From [R,G,B] to Surface Reflectance: Computing Color Constant Descriptors in Images. In *10th international joint conference on Arti cial intelligence*, volume 2, pages 755–758, August 1987.

[41] J.-M. Geusebroek, R. Boomgaard, A. Smeulders, and T. Gevers. Color Constancy from Physical Principles. *Pattern Recognition Letters*, 24(11):1653–1662, July 2003.

[42] T. Gevers and A. W. M. Smeulders. Color-based object recognition. *International Journal of Pattern Recognition and Arti cial Intelligence*, 32(3):453–464, 1999.

[43] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2):127–139, 2010.

[44] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized Gamut Mapping using Image Derivative Structures for Color Constancy. *International Journal of Computer Vision*, 86(2–3):127–139, January 2010.

[45] A. Gijsenij, R. Lu, and T. Gevers. Color Constancy for Multiple Light Sources. *IEEE Transactions on Image Processingn*, 2011.

[46] Arjan Gijsenij, Theo Gevers, and Joost van de Weijer. Computational Color Constancy: Survey and Experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, September 2011.

[47] Arjan Gijsenij, Theo Gevers, and Joost van de Weijer. Improving color constancy by photometric edge weighting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2012.

[48] E.B. Goldstein. *Sensation and Perception*. Thomson Learning, 2002.

[49] R. Gonsalves. Method and apparatus for color manipulation. *United State Patent 6,351,557*, Feb 26, 2002.

[50] R. Grosse, M.K. Johnson, E.H. Adelson, and W.T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2335–2342. IEEE, 2009.

[51] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand. Light Mixture Estimation for Spatially Varying White Balance. *ACM Transactions on Graphics*, 27(3):70:1–70:7, August 2008.

[52] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand. Light mixture estimation for spatially varying white balance. In *ACM SIGGRAPH 2008 papers*, SIGGRAPH '08, pages 70:1–70:7, New York, NY, USA, 2008. ACM.

[53] Henrik Wann Jensen. *Realistic image synthesis using photon mapping*. A. K. Peters, Ltd., Natick, MA, USA, 2001.

[54] R. Kawakami, K. Ikeuchi, and R. T. Tan. Consistent Surface Color for Texturing Large Objects in Outdoor Scenes. In *IEEE International Conference on Computer Vision*, pages 1200–1207, 2005.

[55] G.J. Klinker and S.A. Shafer. A physical approach to color image understanding. *IJCV*, 4, 1990.

[56] Pushmeet Kohli, L?ubor Ladick?, and Philip Torr. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82:302–324, 2009. 10.1007/s11263-008-0202-0.

[57] V. Konushin and V. Vezhnevets. Interactive image colorization and recoloring based on coupled map lattices. *Graphicon*, 2006.

[58] V. Kravtchenko and J.J. Little. Efficient color object segmentation using the dichromatic reflection model. *Communications, Computers and Signal Processing, 1999 IEEE Paci c Rim Conference on*, 1999.

[59] E. H. Land. Lightness and the Retinex Theory. *Scienti c American*, 237(6):108–129, December 1977.

[60] E. H. Land. The retinex theory of color vision. *Scienti c American*, 1977.

[61] H.-C. Lee. Method for Computing the Scene-Illuminant Chromaticity from Specular Highlights. *Journal of the Optical Society of America A*, 3(10):1694–1699, 1986.

[62] T. M. Lehmann and C. Palm. Color Line Search for Illuminant Estimation in Real World Scene. *Journal of the Optical Society of America A*, 18(11):2679–2691, 2001.

[63] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *PAMI*, 30, 2008.

[64] Bing Li. Evaluating combinational color constancy methods on real-world images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2011.

[65] J. Liebelt and C. Schmid. Multi-view object class detection with a 3d geometric model. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1688–1695. IEEE, 2010.

[66] A. D. Logvinenko, E. H. Adelson, D. A. Ross, and D. Somers. Straightness as a Cue for Luminance Edge Classification. *Perception and Psychophysics*, 67:120–128, 2005.

[67] R Beau Lotto and Dale Purves. The empirical basis of color perception. *Consciousness and Cognition*, 11(4):609 – 629, 2002.

[68] Satya P. Mallick, Todd Zickler, Peter N. Belhumeur, and David J. Kriegman. Specularity removal in images and videos: A pde approach. In *ECCV*, 2006.

[69] J. Marin, D. Vázquez, D. Gerónimo, and A.M. López. Learning appearance in virtual scenarios for pedestrian detection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 137–144. IEEE, 2010.

[70] B.A. Maxwell, R.M. Friedhoff, and C.A. Smith. A bi-illuminant dichromatic reflection model for understanding images. In *CVPR*, 2008.

[71] B.A. Maxwell and S.A. Shafer. Physics-based segmentation of complex objects using multiple hypothesis of image formation. *CVIU*, 65, 1997.

[72] B.A. Maxwell and S.A. Shafer. Segmentation and interpretation of multicolored objects with highlights. *CVIU*, 77, 2000.

[73] M. Mosny and B. Funt. Cubical gamut mapping colour constancy. In *IS&T s European Conference on Color in Graphics, Imaging and Vision*, 2010.

[74] Ido Omer and Michael Werman. Color lines: Image specific color representation. *CVPR*, 2, 2004.

[75] C. A. Parraga, R. Baldrich, and M. Vanrell. Accurate Mapping of Natural Scenes Radiance to Cone Activation Space: A New Image Dataset. In *European Conference on Colour in Graphics, Imaging, and Vision*, 2010.

[76] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. The Morgan Kaufmann series in interactive 3D technology. Elsevier Science, 2010.

[77] François Pitié, Anil C. Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *CVIU*, 107, 2007.

[78] Erik Reinhard, Michael Ashikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21, 2001.

[79] C. Riess, E. Eibenberger, and E. Angelopoulou. Illuminant Color Estimation for Real-World Mixed-Illuminant Scenes. In *IEEE Workshop on Color and Photometry in Computer Vision*, 2011.

[80] Ch. Riess, J. Jordan, and E. Angelopoulo. A common framework for ambient illumination in the dichromatic reflectance model. *ICCV Workshops*, 2009.

[81] J. Rodriguez-Serrano, H. Sandhawalia, R. Bala, F. Perronnin, and C. Saunders. Data-driven vehicle identification by image matching. In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 536–545. Springer, 2012.

[82] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM SIGGRAPH*, 23, 2004.

[83] R. Schettini, G. Ciocca, and I. Gagliardi. Content-Based Color Image Retrieval with Relevance Feedback. In *International Conference on Image Processing*, volume 3, pages 75–79, 1999.

[84] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell. Names and shades of color for intrinsic image estimation: Supplementary material. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012.

[85] Marc Serra, Olivier Penacchio, Robert Benavente, and Maria Vanrell. Names and shades of color for intrinsic image estimation. In *Computer Vision and Pattern Recognition (CVPR)*, 2012.

[86] A. Shafer and D. Lischinski. Using color to seperate reflection components. *Color Research and Application*, 10, 1985.

[87] S. A. Shafer. Using Color to Separate Reflection Components. *Journal Color Research and Application*, 10(4):210–218, 1985.

[88] H.L. Shen and J.H. Xin. Transferring color between three-dimensional objects. *Applied Optics*, 44(10), 2005.

[89] Ping Tan, Long Quan, and Stephen Lin. Separation of highlight reflections on textured surfaces. In *CVPR*, 2006.

[90] Robby T. Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. In *PAMI*, 2003.

[91] Robby T. Tan, Ko Nishino, and Katsushi Ikeuchi. Color constancy through inverse-intensity chromaticity space. *Journal Optical Society of America A*, 21(3):321–334, Mar 2004.

[92] Marshall F. Tappen, William T. Freeman, and Edward H. Adelson. Recovering intrinsic images from a single image. *PAMI*, 27, 2005.

[93] S. Tominaga and B.A. Wandell. Standard surface-reflectance model and illuminant estimation. *JOSA*, 6(4), 1989.

[94] J. van de Weijer, T. Gevers, and J.-M. Geusebroek. Edge and Corner Detection by Photometric Quasi-Invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):625–630, April 2005.

[95] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *Image Processing, IEEE Transactions on*, 16(9):2207 –2214, sept. 2007.

[96] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge Based Color Constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, September 2007.

[97] J. van de Weijer, T. Gevers, and A. W.M. Smeulders. Robust photometric invariant features from the color tensor. *Trans. Img. Proc.*, 15(1):118–127, January 2006.

[98] J. van de Weijer and R. van den Boomgaard. Least squares and robust estimation of local image structure. *IJCV*, 64(2-3), 2005.

[99] Joost van de Weijer, Cordelia Schmid, and Jan Verbeek. Using High-Level Visual Information for Color Constancy. In *IEEE International Conference on Computer Vision*, Rio de Janeiro, Bresil, 2007.

[100] D. Vázquez, A.M. López, and D. Ponsa. Unsupervised domain adaptation of virtual and real worlds for pedestrian detection. In *Proc. Int. Conf. on Pattern Recognition*, 2012.

[101] J. von Kries. Influence of adaptation on the effects produced by luminous stimuli. *Sources of Color Vision, D. L. MacAdam, ed.*, 1970.

[102] Y. Weiss. Deriving intrinsic images from image sequences. *ICCV*, 2001.

[103] Yu wing Tai, Jiaya Jia, and Chi keung Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *CVPR*, 2005.