Illumination and Object Reflectance Modeling

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

Bellaterra, April 4, 2013

Director	**Dr. Joost van de Weijer** Centre de Visió per Computador
Tutor	**Dr. Ernest Valveny** Dep. Ciències de la Computació 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: 978-84-940530-9-2

Printed by Ediciones Gráficas Rey, S.L.

. .

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

Acknowledgements

I would like to thank the committee for giving me the opportunity to present the scientific outcome of the last four years of my research life as a member of the Computer Vision Center. The work done in this thesis could not be possible without the support and help of many individuals. I take this opportunity to thank all those people who helped me during these years. Specifically, I would like to single out following people for their wonderful support.

Firstly, I would like to thank Dr. Joost van de Weijer for giving me the opportunity to carry out my postgraduate studies under his supervision, for his scientific insight, and for being always my hardest critic, without whom this work could not have achieved its scientific level. And I appreciate the support of my great friends and colleagues who collaborated with me on various scientific projects and publications during these four years. Especially Dr. Elli Angelopoulou who mentored me during my research visit to Friedrich-Alexander-Universität Erlangen-Nürnberg for all her generous support not only as a teacher, but also as a great friend. And Christian Riess who has been like a family to me. Working with them was a great honor which I hope would continue in the future. Furthermore, my brief collaboration with my good friends, Fahad Khan and Marc Serra, was a pleasure.

I owe the completion of this work also to those great people in my life who with their unconditional friendship and support gave me the courage and power to overcome the hardships of these years. Words are not enough to thank Sven for always being there for me and for all his love and support during the hardest moments of my life which made it possible to carry on and focus on my work. Also I am grateful to my father who until the last moments of his life tried to support me and not let his acute health condition keep me from devoting my mind and energy to my studies and research. And to my sister who introduced me to the world of science and its fascinating power.

Moreover, I would like to thank my best friends Mahsa and Claudia who have always inspired me and made my life a happier experience. I would like to express my gratitude to my good friend Ramon who always guided me with his wisdom, for his unconditional friendship and for all his kindness through all these years of my life in Barcelona. Also I thank the kindness of the lovely Kunz family, especially Iris and Klaus, for accepting me as their own. Many thanks to my amazing friend Negar with whom I have made some of the best memories of my life during my undergraduate studies and to my good friend David for all his generous support during the past four years.

Among those kind individuals who helped me during the toughest time in my life, I would like to especially thank my friends Nacho, Estrella, Conchi, Veronica, Rosa, Tulio, and Tressa for helping me believe in myself and overcome the challenges in my life.

Finally I would like to acknowledge the support of my good friends Fernando, Naila, Jose, Jorge, Camp, Andrés, Olivier, Mirta, Jordi, Noha, Joselina, Marta, Beata, Montse, Hilda, Alejandro, Nadine, Bhaskar, Andy, Murad, Naveen, Carlos, Javier, Angela, Jan, Esther, Yainuvis, Annika, Shahbaz, Marc, Ariel, Montse, Pilar, all my colleagues in the Computer Vision Center, Universitad Autonoma de Barcelona, the Pattern Recognition Lab, and all my friends in Tehran, Sant Cugat del Valles, Barcelona and Erlangen who made each day of these four years count.

Abstract

More realistic and accurate models of the scene illumination and object reflectance can greatly improve the quality of many computer vision and computer graphics tasks. Using such model, a more profound knowledge about the interaction of light with object surfaces can be established which proves crucial to a variety of computer vision applications. In the current work, we investigate the various existing approaches to illumination and reflectance modeling and form an analysis on their shortcomings in capturing the complexity of real-world scenes. Based on this analysis we propose improvements to different aspects of reflectance and illumination estimation in order to more realistically model the real-world scenes in the presence of complex lighting phenomena. As applications of improved reflectance and illumination estimation we show examples of improved automatic white balance, scene relighting, and object re-coloring. The proposed theory can be employed in order to improve other fields of computer vision, such as color naming, object detection, recognition, and image segmentation.

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. To address such complex physical phenomena, we extend the standard 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 inter-reflections. 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 scenes with non-uniform illumination (e.g., an outdoor scene with a blue sky and a yellow sun, or a scene with indoor lighting combined with outdoor lighting through a window). 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 problem. This method provides us with local illuminant estimations which significantly improve 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 that our approach achieves improvement over state-of-the-art methods for local illuminant estimation. We further propose the use of synthetic data to facilitate the construction of datasets and improve the process of obtaining ground-truth information. We proceed

by comparing state-of-the-art intrinsic estimation and decomposition methods using our computer-generated dataset for the case of more complex objects and illumination conditions (e.g., specularities, inter-reflections and shadows).

Resum

Poder tenir models de la il·luminació de la escena i la reflectància dels objectes més realistes podrien millor de manera significativa moltes de les tasques de la visió per computador i els gràfics per computador. Usant un d'aquests models es podria obtenir un coneixement més profund sobre la interacció de la llum i les superfícies dels objectes en la vida real, que és crucial per a una gran varietat de aplicacions en visió per computador. En el present treball, investiguem les diferents aproximacions existents al modelatge de la il·luminació i reflectàncies, fen-te un anàlisis dels seus defectes a l'hora de capturar la complexitat de les escenes del món real. Basant-nos en aquest anàlisi, proposem millores en diferents aspectes de l'estimació de la il·luminació i la reflectància per tal de modelar més realistament les escenes reals en casos d'efectes d'il·luminació complexa. Com a aplicacions de la millora de la estimació de la reflectància i il·luminació, mostrem exemples de millora de balanç de blanc automàtic, re-il·luminació d'escenes, i re-acoloriment d'objectes. La teoria proposada pot ser usada per a la millora en altres camps de la visió per computador com ara donar noms als colors, detecció d'objectes, reconeixement i segmentació d'imatges.

La gran majoria de les aplicacions existents en visió per computador basen els seus mètodes simplificant les assumpcions, com ara assumint reflectàncies Lambertianes o il·luminació uniforme, per tal de ser capaços de resoldre el problema. No obstant, en escenes del món real els objectes solen mostrar reflexions més complexes (difuses i especulars) i a més, estan més afectades per les característiques i cromaticitats dels il·luminants. En aquesta tesi incorporem un model de reflexió més realistes. Per tal d'abordar aquest fenomen físic complex, estenem els models de reflectància d'objectes estàndards, introduint un model de reflexió dicromàtica multi-il·luminant (MIDR). El fet d'usar MIDR ens permet ser capaços de modelar i descomposar les reflectàncies d'un objecte amb especularitats complexes sota varies il·luminants, presentant ombres i inter-reflexions. Demostrem que d'aquesta manera podem realitzar re-colorejats d'objectes il·luminants per llums de colors i per múltiples llums de manera més realista.

A més a més, proposem un mètode d'estimació local de l'il·luminant per tal de modelar escenes que presenten una il·luminació no uniforme (p.ex: una escena exterior amb un cel blau i un sol groc, o una escena d'il·luminació d'interior combinada amb una llum d'exterior provinent d'una finestra). El mètode proposat aprofita un model probabilístic basat en grafs i resol el problema redefinint la estimació del problema com a un problema de minimització d'energies. Aquest mètode en dona estimacions locals de l'il·luminant que, de forma significativa, milloren l'estat de l'art dels mètodes de constància de color.

Una altra aportació és la captura d'un nou conjunt de dades multi-il·luminants que consisteix d'escenes complexes i amb condicions d'il·luminació tant exterior com en

condicions de laboratori. Mostrem que la nostra proposta aconsegueix millores sobre l'estat de l'art per a l'estimació local de l'il·luminant. A més proposem l'ús de dades sintètiques per a facilitar la construcció de conjunts de dades i així millorar el procés d'obtenció de dades amb informació de comparació. A continuació comparem els metodes de estimació d'imatges intrínseques i descomposició de l'estat de l'art usant el nostre conjunt de dades generat per ordinador per al cas de condicions més complexes de objectes i il·luminació (p.ex.: especularitats, inter-reflexions, i ombres).

Resumen

Los modelos más realistas y precisos de la reflectancia de objetos y la iluminación de escenas puede mejorar significativamente la calidad de muchas tareas en campos de gráficos por ordenador y visión por computador. Empleando tales modelos, se puede alcanzar un conocimiento más profundo sobre la interacción de la luz con las superficies de los objetos, que prueben ser cruciales para una variedad de aplicaciones de visión por ordenador.

En el presente trabajo, investigamos los diferentes enfoques existentes para el modelado de la reflectancia e iluminación, y analizamos sus deficiencias para capturar la complejidad de las escenas del mundo real. En base a este análisis se proponen mejoras para los diferentes aspectos de la estimación de la reflectancia y de la iluminación con el fin de modelar de manera más realista las escenas del mundo real en presencia de fenómenos complejos de iluminación. Como aplicaciones de la estimación mejorada de la reflectancia e iluminación, se muestran ejemplos en la mejora del balance automático de blanco, la re-iluminación de escenas, y la recolorización de objetos. La teoría propuesta se puede emplear con el fin de mejorar otros campos de la visión por computador, tales como el nombrado automático del color, la detección de objetos, reconocimiento y segmentación de imágenes.

La gran mayoría de las aplicaciones existentes de visión por computador basan sus métodos en la simplificación de supuestos tales como la reflexión Lambertiana o iluminación uniforme para poder solucionar sus problemas. Sin embargo, en las escenas del mundo real, los objetos tienden a exhibir reflexiones más complejas (difusas y especulares), y además se ve afectada por las características y la cromaticidad de los iluminantes. En esta tesis, se incorpora un modelo de reflexión más realista. Para hacer frente a estos fenómenos físicos complejos, extendemos los modelos estándares de reflectancia de objetos mediante la introducción de un modelo de reflexión multi-Iluminante dicromático (MIDR). Usando MIDR somos capaces de modelar y descomponer la reflectancia de un objeto con especularidades complejas bajo múltiples iluminantes que presentan sombras e interreflecciones. Se demuestra que esto nos permite realizar la recolorización de objetos iluminados por luces coloreadas e iluminantes múltiples.

Por otra parte, se propone un método de estimación local del iluminante para modelar escenas con iluminación no uniforme (por ejemplo, un escenario al aire libre con un cielo azul y un sol amarillo, o una escena con iluminación interior combinado con la iluminación al aire libre a través de una ventana). El método propuesto aprovecha un modelo probabilístico basado en grafos y resuelve el problema volviendo a definir la estimación como un problema de minimización de energía. Este método nos proporciona estimaciones de iluminante locales que mejoran significativamente respecto a los métodos del estado-del-arte de la constancia de color.

Además hemos capturado nuestra propia base de datos multi-iluminante, que consta de escenas complejas y condiciones de iluminación al aire libre y de laboratorio. Se demuestra que nuestro enfoque logra mejoras sobre los métodos estado-del-arte para la estimación iluminante local. Asimismo, proponemos el uso de datos sintéticos para facilitar la construcción de bases de datos y mejorar el proceso de obtención de información real. Se procede comparando los métodos del estado-del-arte en la estimación y descomposición intrínseca usando nuestra base de datos generado por ordenador para el caso de objetos y las condiciones de iluminación más complejas (por ejemplo, especularidades, interreflecciones y sombras).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Recent advances in imaging technologies result in high quality digital images and videos. Everyday, more and more users are encouraged to use these technologies in their work and free-time. The smartphone trends and social networks are driving people to rely on this visual medium to interact with each other. The importance of digital photography has widely increased the demand for applications which are capable of performing automatic image understanding.

The light we observe is a mixture of the object color and the illuminant color. Disentangling illumination and object reflectance is one of the fundamental challenges of computer vision. It is well-known that humans have this capacity, and it is believed to be fundamental for scene understanding. Disentangling the scene accidental illuminant from the measured light and estimating the true material reflectance is essential for accurate object recognition. Furthermore, estimating the illuminant in a scene allows for white-balancing of scenes, which is an important operation in digital cameras.

The interaction of light and object materials is modeled by reflection models. The underlying physics of these interactions has been long known [57] . However, due to the high number of unknowns of these models, solving them even for simple scenes is an under-constrained problem. Therefore, additional constraints are required to estimate the illuminant and reflectances of a scene. The most used constraint in computer vision is the Lambertian reflectance model assumption combined with a single illuminant assumption. Later, the model was extended to also include specular reflectance by Shafer [88], however the single illuminant estimate was preserved. In many situations however these basic models fail to describe the reality of the reflectances in the scene. Therefore in this thesis we investigate reflectance estimation and illuminant estimation in complex lighting situations.

Figure 1.1: Examples of multi illuminant scenes. (left) A person blocking the white sunlight, resulting in a blue shadow illuminated by the sky. (middle) Mixture of indoor and outdoor illuminants. (right) Artificial lighting environment with various colored light sources.

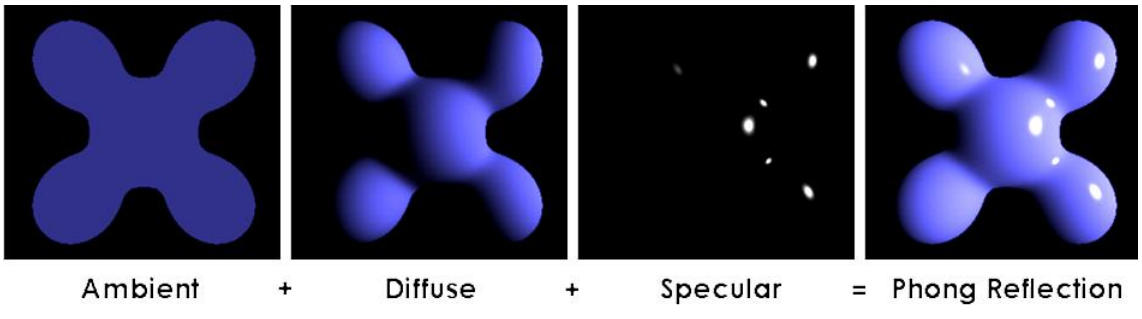


Figure 1.2: An example of image formation using the Phong model. From left to right: the ambient, diffuse, and specular components and the final image formed by their summation (the image is taken from a work by Bradley Smith).

1.1 Illumination and Object Reflectance Modeling

In real world applications, objects tend to exhibit more complex reflections due to shadows, highlights, inter-reflections, and are furthermore affected by the characteristics and chromaticity of their illuminant (see Figure 1.1). To address such complex physical phenomena, more complex reflection models are required. Using these models, more profound knowledge of light interaction with objects surfaces can be established which is crucial to a variety of computer vision applications.

Even though complex lighting situations have received more attention over the last years ([8,46,70]), there are still many open research questions. In this thesis we focus on three aspects related to computer vision in complex - more realistic - lighting situations.

1.1.1 Reflectance and illumination decomposition in complex scenes

When a ray of light hits an object surface, various phenomena occur regarding the scene optics. Part of the energy of the light is absorbed by the object and part of it is reflected. The reflection from the object mainly occurs in two ways, namely: *specular*, that is the

mirror-like reflection directly from the surface according to the surface normal direction; and *diffuse* which is the scattering of the light from the rough surface. Often, for the sake of simplicity, in the literature, the surface is assumed to be completely opaque and not a complete mirror. For the both types of reflection, the angle between the ray and the surface normal, plays an important role. Further, the viewing angle to the surface normal, controls the image formation.

Maxwell *et al.* [70] have extended the physics-based object reflectance model to account for secondary lights using a four-term equation. Their work is mainly focused on ambient light which is the cause of colored (bluish) shadows. The authors argue that due to the high complexity of such model, at the time of illuminant estimation, simplifications are needed in order to solve the equation in practice. In this sense, they approximate the model with the assumption of *Lambertian* surfaces with a two-term equation and define the ambient lighting to be constant over the whole surface. One of the popular models for object reflectance used mainly in computer graphics is the Phong model [77] in which the object image is formed by the diffuse, specular, and ambient reflection components. Figure 1.2 shows an example of an image constructed using this model.

Even though some more complex reflection models exist, estimating these models in real-world scenes is still an open problem. Finding algorithms which can estimate reflection and illumination in complex scenes outside the controlled laboratory settings is believed to be important for many computer vision applications, such as color description, photo editing and object recognition.

1.1.2 Color constancy in multi-illuminant scenes

The amount of light for any given wavelength reflected from a surface changes with regard to the different types of light illuminating the scene. We perceive the color of a surface to be the same, even though there are changes in the brightness and color of the light. It would be confusing if the objects would change color when seen under different lights. We often take advantage of object colors to classify them (e.g. a ripe banana is yellow or grass is green no matter if viewed in the middle of a sunny day or at the sunset). Our brain is able to discount the effect of the scene illumination on the object colors reflected from their surfaces. This relative stability of the colors in human perception is referred to as *color constancy*. In computational approaches, this phenomenon is simplified to estimation of the scene illuminant and correcting the image colors to look as if captured under a canonical light (e.g., white light). Figure 1.3 shows an example of a scene obtained under varying lighting.

There are various color constancy methods in the literature. One of the most popular categories is the *static* approaches in which the scene is modeled by preset parameters. These methods use the color distribution of the scene pixels to infer the illuminant (e.g., averaging the scene colors as in *grey-world* or using the color of the specular pixels as in the physics-based approaches). Another popular method is to infer chromaticity of the illuminant by choosing the light which most likely produces the subspace of color space

Figure 1.3: Example of scene under various illuminations [6]. Computational color constancy aims to estimate the illuminant color in a scene.

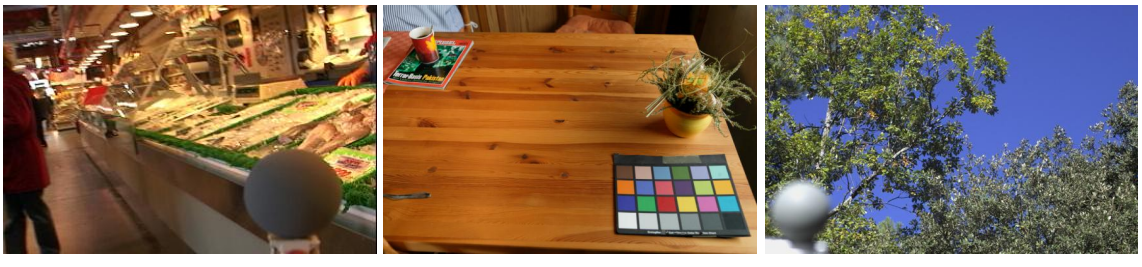


Figure 1.4: Example of existing data sets for color constancy. (left) Example from the SFU data set [19], (middle) Example from Gehler *et al.* [37], (right) example from the Barcelona data set [75].

formed by the input image pixel colors. And finally, there has been recent advancements in this field using training on image statistics or categories as well as probabilistic approaches. The *Automatic White Balance* (AWB) embedded in digital cameras is a very popular application in digital photography. Most digital cameras use statistical methods to correct for the scene illumination. While existing color constancy methods appear to perform well on many image datasets, they mostly fail in the case of complex scenes. An example, is the case of having multiple illuminants with strong effects in an image. This will confuse the automatic white balancing method that is trying to estimate one illuminant for the whole image.

Real-world scenes often exhibit more complex illuminant conditions than what is commonly assumed in many color constancy algorithms. That is, the illumination is not completely uniform over the scene. This could be due to multiple lights present in the scene or simply the ambient light and inter-reflections between objects. In the simplest case with no specular surfaces and no extra lights, the bounces of the light emitted from the light source over the various surfaces present in the scene can act as different illuminants since the interaction with the surface can affect the chromaticity of the re-emitted light[1]. Another example of a 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.

As explained above, in many indoor and outdoor scenes multiple illuminants are

[1]For more details on this phenomenon refer to Section 5.3.3

present. But due to the complexity of multi-illuminant estimation in practice, so far only very little work has been dedicated to this problem (e.g., [46]). We believe that proposing new algorithms for local illuminant estimation is important for scene understanding and improved white balancing in particular.

1.1.3 Ground Truth data sets in complex lighting

Humans strongly rely on their memory for many cognitive behaviors in their daily life. The process of learning in human-beings starts from very early age and continues for all their lives. Using various examples, human brain can extract the main features that members of the same class share with each other. Similarly for computers, *image datasets* play the crucial role as memory for many computer vision algorithms. While these methods use training data to estimate their parameters and construct their models of real world, the performance of the majority of the algorithms needs to be verified using test data. In this context, a large collection of data with variety in context and imaging conditions along with an accurate and reliable ground-truth information is a valuable asset.

Currently only very few medium size or relatively large datasets are available with illumination or reflectance ground-truth data. Figure 1.4 shows some examples of color constancy datasets([6, 19, 37]). In most existing color constancy image datasets, the illumination is assumed to be uniformed and given by a reference in the scene (e.g., gray ball or Macbeth color checker). Intrinsic image decomposition methods are often trained and tested on the MIT dataset [49] which is limited only to white light.

The main challenge in constructing datasets of complex lighting scenarios is the complicated process of obtaining the ground-truth data for a scene with non-uniform illumination. Capturing local changes in the scene illumination requires accurate and reliable pixel-wise illuminant calculation. Therefore, so far, the majority of these datasets are only focused on simpler scenarios and there is a need for large scale image datasets.

1.2 Objectives and approach

Above we have indicated three aspects of computer vision in complex scenes. In this thesis, we aim to tackle these issues in order to advance our understanding of illuminant and object reflectance modeling in complex real-world scenes.

1.2.1 Reflectance and illumination decomposition in complex scenes

While the previous works in this field were limited to rather simple scenes (e.g., laboratory data) or used assumptions like Lambertian reflectance or a known illuminant, our goal is to extend existing physics-based models to account for complex illumination, shapes, and surface properties. Therefore, in Chapter 3, we attempt to solve the reflectance model with

minimal assumptions and to form an accurate decomposition of the object reflectance in an arbitrary uncalibrated image.

The key ingredient in our approach is the use of segmentation based reflectance and illumination estimation. That is to extract pixels which belong to a single-colored surface, but still present a complex distribution in the color space as their brightness is effected by the surface geometry and specularities and their color is influenced by the different chromaticities of the illuminants in the scene. Using this extra cue we are able to model multi-illuminant scenarios which can also include inter-reflections and colored shadows. Achieving a realistic decomposition enables us to perform highly realistic object recoloring and physics based color transfer.

1.2.2 Color constancy in multi-illuminant scenes

Many Computer Vision applications use color properties as discriminative features to conduct fundamental tasks like object segmentation, classification, and other cues for scene understanding. To this end, it is crucial to estimate the colors of objects in the scene independent of their illumination condition. Conventional methods often limit themselves to simplifying assumptions over the illumination (e.g., uniform illumination). In real world images, objects tend to exhibit more complex reflections due to shadows, highlights, and are furthermore affected by the characteristics and chromaticity of their illuminants. Therefore, a more general model of illumination and reflectance which can handle multiple illuminants is required.

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 enables 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 are able to develop a more intelligent automatic white balancing method.

1.2.3 Ground Truth data sets in complex scenes

Scene decomposition into its illuminant, shading, and reflectance intrinsic images is an essential step for scene understanding. For the evaluation and comparison of algorithms in the field, having an image dataset with reliable ground truth data is essential. However collecting intrinsic image ground-truth data is a laborious task. And the assumptions on which the existing ground-truth procedures are based limit their application to simple scenes which are not always representative of the real world. In Chapter 5 we provide details on data acquisition and ground truth computation.

In the first part of this chapter, we focus on the data captured using digital cameras from both real-world and laboratory scenes which present complex reflectance and illumination conditions. The aim of this dataset is to provide images with accurate pixel-wise illumination ground-truth for multi-illuminant scenes. The scenes include various examples of colored shadows and contain various types of surfaces and materials.

The second part, presents our synthetic image dataset constructed using graphical but physically accurate 3D modeling and rendering software. We investigated synthetic data for intrinsic image research since the extraction of ground truth is straightforward, and it allows for scenes in more realistic situations (e.g, multiple illuminants and interreflections). With this data set we aim to motivate researchers to further explore intrinsic image decomposition in complex scenes. Moreover we compare the two different methods presented for constructing and collecting our datasets and investigate the advantages and disadvantages of each method.

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 image segmentation and object 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 classification*).

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:

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

where b is the surface albedo, $e(\lambda, \mathbf{x})$ 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

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

to denote vectors. 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 $\rho^c(\lambda), c \in \{R, G, B\}$ at position $\mathbf{x}$ is modeled as $f^c(\mathbf{x})$ by integrating over the visible spectrum ω,

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

We define the *body reflectance* as

$${c_b}^c(\mathbf{x}) = \int_\omega b(\lambda, \mathbf{x}) e(\lambda) \rho^c(\lambda) d\lambda \ . \tag{2.3}$$

Therefore,

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

In spite of its 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 [88]. 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.

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

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:

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

The reflection of the light consists 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.

Several methods have been developed to approximate the dichromatic model of an object. Kravtchenko and Little [56] 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. Shen and Xin have solved the model with the assumption of a known illuminant [89].

The original application to which the DRM was applied, was the separation of shading from specularities [88]. 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 [53, 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 [41, 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 [88] models the diffuse light, $\mathbf{a}$, by a third term

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

Later work improved the modeling [70, 82] 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 presence of multiple illuminants in the scene (a more generalized case of ambient light). A typical example of a "multi-illuminant" scenario is the interreflections occurring between objects in 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.

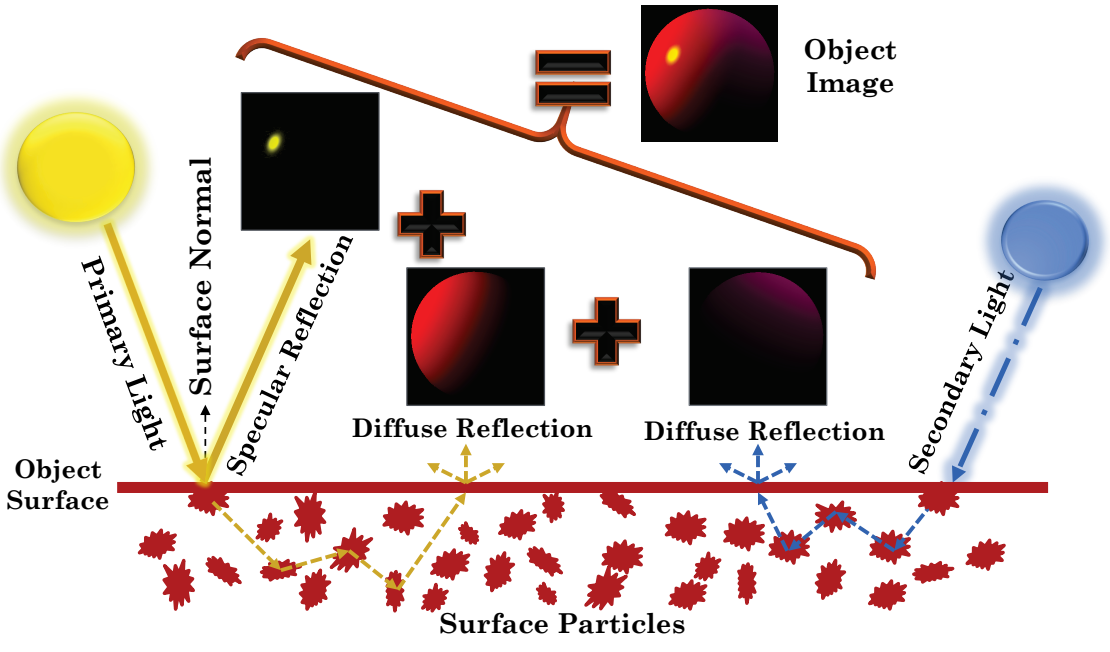


Figure 2.1: Real-world objects often exhibit body and surface reflection under more than just one illuminant (e.g., outdoor scene with blue sky and yellow sun).

2.1.4 Inverse-Intensity Chromaticity Space

Using the Dichromatic Reflection Model, Tan *et al.* [92] formulate 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:

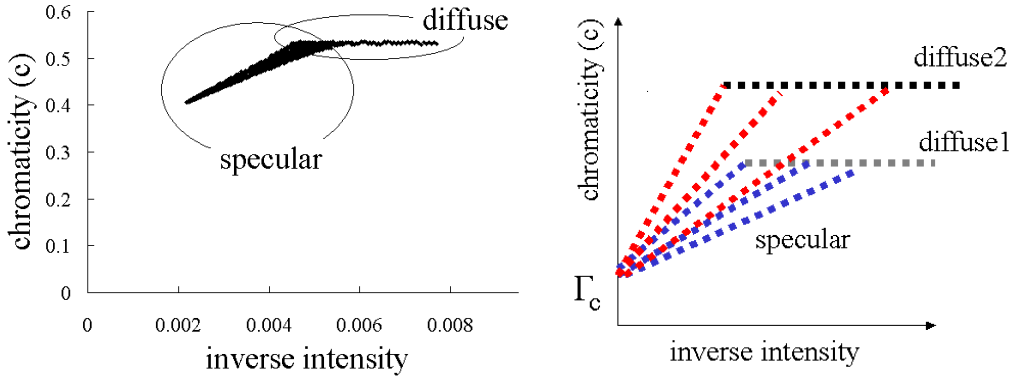
$$B_c(\mathbf{x}) = \int_\omega b\left(\lambda, \mathbf{x}\right) e\left(\lambda\right) \rho^c\left(\lambda\right) d\lambda$$

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

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

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



(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; (b) Diffuse and specular points of an image. Images are taken from Tan *et al.* [92].

where $p = m_b(\Lambda_c - \Gamma_c)$. This demonstrates that we are able to determine the chromaticity of the illuminant, Γ_c, using p since the image chromaticity, σ_c, and total image intensity, $\sum I_i$ 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 $\sum I_i$ 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.

2.2 Intrinsic Images

As mentioned in the previous section, 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 [10] 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 [104]. 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.* [49] 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 (see Figure 2.3). Since then many methods have been developed and tested using this dataset.

Figure 2.3: Example of intrinsic image decomposition. From left to right: the original image, reflectance image, and shading image. Images are taken from the MIT intrinsic image data set [49].

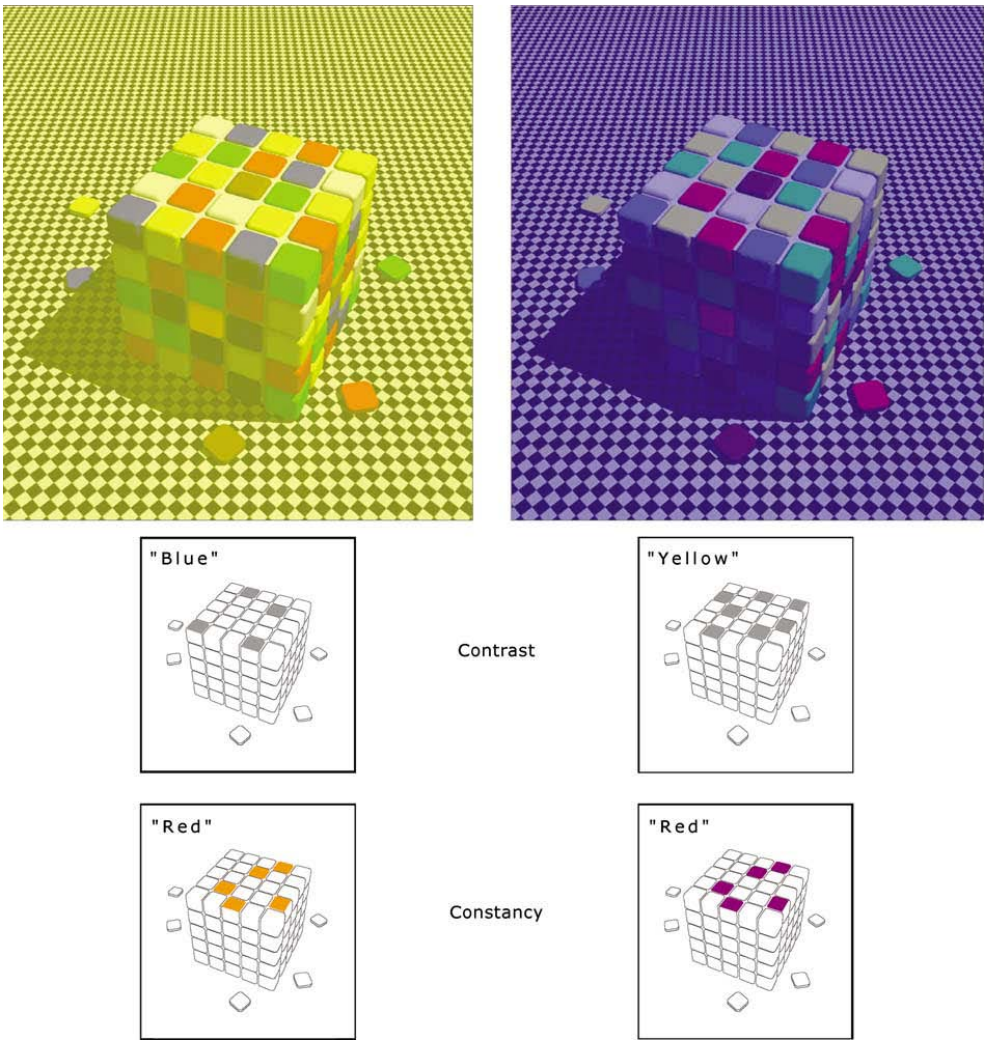


Figure 2.4: This figure demonstrates the effect of color constancy on the perception of colors. The first row is an example of such effect using two different surrounding colors. Images on the second row show that the bluish tiles in the image on the top of the left cube are identical to the yellowish tiles on the top of the cube in the right, and all are in fact gray. Also the red tiles on the top of both cubes, even though appearing identical, are in fact different colors as demonstrated in the third row. Figures are taken from Lotto *et al.* [66].

Gehler *et al.* [38] 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 the 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.* [87] who also show competitive results obtained on the MIT dataset. In this work, *ridges* have been used as extra cues to improve the results [101].

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 estimate and intrinsic image decomposition. 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 has more complex 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 changes in observer position, posture, and movement."* [21]

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" [25] defines this phenomenon as the human ob-

servers' 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, goldfishes, and monkeys (the last are believed to poses the vision which is most similar to humans).

Regarding the reasons why color constancy could occur in human vision, there are various studies in the literature. These studies converge in a number of possibilities. For example Goldstein in "sensation and perception" [47] 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 existence of other 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 2.4 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 image 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 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.

Color Correction

Using the color correction 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 [103] transform which independently scales each of the cone responses or RGB channels:

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

where $d_i = \frac{e_i}{\sqrt{3(e^2{}_R + e^2{}_G + e^2{}_B)}}$, $i \in \{R, G, B\}$. 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 [31] (which ideally becomes zero for von Kries model):

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

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 [96] or that the convex hull of the pixels in a suitably chosen color space encompasses most illumination changes as in gamut mapping [34]). 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 [44].

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 wavelength. 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* [59] 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 [96]. 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 Minkowski norm [32], a generalized formula is introduced in the literature which can incorporate all the above mentioned methods:

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

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. Figure 2.5 shows examples of white-balance performed using the illuminant estimate by different static methods and their relative errors.

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 a higher Minkowski norm will emphasize larger measurements in the image, while lower values will equally distribute 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.

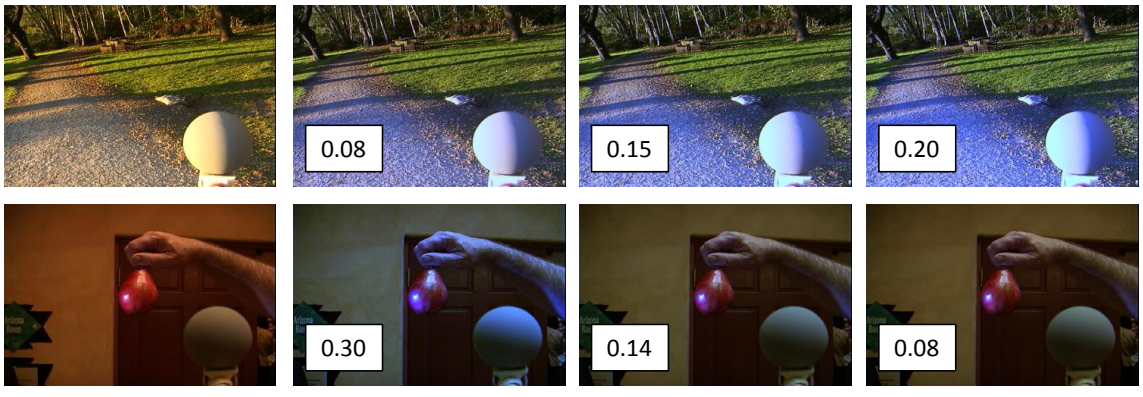


Figure 2.5: Examples of statistic based color constancy for two scenes from the SFU data set [19]. Each row from left to right demonstrates: the original image and results using gray-world, gray-edge, and second order gray-edge.

2.3.4 Physics-based Illuminant Estimation

Although many illuminant estimation method make the 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. 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.* [92]. Although this work has presented a very elegant formulation based on the original work by Shafer [88], 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 the 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 the canonical illluminant.

Existing gamut mapping methods, although being based on the original work pioneered by Forsyth [34], 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* [33] 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, 6]. Other approaches tried to address this problem by enlarging the input or the canonical gamut using different heuristics [3, 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 exploit derivatives to better estimate the illuminant [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. Figure 2.6 shows an example of color constancy using gamut mapping.

2.3.6 Learning-based Methods

In this section we make a brief review of the illuminant estimation methods which learn their model using training data.

Color-by-correlation: In this approach a *correlation matrix* is created for every possible illuminant to be considered. The correlation matrix for a known light l_i is computed by partitioning the color space into a finite number of cells and computing the probability of occurrence of each coordinate under that light. Then the information from the input image is matched to the information from the correlation matrices and the probability of

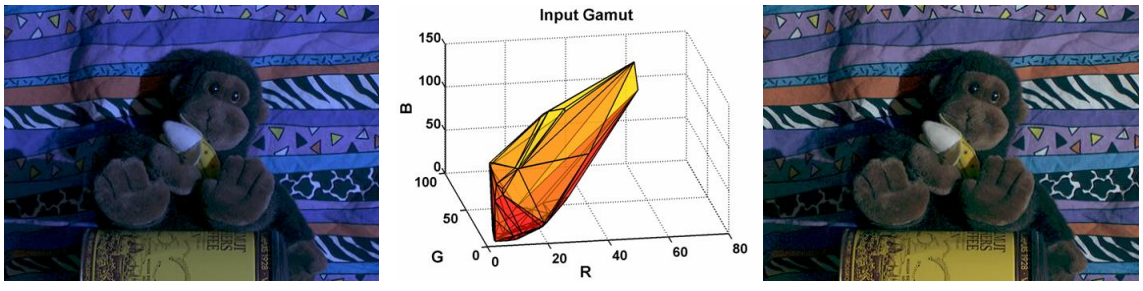


Figure 2.6: An example of a scene from [6] and its gamut (middle). On the right is the image after gamut mapping has been applied.

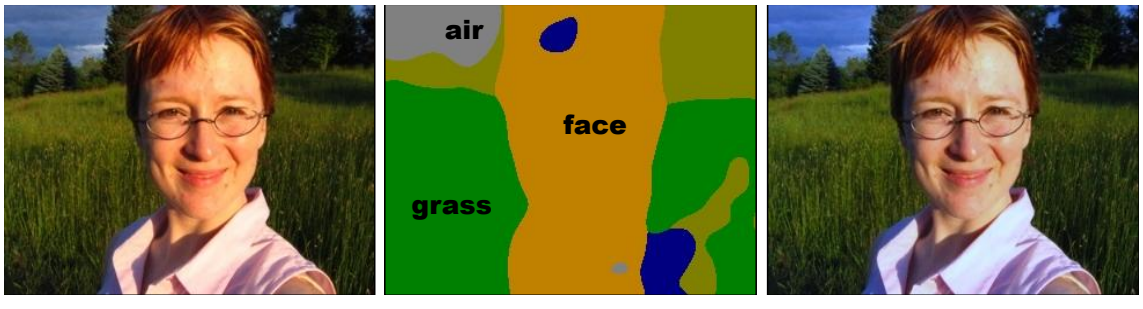


Figure 2.7: Example of top-down color constancy [98]. Based on the correct identification of the face and grass in this scene the algorithm succeeded in estimating the illuminant color.

each light being the match is calculated. This probability defines the likelihood of the image being taken under that light. Using this approach, Finlayson *et al.* [30] then proceed to choose the light using the *maximum likelihood* as the final estimate. Note that although this technique has some similarities with the gamut mapping technique, it is a more general framework which also includes low-level statistics-based methods such as gray-world and white-patch.

Bayesian color constancy: These approaches model the variability of reflectance and light sources as random variables. The illuminant is then estimated from the posterior distribution conditioned on the image intensity data [16]. Using the assumption that nearby pixels are correlated, Rosenberg *et al.* [84] replaced the Gaussian distribution assumption with nonparametric models. Further, Gehler *et al.* [37] were able to achieve high performance using their color constancy dataset proving that the Bayesian approach can strongly benefit from using precise priors for illumination and reflectance.

Natural image statistics: To improve the illuminant estimation results and in order to choose the correct color constancy method, various method have been proposed. Gijsenij and Gevers [42] used the fact that each method makes assumptions on the color distribution in the input image (e.g., average color or edge is achromatic under neutral light) to combine several color constancy approaches. They argue that *natural image statistics* can be used in order to describe these distributions in an image. They state that this fusion

improves the performance over individual methods.

High level cues: Using prior knowledge about the world, van de Weijer *et al.* [98] proposed to select the best illuminant among a set of illuminant hypothesis. In this sense, possible illuminants are considered based on the likelihood of their semantic content in order to generate the most *plausible* scene. Images are modeled as a mixture of semantic classes (e.g., water, grass, or sky) using a distribution over visual words which contain 3 cues namely: color, texture, and position (described using Gaussian RGB average, SIFT [67], and regular grid respectively). The light source which leads to the most likely scene (i.e. the one in which the grass is green and the sky blue) is selected among the candidate light sources (see Figure 2.7). A more specific high-level information source is used by Bianco and Schettini [13], who use the color of skin tones (predominantly faces) to estimate the illuminant in a scene. Finally, Vazquez et al. [102] show that color categories can also be applied as a top-down mechanism to improve illuminant estimation.

In Chapter 4 we explain how we use *Conditional Random Field* (CRF) to combine bottom-up color constancy approaches in order to tackle the problem of illuminant estimation in complex multi-illuminant scenes.

2.3.7 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.* [24] proposed a method for solving the case of non-uniform illumination by computing the local average color in the image. Recently Bleier *et al.* [14] 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 superpixel. Also using a similar approach Riess *et al.* [81] have extended the [92] to obtain an estimate for local illumination. The work published by Gijsenij *et al.* [46] 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.

2.4 Conclusion

In this chapter, we have presented an overview of the different topics addressed in this thesis and briefly discussed some of the main studies and methods in the literature regarding them. Firstly, we explained object reflectance modeling and reviewed various existing models. Then we described the concept of intrinsic images followed by a summery of methods for image decomposition into its intrinsic components. Lastly, we reviewed the color constancy phenomenon in human vision and the main computational approaches for illuminant estimation. In each section, we pointed out the strengths and shortcomings of the state-of-the-art approaches. In the next chapters, we will focus on the possibilities for improvement in each of these aspects.

Chapter 3

Object Recoloring based on Intrinsic Image Estimation[1]

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 inter-reflections, results in unrealistic recoloring of objects.
In this chapter, 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.

[1]This chapter has appeared in ICCV 2011 [11].

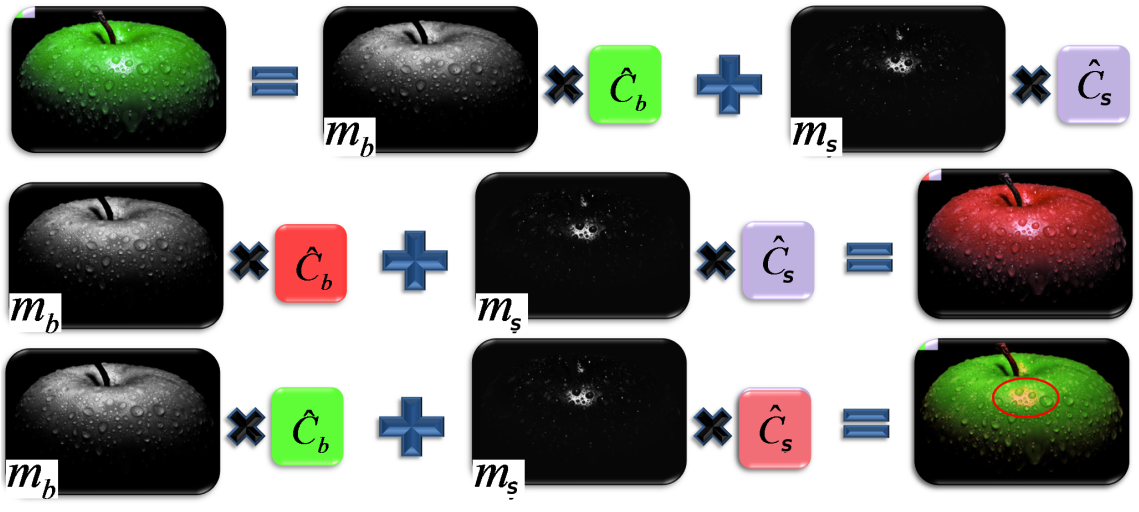


Figure 3.1: The first row is an example of the reflectance decomposition achieved by DRM [88]. Using this decomposition, *object recoloring* is performed by changing the *body reflectance* (the second row), and *illuminant recoloring* is achieved by changing the *specular reflectance* (the third row).

One of the most popular color modification applications is the recoloring of a specific 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 [10]. 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) [88] 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 chapter, 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 [62, 85]. 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 [104] 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. [93] 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 $m_s \neq 0$. Klinker et al. [53] 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. [91] proposed an iterative method for reflectance decomposition of textured surfaces. Tan at al. [90] improve

the previous methods by adding spacial distribution and texture constraints when available. Mallick et al. [68] use 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 [23, 55]. 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 [78,80]. Local color transfer [36,105] and user-interactive methods [2] 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 [48]. 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 illuminants.

Omer et al. [74] present an *image specific* 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 [89] in which the object (body) color is estimated and transfered between images. Realistic results on lab conditioned high quality images of objects under single *known* illuminant are presented.

Hsu et al. [50] 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 $\mathbf{c_s}$ [88]. Validity of the DRM has been proven for a variety of inhomogeneous dielectric materials commonly observed in natural scenes [94]. In this chapter, 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 propose 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

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

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 [88] 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 [53, 68, 72, 91]. In this chapter, 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 [53, 72], and known illumination [68,91]. 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 [53, 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 $\hat{\mathbf{c_b}}$ 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* [99] 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, $\mathbf{cL}$ and $\mathbf{l}$. The error, Equation 3.6, can now be rewritten as (we will omit the spatial arguments):

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

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

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

and G^m stands for the Gaussian function at scale m.

The main difference with the ordinary LS estimator is that here the matrix η is dependent on $\hat{\mathbf{c_b}}$. Eq 3.10 can be solved by a *fixed 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 visualizes this algorithm on an example data.

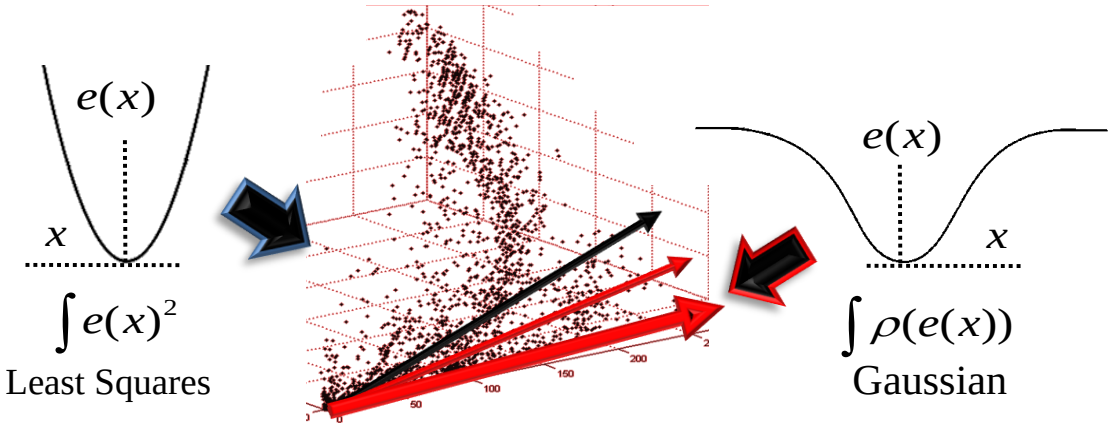


Figure 3.2: Visualization of RBRE algorithm on an example data (the red car in Figure 3.4-(a)). Object pixels of the non-specular ($m_s = 0$) form a line ($c_b = Lc$) passing origin. A least square estimation (the black arrow) will mix the two main directions in data. In a Gaussian error norm large derivations from the model are considered outliers and are not taken heavily into account. Therefore using Gaussian error norm we iteratively converge to the c_b direction (the thick red arrow)

Figure 3.3: 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 Confined 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 [27]. We propose to use this constraint to estimate the illuminant.

We sample Planckian colors ($T \subset 1000K \sim 40000K$) which vary from orange to yellow to white to blue, resulting in a set of illuminants $\{\mathbf{l}_1, ..., \mathbf{l}_m\}$.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.

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

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 $\hat{\mathbb{C}}$, is based on the convex optimization problem:

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

Fig 3.3 demonstrates an example of intrinsic images recovered for an object. Note that the specular reflectance is correctly separated from the body reflectance.

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 $\mathbf{x} \in Mask$
3: Estimate the Planckian illuminant $\mathbf{l_1}$ using CIE method
4: $\mathbf{c} \leftarrow diag\left(\mathbf{c_b}\ \mathbf{L_1}^{-1}\right)$
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 $\mathbf{x} \in Mask$
8: Estimate the Planckian illuminant $\mathbf{l_1}$ using CIE method
9: $\mathbf{c} \leftarrow diag\left(\mathbf{c_b^1}\ \mathbf{L_1}^{-1}\right)$
10: Estimate $\mathbf{c_b^2}$ using RBRE for the object pixels $f(\mathbf{x}) \notin Mask$
11: $\mathbf{L_2} \leftarrow diag(\mathbf{c}_b^2)/diag(\mathbf{c})$ (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 ${m_s}^1$.

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

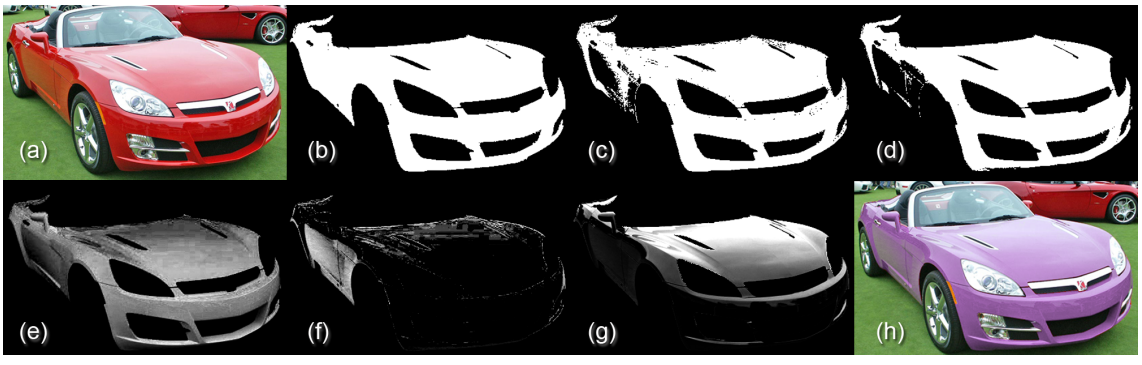


Figure 3.4: An example of the MIDR algorithm performance: (a) Original image; (b) Object mask as the initial *Mask* for the illuminant l_1; (c) The Mask after 1st iteration; (d) The Mask at 3rd (final) iteration; (e) estimated m_b^1; (f) estimated m_b^2 (the interreflection area,l_2, has been correctly detected); (g) estimated m_s^1; (h) An example recoloring (the interreflection is preserved).

$m_s^1(\mathbf{x})m_b^2(\mathbf{x}) = 0$. 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 7.1, and estimates m_b^1 and m_b^2 constraining them to be positive.

In Fig 3.4 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.

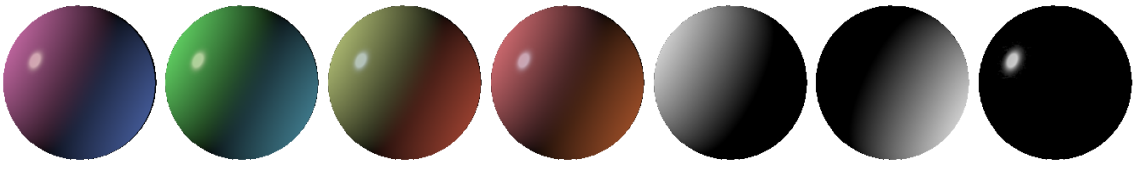


Figure 3.5: The first four images are examples of the synthetic images. The last three images are the m_b^1, m_b^2,and m_s^1 ground truth.

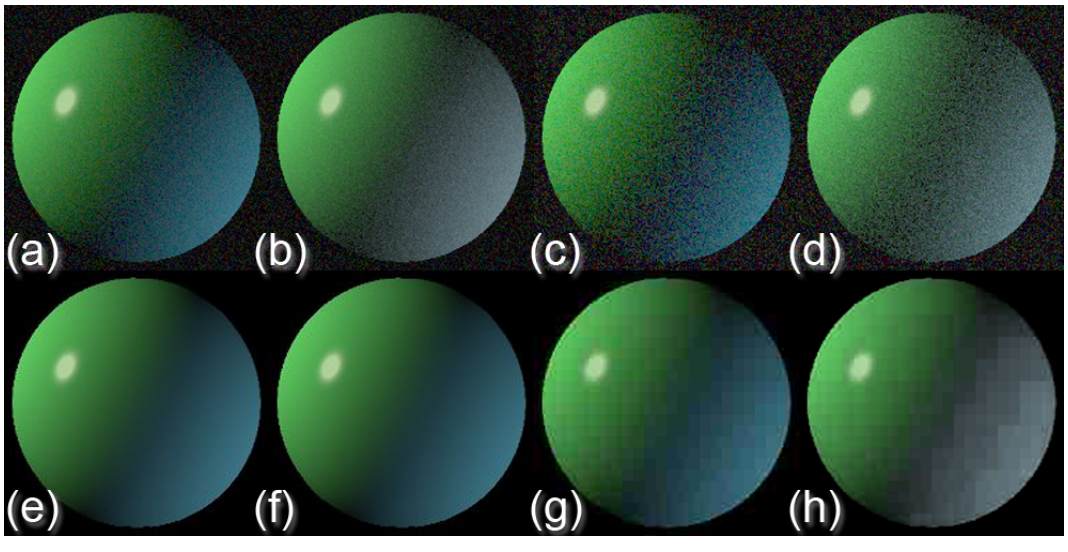


Figure 3.6: Effect of noise and JPEG compression: (a) and (c) examples of applying noise by sigma 4.0 and 9.0; (b) and (d) their corresponding reconstructions; (e) and (g) examples of applying JPEG compressions of 20% and 80%; (f) and (h) their corresponding reconstructions.

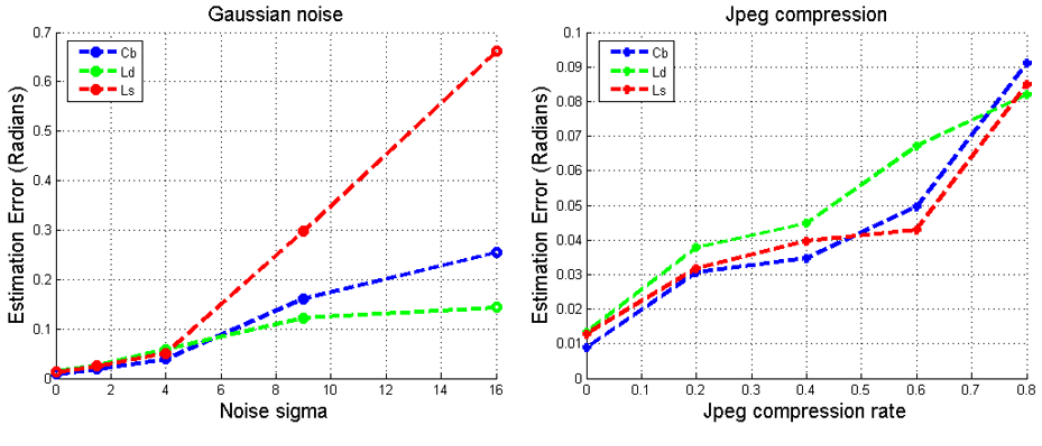


Figure 3.7: Median angular error (in radian) as a function of: Gaussian noise sigma (left) and JPEG compression (right) for c_b, l_1 and l_2 estimates.

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 3.5). With these we generate a set of 60 test images by varying the illuminants and the object color. Some examples are given in Fig 3.5. 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 3.7). 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 3.6.

3.6.2 Real-world Images

Fig 3.9 compares MIDR-based recoloring with the one done by hue-saturation shift method. The secondary illuminant (greenish interreflection) is correctly preserved by MIDR while wrongfully changed to blue by the professional photo-editor. In Fig 3.10 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 3.8.

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: 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.

Figure 3.9: 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 3.10: 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 3.11 compares the physics-based color transfer performed using MIDR and DRM models with the methods from [78, 80]. 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 [78, 80] mixing the illuminants and object colors resulted in unrealistic images. Furthermore, the resulting objects exhibit different colors than the target objects.

Photo Fusion. Fig 3.12 is an interesting example made possible by our method. The car in Fig 3.4 is copied into another scene. Here the object is recolored using the estimated

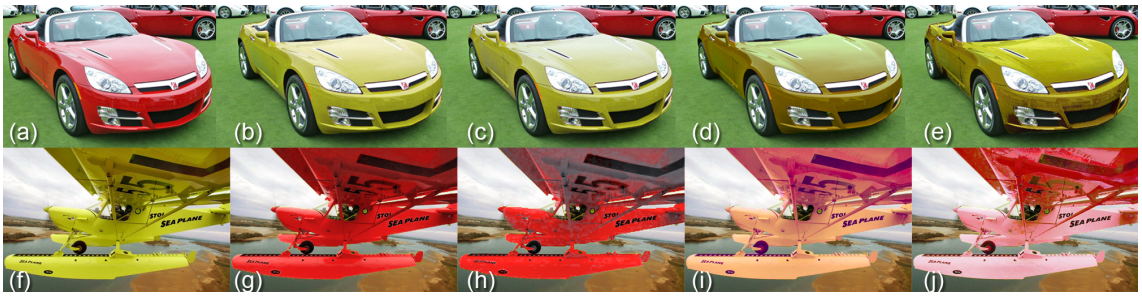


Figure 3.11: Comparing the *Color Transfer* results by DRM, and [78, 80]. (a) and (f) Original images;(b) and (g) MIDR results; (c) and (h) DRM results; (d) and (i) results by [80]; (e) and (j) results by [78] . Note that the secondary illuminants (interreflections) on the side of the car and the plane wing are lost in (c) and (h), wrongfully transformed in (d), (e),(i), and (j), while being preserved in (b) and (g).

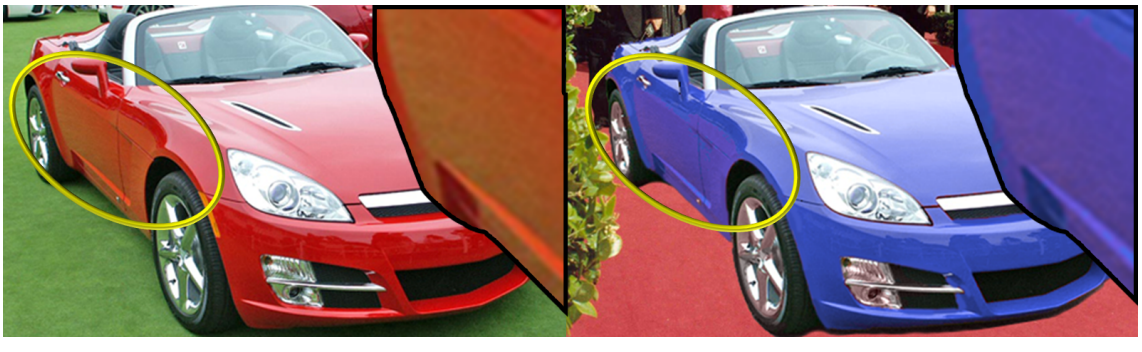


Figure 3.12: An example of photo montage: The interreflection of the green grass (zoomed area) in the original image is re-lighted by the red color of the carpet to match the target scene.

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 [1]

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

[1] The materials in this chapter are used in a journal submission with the same name by Shida Beigpour, Christian Riess, Joost van de Weijer, and Elli Angelopoulou

Figure 4.1: Sample images from our database. The first row contains the original photographed images (transformed from their raw format to sRGB only for the purpose of visualization). The bottom row shows the relative influence of the two incident illuminants color-coded in blue and red.

scenes. Examples of multi-illuminant pictures, and the color-coded pixel-wise influence of each illuminant, can be seen in Fig. 4.1.

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 [17, 32, 45, 96] depends on the size of the statistical sample. Applying these methods to small image regions introduces inaccuracies [14] and is unlikely to yield stable results. Physics based methods either assume purely diffuse scenes, e.g. [16, 40] which is not often applicable in real scenes, or exploit the presence of specularities in an image, e.g [92], 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. Fur-

thermore, 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, our main contributions in this chapter 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.

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 [17] 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 [58]. It was noted by Gershon *et al.* [39] 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 [32] 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.* [45] 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. [16, 40] assume the scene is entirely composed of diffuse surfaces, while others e.g. [60, 92] exploit the presence of specular highlights. These latter methods are based on the dichromatic reflection model [88] 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., [88, 92]).

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.* [43] 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 [35, 44, 63].

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 [24] 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 [50]. More recently, Kawakami *et al.* [52] 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.* [46] 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.* [50] 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 [54]. 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 $\mathcal{V} = \{1, 2, ..., N\}$ 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 $i \in \mathcal{V}$ is associated with a random variable $\mathbf{X}_i \in \mathcal{X}$, which can take on a value $\mathbf{x}_i$ from the illuminant-color label set $\mathcal{L} = \{\mathrm{l}_1, \mathrm{l}_2, ..., \mathrm{l}_k\}$. At each node $i \in \mathcal{V}$ 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 $\breve{\mathbf{x}} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ 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

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

where $\xi_c(\breve{\mathbf{x}}_c|\mathcal{F})$ are potential functions defined over the observations $\mathcal{F}$ and the variables $\breve{\mathbf{x}}_c = \{\mathbf{x}_i, i \in c\}$ 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:

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

where ϕ denotes the unary potential and ψ the pairwise potential. The unary potentials ϕ 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 ψ 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 [17,32,96]. 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 $P = \{p_1, p_2, ..., p_N\}$ 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

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

where $\|.\|$ 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

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

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

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

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

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

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 7.2 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

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

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

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

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 this chapter, 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:

$$\phi^s(\mathbf{x}_i|\mathcal{F}_i) = \left(w^i\right)^q \rho_{\sigma_r}\left(\varphi\left(\mathbf{i}_i, \mathbf{x}_i\right)\right) \ . \tag{4.12}$$

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

$$\phi^s(\mathbf{x}_i|\mathcal{F}_i) = \left(1 - \cos\left(\varphi(\mathbf{x}_i, \mathbf{i}_i)\right), \right) \ . \tag{4.13}$$

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 $q \in \left\{0, \frac{1}{2}, 1\right\}$.

4.4.2 Physics-based Color Constancy

Another family of color constancy methods is based on the physics of light and surface interactions [16, 40, 60, 92]. In this section, we focus on the approach by Tan *et al.* [92] 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.* [81] 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.* [92] exploits the inverse-intensity chromaticity (IIC) space. IIC is a two-dimensional space where the horizontal axis represents the inverse intensity $1/\|\mathbf{f}_j\|_1$ and the vertical axis is given by a pixel's chromaticity. Thus, a pixel $\mathbf{f}_j = (f_j^R, f_j^G, f_j^B)^T$ is mapped to

$$f_j^c \rightarrow \left(\frac{1}{\|\mathbf{f}_j\|_1}, \frac{f_j^c}{\|\mathbf{f}_j\|_1} \right) \quad . \tag{4.14}$$

where $\|\cdot\|_1$ is the L_1-norm and $c \in R, G, B$.

The advantage of IIC is that the relationship between image chromaticity and illumination chromaticity becomes linear. According to [92], one generates per color channel a scatter plot 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_j^c.

Tan *et al.* [92] identify potential specular regions by thresholding on brightness and saturation values, an approach that was originally proposed by Lehmann and Palm [61]. 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 [81]. 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 [81]. 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

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

where $\mathcal{H}_{\mathrm{IIC}}^c(i^c)$ denotes the count for i^c in H_{IIC}^c. For further details, please refer to [81].

Our physics-based unary potential is then defined as:

$$\phi^p\left(\mathbf{x}_i|\mathcal{F}_i\right) = w_p \rho_{\sigma_r}\left(\varphi\left(\mathbf{i}_p, \mathbf{x}_i\right)\right) \tag{4.16}$$

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

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

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

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

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 work, we use a simple constraint where we exclude illuminants which are too saturated, i.e., $\{\forall i|\varphi\left(\mathbf{l}_i, \mathbf{i}_w\right) < \phi_d\}$ where $\mathbf{i}_w = \frac{1}{\sqrt{3}}(1,1,1)^{\text{T}}$ 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, $\psi\left((\mathbf{x}_i, \mathbf{x}_j)|(\mathcal{F}_i, \mathcal{F}_j)\right)$ is to ensure, when appropriate, the smooth transition of labels in neighboring vertices. Similar to Boykov *et al.* [15] we consider pairwise potentials that resemble a well. In MRFs, especially as described in [15], $\psi(\mathbf{x}_i, \mathbf{x}_j) = u(1 - \delta_{ij})$, where u is the well "depth" and the function $(1 - \delta_{ij})$ controls the shape of the well. In [15], 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 [**?**]) the "depth" depends on the observations $h(\mathcal{F}_i, \mathcal{F}_j)$. Thus, our pairwise potential function has the form:

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

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:

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

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.* [65] 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 [58] 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 [95] 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 (MIMO) Dataset

Several datasets are available for single illuminant estimation. The first datasets where taken under laboratory settings with controlled illumination [4, 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 [37]. Gijsenij et al. [46] 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[2].

[2]From correspondence with the authors of [46], 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 $\mathbf{l}_j \in \mathbf{L}$ **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 [46], 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.

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 single illuminant. The basic idea is that we can use the single illuminant images to derive

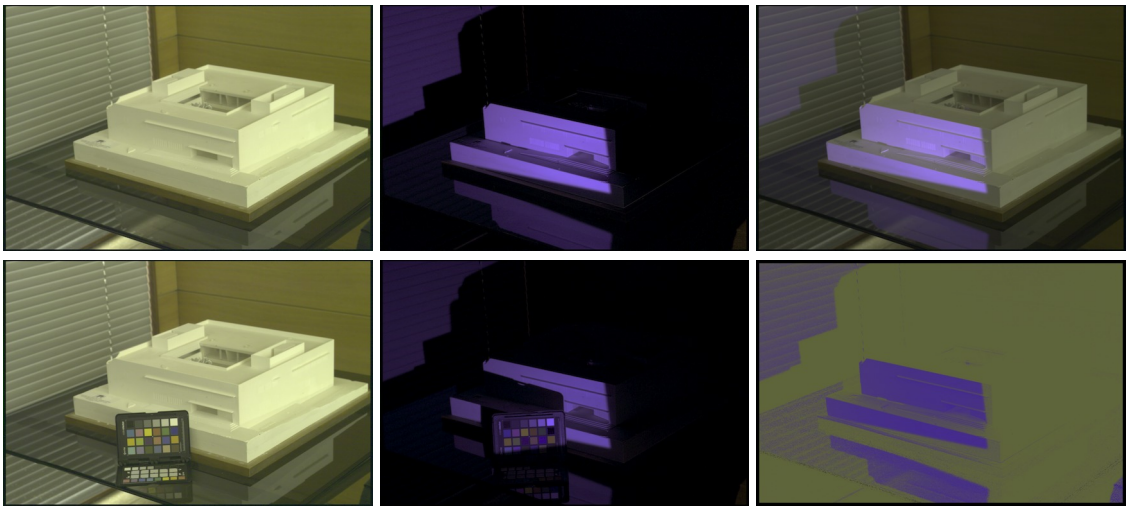


Figure 4.2: Example input images to compute the ground truth. In the top row, from left to right: scene under illuminant 1, 2, and a mixture of 1 and 2. In the bottom row: scene under separate illuminants 1 and 2 again, but with a Macbeth color chart to estimate the illuminant chromaticity. On the bottom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

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 lighting 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 $\hat{f}_{\mathbf{a},g}$ and $\hat{f}_{\mathbf{b},g}$, respectively. The relative difference for illuminant $\mathbf{a}$ in the green channel is obtained for each pixel as

$$r_{\mathbf{a},g} = \frac{\hat{f}_{\mathbf{a},g}}{\hat{f}_{\mathbf{a},g} + \hat{f}_{\mathbf{b},g}} \quad , \tag{4.24}$$

See appendix 7.3 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 $5.9°$ between blue and white, $6.1°$ between white and red, and $11.4°$ 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 both illuminants. A stronger blue component denotes stronger influence of the left illuminant, while red represents 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 images are processed without gamma to preserve the linearity of the data). 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 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

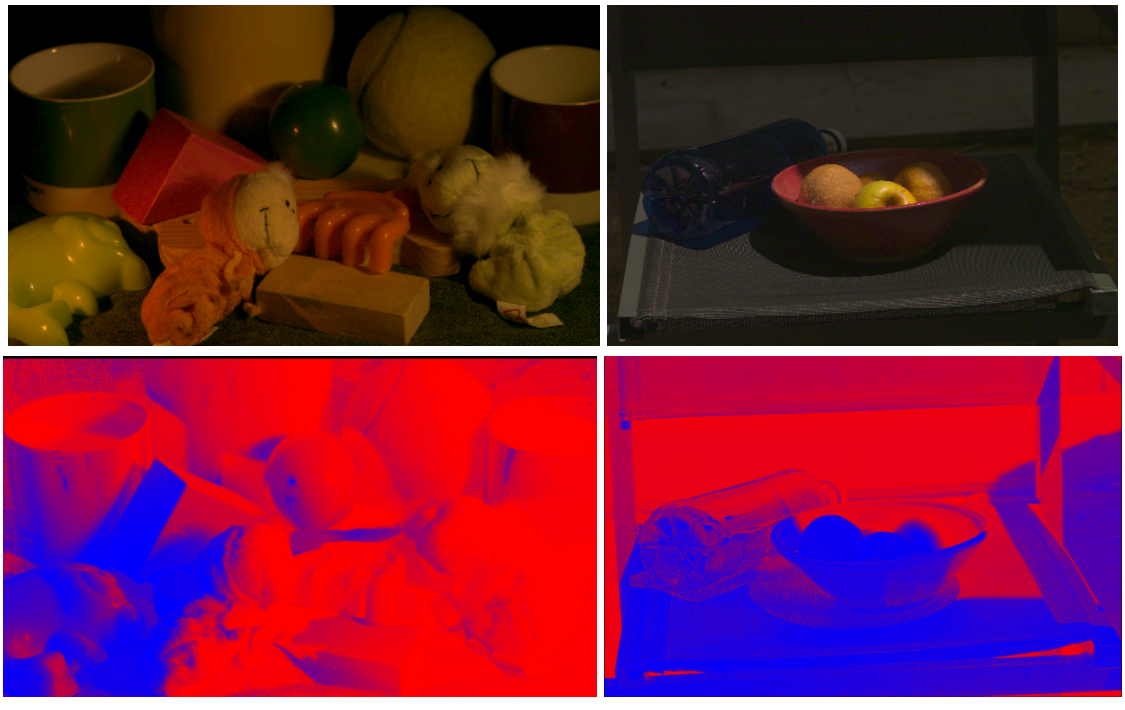


Figure 4.3: Example images from our laboratory and real-world dataset. The first row contains the original photographed images (transformed from their raw format to sRGB only for the purpose of visualization). In the bottom row, the relative influence of the two illuminants is shown, color-coded from blue to red.

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.* [46].

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.* [46], as it showed very competitive performance in a number of experiments.

Both, the method by Gijsenij *et al.* [46] and MIRF use as input illuminant estimates with small spatial support. Such illuminant estimates can be obtained from different es-

timators. 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 $n = 0$, $m = \infty$, $\sigma_{\mathrm{GW}} = 0$), 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.* [46]. 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[3], as in [81]. The settings for the CRF framework were as follows: the saturation threshold ϕ_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 methods by Gijsenij *et al.* [46] and our proposed algorithm "Multi-Illuminant Random

[3]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.

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.* [46]. 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 [46] was conducted on the non-gamma-corrected images[4]. In our implementation, we performed gamma correction on the input images, as it was also originally intended by [46]. 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 in our method versus the method of Gijsenij *et al.*, by providing them with 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.

[4]Without gamma correction, we obtain the same numbers as reported in [46].

	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	$10.1°$	$10.1°$	$8.5°$	$7.6°$	$\mathbf{7.1}°$(-16%)	$\mathbf{4.7}°$(-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.* [46]

	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 ϕ_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.* [46] (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-

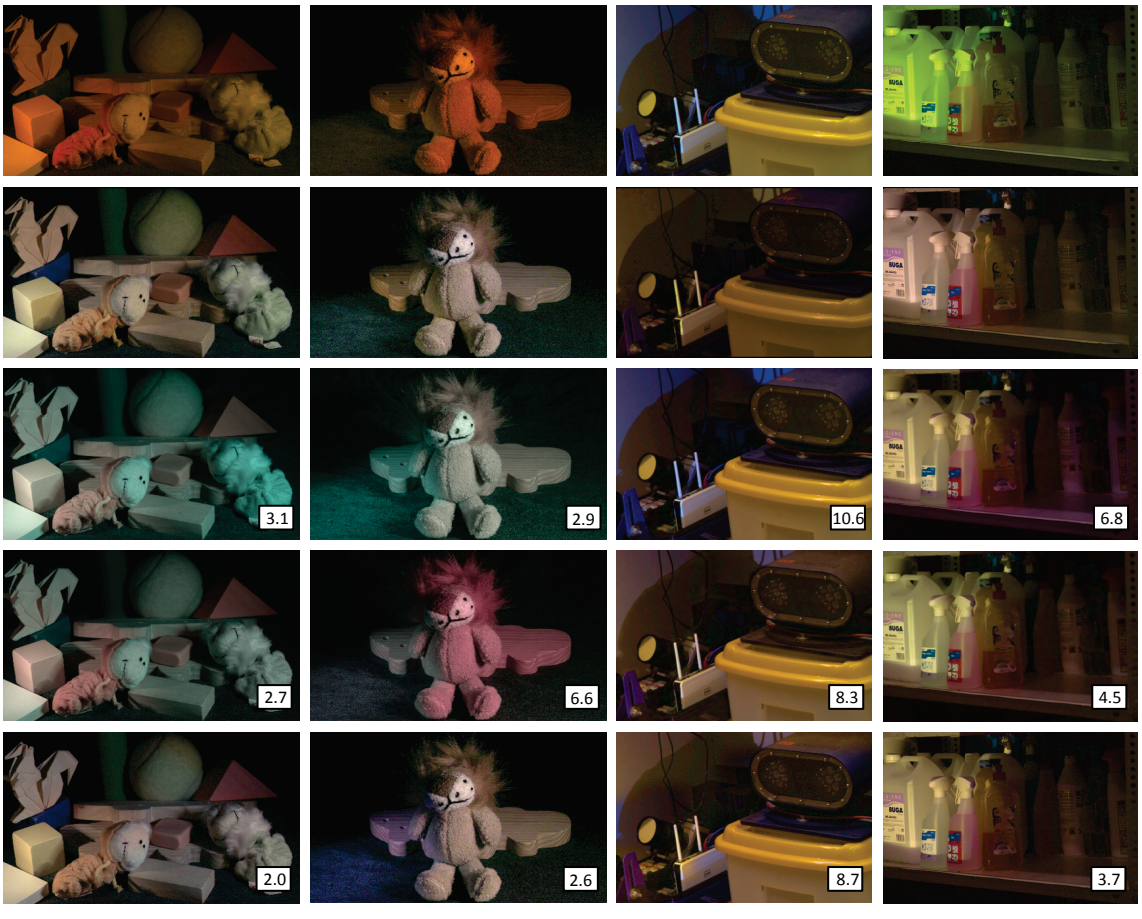


Figure 4.4: Example for the automatic white balance (WB). From top to bottom the rows present: original image from the camera, the WB images using the ground truth, global grey world, Gijsenij *et al.* [46], and MIRF. Note that here all the images are transformed from their original raw format to sRGB 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.* [46], 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 [86,98]. Recognition of common materials in the scene such as grass, stone, and faces could further improve multi-illuminant estimation.

Chapter 5

Dataset

Datasets play an important role in many computer vision and image processing applications. Illuminant and reflectance estimation algorithms need to be validated not only qualitatively but also quantitatively using a relatively large number of examples to prove their performance. In this thesis, we are interested in more complex reflectance and illumination situations. We found that ground-truth datasets are lacking for such scenarios. Therefore, we present our *Multi-illuminant Multi-object* (MIMO) Dataset for color constancy as well as our *synthetic scenes* dataset for intrinsic image decomposition.

5.1 Related Work

Figure 5.1 are examples for the existing image dataset for reflectance and illuminant estimation methods. In the following, we present a brief review of these datasets.

Color Constancy Datasets

One of the most popular datasets in the field of color constancy is the one proposed by Barnard *et al.* [6] which consists of 30 scenes captured under 11 different illuminants[1]. Each scene has been captured first by placing a white reference in the center (perpendicular to the illumination direction) to adjust the exposure and calculate the ground-truth. Then 50 successive images of the scene were captured and averaged to make the final image used in the dataset. This process is then repeated for each illuminant/scene.

To perform a more thorough study on statistics-based approaches for the problem of color constancy Ciurea and Funt [19] have proposed a large dataset of approximately 11,000 images of various indoor and outdoor scenes using a video camera. The ground-truth illuminant for each image is then calculated using a matte gray-ball connected to the camera which is visible in the image and assumed to contain the illumination of the

[1] According to the authors, 9 images were removed due to problems.

scene. Authors argue that even though calculating the illuminant's standard coordinate (e.g., CIE XYZ) would be more desirable than the camera's RGB sensor responses, due to the complicated task of camera calibration the camera RGB coordinate is reported for illuminant color. In this sense, Parraga *et al.* [75] present a calibration for their camera in order to accurately map the camera responses to the cone activation space of human vision (LMS) and published their dataset for color constancy [2]. To calculate the ground-truth a similar approach is used as the one by Ciurea and Funt.

Bianco *et al.* [12] showed that there is a significant amount of correlation between the images in the Ciurea and Funt data set as the images are taken from video clips. They applied a video-based analysis to select images with less correlation and reported a subset of 1,135 images. Gehler *et al.* [37] further argue that the low quality of the Ciurea and Funt dataset make it less suitable for training and testing color constancy methods. Therefore, they have proposed a high quality dataset consisting of 568 images from a variety of indoor and outdoor scenes. *Macbeth color checker* is used in order to improve the accuracy of the ground-truth illuminant calculation.

While these datasets provide an important resource for training and evaluation of single-illuminant color constancy methods, so far, the problem of the more realistic multi-illuminant scenes has not been widely targeted by the computer vision research community. One of the main reasons is the lack of a relatively large-scale dataset with accurate ground-truth data of local illumination. In order to provide accurate pixel-wise illuminant chromaticity map per image for mult-illuminant scenes, having just a color reference is not enough to produce the ground-truth. In this sense, Gijsenij *et al.* [46] placed gray balls in different parts of the scene and used manual annotation and segmentation to create illumination maps of their dataset. We on the other hand, use extra images captured from the scene in single-illuminant condition as an additional cue to accurately calculate the ground-truth. Chapter 4 contains details on ground-truth calculation for our multi-illuminant scene dataset. In Section 5.2, we explain further details on the acquisition of our dataset.

Intrinsic Image Datasets

Unlike color constancy, the intrinsic image estimation methods always required an accurate pixel-wise ground truth for both shape and reflectance features. A common ground truth calculation technique used in developing existing shape and intrinsic image estimation datasets is gray paint spray. 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 scene's *shading* image from which it is possible to obtain also the shape of the objects [79]

Often this task is done for scenes which consist of only one single object in order to simplify the problem. One popular example of datasets using this technique is the

[2] http://www.cvc.uab.es/color_calibration/Database.html

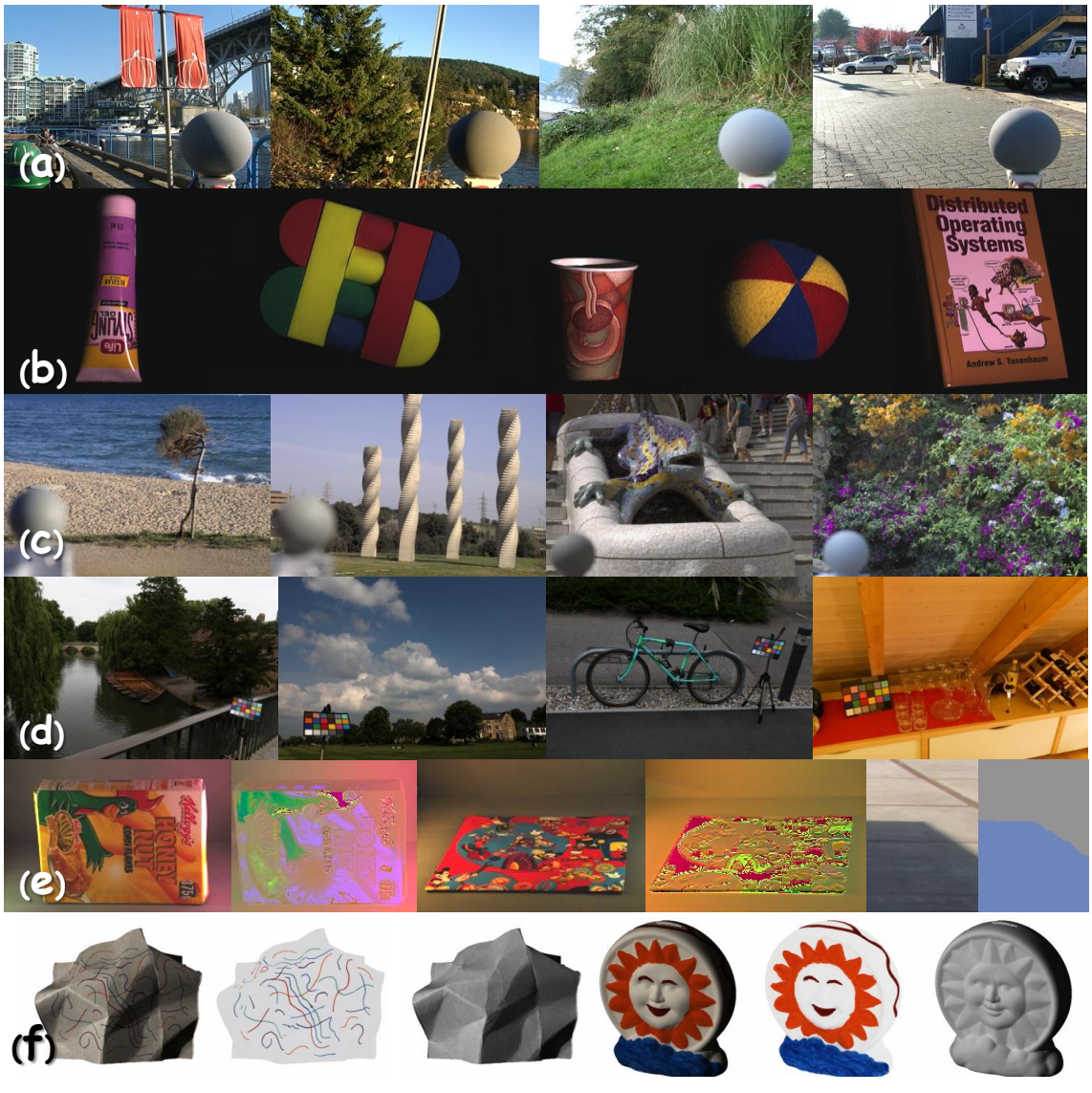


Figure 5.1: The rows (a) to (f) are examples of datasets: Ciurea and Funt [19], Barnard *et al.* [6], Parraga *et al.* [75], Gehler *et al.* [37], Gijsenij *et al.* [46], and Grosse *et al.* [49]. The last two rows include the ground-truth images provided by the authors of the respective datasets

widely used collection of MIT by Grosse *et al.* [49]. Also recently Bleier *et al.* [14] 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 the 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 do 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 acknowledged 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 which will be introduced in Section 5.3.

Dataset	Quality	Lighting	Scale	Content	Groundtruth
Barnard *et al.* [6]	High	Colored	30×11	Lab	GrayBall
Ciurea & Funt [19]	Medium	Ambient	∼11,000	Real	GrayBall
Parraga *et al.* [75]	High	Ambient	210	Real	GrayBall
Gehler *et al.* [37]	High	Ambient	568	Real	Macbeth
Gijsenij *et al.* [46]	High	Multi-color	59+9	Lab/Real	GrayBall
MIMO (§Chapter 4)	High	Multi-color	58+20	Lab/Real	Macbeth
Grosse *et al.* [49]	High	White	20×10	Lab	GrayPaint
Our intrinsic DS	High	Multi-color	17×4	Synthetic	Auto

Table 5.1: Comparing popular color constancy and intrinsic image datasets with our datasets.

5.2 Multi-illuminant Multi-object (MIMO) Dataset

In Chapter 4, we introduced our *Multi-illuminant Multi-object* (MIMO) Dataset developed for local illuminant estimation and color constancy. In this section we present more details on the acquisition of the images and their corresponding ground truth.

5.2.1 Lab set

Figure 5.2 presents example scenes captured for this dataset. To capture the scenes we used a booth and placed the objects, lights and the camera inside. We covered the interior surface of the booth completely with black matte filter-like fabric which damps the lights in order to reduce the effect of ambient light and to avoid unwanted inter-reflections. We used two $100W$ lamps placed on the front left and right side of the scene to illuminate the objects (lighting is directional from up and front toward the center of the scene). Here on, we refer to the chromaticity of the lamps as white (in practice they are slightly reddish).

To achieve colored lights, we used two *Cokin* glass filters, namely number 30 and number 22, for red and blue light respectively.

Each scene in the set is captured under 14 different lighting combinations. That is 6 single illuminant (white, red, and blue from either left or right), and 8 multi-illuminant scenes (white with white, red, and blue; red with white and blue; and blue with red, white and blue). During the capturing of the scenes we kept the position of the lights and the camera completely fixed to facilitate the matching of the images for the ground-truth computation step (an external switch is used for the camera to avoid any undesired movement).

The scenes presented in this set typically contain both specular and diffuse objects made of different materials (plastic and porcelain or textile and wood respectively). We have included some scenes of almost identical porcelain coffee mugs with different colors placed in exact same positions for future use for evaluation of recoloring methods. We calibrate the lights for ground-truth computation using an empty scene with only a gray-level reference board placed in the center photographed in all the 14 different lightings. There are many clear cases of colored shadows present in this set which provides the desired complexity of the illumination conditions in order to better evaluate the estimation methods.

In practice we only used 6 different multi-illuminant lighting conditions in the evaluation of our method in Chapter 4 to assure that the chromaticities of the lights for each scene are at least 5 degrees apart. The single-illuminant images are captured to provide additional cues necessary for the ground-truth computation step (Section 7.1).

5.2.2 Real-world set

Capturing the real-world scenes is indeed harder for various reasons such as inevitable motion or changes in outdoor scenarios or having much less control on the illumination and light sources in the scene. For example, in case of the images with natural sunlight, we have little control over the main illuminant, and we have to assume that the acquisition process is fast enough to avoid any changes on the direction of the sun (some scenes needed to be captured again due to variations in the clouds). Also it is almost impossible to capture a real-world outdoor scene and be able to control or perfectly estimate the ambient or secondary lights present in the scene at each position. For the same reasons, using the gray paint spray technique for ground-truth computation can not be accurate.

For these reasons we captured a sum of 28 scenes, while in the end we could only use 20 scenes for our real-world set after removing the images which were mis-aligned, moved, etc. The set overall consists of 5 indoor scenes lit by indoor lamps and a color projector, 3 outdoor scenes lit by yellowish sunlight and bluish skylight (ambient), and 12 indoor scenes lit by the sunlight from windows in the rooms and a projector (6 scenes) or a lamp (6 scenes). For each scene we captured two images: one for single illuminant and one for multi-illuminant, and also two images from exactly the same illumination

Figure 5.2: Here we demonstrate the process of data acquisition for our multi-illuminant scene images. The two images on the bottom right are examples of a complex scene (e.g, specular highlights and shadows) captured under distinct illuminants (note that these images are enhanced for visualization).

Figure 5.3: Some examples from the real-world scenes for the multi-illuminant multi-object dataset. The scenes demonstrate complex objects and illumination conditions (e.g., specularities and shadows) captured under distinct illuminants. Note that here the images are transformed from their original raw format to sRGB for visualization.

conditions but adding a Macbeth color checker to obtain illuminant chromaticities. In practice, only one image per scene, that is the multi-illuminant scene without the color reference, is used for our evaluations in Chapter 4. Figure 5.3 shows some examples of the dataset.

5.2.3 Technical Challenges

In this section we have demonstrated an example of data acquisition to show the challenges of capturing an entire dataset. As much as the need for large datasets with variety in the context has been demonstrated in various areas of computer vision, so far, there are not many reliable datasets are available. The reasons for that can be found in numerous drawbacks of traditional data acquisition using cameras:

- The over all task is quite slow and cumbersome which limits the number of images in the set. And the need of proper equipments makes data acquisition financially costly which could be a problem for many researchers.
- The procedure for capturing the data and calculating the ground truth is often so complicated that it limits the characteristics of the scenarios.
- Often there is not much variety in the content of the scenes. This could be a huge disadvantage in the case of parameter training and result in over-fitting of the model.
- It is often hard to repeat or re-use the data with modifications (e.g., to obtain the scene under different illuminants or different view-points).

Therefore, in the next section, we focus on the use of photo-realistic synthetic data to make up for these shortcomings.

5.3 Synthetic Intrinsic Image dataset [3]

Obtaining a precise ground truth for real complex scenes, such as a landscape, would be hardly possible using the procedure described in [49]. Recently, the use of synthetic data to train and test complex computer vision tasks has attracted growing attention due to the increased accuracy with which 3D renderers visualize the world. In addition synthetic data allows for easy access to the ground truth, making it possible to prevent the costly manual labeling process. Marin *et al.* [69] and Vazquez *et al.* [100] show that a pedestrian detector trained from virtual scenarios can obtain competitive results on real-world data. Liebelt and Schmid [64] use synthetic data to improve multi-view object class detection. Finally, Rodriguez *et al.* [83] generate synthetic license plates to train recognition system.

[3] The materials in this section are used for a conference submission: "Intrinsic Image Evaluation on Synthetic Complex Scenes" by Shida Beigpour, Marc Serra, Joost van de Weijer, Robert Benavente, Maria Vanrell, Olivier Penacchio, and Dimitris Samaras, that is currently under revision

We propose a synthetic dataset for intrinsic image evaluation which includes not only single object scenes with white illumination, but also multi- object scenes and multiple non-white illuminants with complex surrounding leading to interreflections. Multispectral sensors have been simulated in this work in order to emulate a realistic visualization as described in [20, 22]. The objective of this new ground truth collection is to overcome the shortcomings of the existing dataset in intrinsic image evaluation and show an easy way to build ground truth for reflectance, shading and illumination from synthetic data which allows the collection of a larger and more complex set of scenes. This dataset will be publicly available to further motivate research into more complex reflectance models.

To validate our dataset, we evaluate and compare three existing methods [9,38,87] and show that similar results are obtained on synthetic data as for real-world data in similar settings (one illuminant, one object). Our experiments also show that current intrinsic image methods are limited in the presence of multiple illuminants and interreflections usually found in complex scenes.

5.3.1 Motivation

Intrinsic image algorithms and datasets can be distinguished by their assumptions on the underlying reflectance models. Consider the reflection model [88] which models the color observation f^c with $c \in \{R, G, B\}$ as:

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

where the integral is over all wavelengths λ of the visible spectrum ω. The material reflectance is given by $s(\lambda, \mathbf{x})$, $e(\lambda, \mathbf{x})$ is the spectrum of the illuminant, and m is a scalar 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 [49] 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 [8] [38] [87]. Recently, Barron and Malik [9] 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 [9].

Here, we go one step further and create a synthetic dataset by using rendering techniques from the computer graphics field. This allows us to remove the restriction other datasets 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 [49] applies polarizing filters to provide both images with and without specular reflection.

5.3.2 3D Object Modeling and Photo-realism

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. Rendering programs are moving toward achieving more realistic results and accuracy using physics-based models in optics. There are currently various applications available which embed the known illumination and reflectance models [76].

In the current work, we have used Blender to model the 3D scenes. Yafaray is used as a rendering software for its photo-realism and physically plausible results. Both of these applications are free and open source which makes them suitable for scientific use.

5.3.3 Global Lighting for 3D scene rendering

In order to obtain more photo-realistic lighting results for 3D scene rendering, a group of rendering algorithms have been developed which are referred to as *global illumination*. These methods, in addition to taking into account the light which reaches 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 latter is also known as *indirect lighting*. This indirect lighting is what causes the reflections, refractions, shadows, ambient lighting, and inter-reflections. Figure 5.4 presents the competitive quality and photo-realism for synthetic scenes rendered using global illumination.

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* [51] 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 5.5 shows the importance of indirect lighting 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*.

Figure 5.4: 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). Note that the gamma is added to the synthetic images for visualization.

Figure 5.5: Comparing different rendering methods: *direct lighting* (left) and *photon mapping* (right) on an example scene from our dataset. Note that the gamma is added for visualization.

Photon Mapping

Photon mapping consists of two steps: *photon tracing* and *rendering*. Below we briefly review the steps of this two pass algorithm.

Photon tracing: This method can handle different types of light sources depending on the necessity of the scene. But to reduce computational time, *photon maps* are generated which optimize the emission by directing photons to the important objects in the scene by estimating at each point of the light source, if geometry of importance is in that direc-

tion. Then *Russian Roulette* probability is used to decide whether photons are reflected, refracted or absorbed. Finally, as suggested by Jensen, the photon map is stored using a kd-tree structure as it is ideal for handling non-uniform distributions of photons.

Rendering: At this step, the photon map created in the previous step is used in order to estimate the radiance of each pixel in the final image. In this scene, each pixel is *ray-traced* until the nearest intersection surface is found. Then the surface radiance leaving the intersection point in the direction of the ray that struck it is calculated using the rendering equation. To improve the efficiency and to decrease the costs of the rendering process, the equation is decomposed into: direct illuminant, specular reflection, diffuse indirect illumination (inter-reflections), caustics. The first two parts are often handled using *ray tracing* since it is quite accurate for direct lighting and reflections. The caustics are calculated using the caustics photon map[4]. And finally the diffuse indirect illumination is calculated using global photon map. To do so, first using the nearest neighbor search on photon map, N nearest photons are gathered. Then for each photon, the amount of flux that photon represents is divided by the area of the sphere which contains these N photons and multiplied by the BRDF of that photon. The sum of those results for each photon represents the total surface radiance returned by the surface intersection in the direction of the ray that struck it.

To demonstrate the importance of indirect lighting in scenes using quantitative results we compare the final renderings of our complex scenes to the renderings which only consider direct illumination (rendering programs allow for this separation). We compare the total energy in both with the ratio $r = \sum_{\mathbf{x}} \|\mathbf{f}^1(\mathbf{x})\| \Big/ \sum_{\mathbf{x}} \|\mathbf{f}^\infty(\mathbf{x})\|$ where $\mathbf{f}^\infty$ is the final rendering and $\mathbf{f}^1$ is the single bounce rendering. For the nine complex scenes we found that the average $r = 0.86$, showing that a significant amount of the lightning in the scene is coming from interreflections between the objects.

5.3.4 Analysis of 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 typically done in graphics. To prevent propagating the full multispectral data, which is computationally very expensive, rendering engines approximate Eq. 5.1 with

$$\hat{f}^c = \int_\omega s(\lambda)\rho^c(\lambda)\,d\lambda \int_\omega e(\lambda)\rho^c(\lambda)\,d\lambda. \tag{5.2}$$

[4]Here we do not focus on the caustic effect

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

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

where we use bold to denote vectors, and $\circ$ is the Hadamard product, and we replaced $\mathbf{s} = \int_{\omega} s(\lambda)\rho^c(\lambda)\,d\lambda$ and $\mathbf{e} = \int_{\omega} e(\lambda)\rho^c(\lambda)\,d\lambda$. 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:

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

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 $c \in \{R, G, B\}$, however it is known that the rendering accuracy increases when considering more sensors [22] [20].

To test the accuracy of $\hat{f}^c$ we perform a statistical analysis. We use the 1269 Munsell color patches [1] and we compute both f^c and $\hat{f}^c$. For sensors ρ^c we use Gaussian shaped sensors which are equally spaced over the visible spectrum. We compare the reconstruction error

$$\varepsilon = \left\|\mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\mathbf{x})\right\| \Big/ \|\mathbf{f}(\mathbf{x})\|$$

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.

	One bounce		**Two bounces**	
sensors	Mean (%)	Max (%)	Mean (%)	Max (%)
3	0.58	2.88	1.38	23.84
6	0.19	1.25	0.55	9.06
9	0.12	0.86	0.34	3.77

Table 5.2: Reconstruction error for single and two bounce reflection for 3, 6, and 9 sensors.

In Table 5.2 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 23.84%. 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 9.06%. This can be conveniently achieved by running existing rendering software (built for 3 channel propagation) twice for three channels [22] [20]. The final

6-D result image is projected back to a RGB image using linear regression. In the only available intrinsic image data set for multi-illuminants [9], 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 relighting of real-world scenes would introduce significant errors.

5.3.5 Proposed dataset

Our dataset consists of two set of images: single objects and complex scenes. In the first set, the aim is to simulate the work on MIT dataset The second set is to our knowledge the first set of complex scenes for intrinsic image estimation which has an accurate ground truth, not only for the typical reflectance and shading decomposition, but also for the illuminant estimation. There are 8 objects in the first set which vary in complexity for their shape and color distribution. The complex scenes on the other hand consist of various complex objects (e.g, furniture) which result in diffuse inter-reflections and complex shadows.

Over all, there are 9 images in the second set. All the colors of the objects present in the scenes are taken from the Munsell colors since the multi-spectral reflectance values for them is recorded. Figure 5.6 shows examples of ground-truth we provide with the dataset. All the single-object and complex scenes in our dataset are rendered under 4 different illumination conditions (i.e., white light, colored light, and 2 cases of multiple illuminants with distinct colors). This leads to a total of 32 images in the single object set and 36 in the complex scene set. The illuminants are randomly chosen from a list of Planckian and non-planckian lights from the Barnard dataset [7].

5.3.6 Experiments

Here we show results for three state-of-the-art methods to intrinsic image estimation on the data set [9, 38, 87]. For the experiment, the available codes provided by the authors of the tested methods have been used, and the parameters for each method have been set to the values provided in their works where the methods were mostly trained and tested on the MIT dataset. Moreover image size was reduced by a factor of two for the method of Gehler et al. [38] due to the high computational cost of the method.

For each of the subsets in our dataset, namely single objects and complex scenes, we have analyzed the three methods on three illumination conditions: white light(WL), one non-white illuminant(1L), and two non-white illuminants(2L). Mean results on each illumination condition have been computed. Figure 5.7, presents an example of the reflectance estimation results by each method for a scene under various lighting conditions.

Errors have been evaluated using the local mean squared error (LMSE) and considering the three RGB channels of the color image [49]. As reflectance images can be recovered only up to a scale factor, we have multiplied the estimated reflectance images

Figure 5.6: Some examples for the dataset and its ground-truth. From left to right: the rendered scene, reflectance component, and illumination. Note that the chromaticity of the illumination image presents the illuminant color at each pixel and its brightness is the shading image. The images are enhanced for visualization.

by an α factor which has been fitted for each local patch to minimize the MSE. Table 5.3 and Table 5.4 summarize the results obtained for reflectance and shading respectively. As expected, the error for all the three methods increases when the illuminant is not white. The shading evaluation is relatively invariant to illuminant changes because it discards

color information. The lower errors on the complex scenes are caused by large uniform colored objects which result in low LMSE errors. Overall the method of Serra et al. [87] obtained the best results on our data set. The expected robustness of the method of Barron and Malik [9] is not reported on this data set. This could be due to the fact that the assumptions of direct lighting is broken in our scenes.

Methods	**Single Objects**			**Complex scenes**		
	WL	1L	2L	WL	1L	2L
Barron & Malik	0.098	0.115	0.120	0.029	0.068	0.050
Gehler *et al.*	0.111	0.135	0.138	0.028	0.078	0.052
Serra *et al.*	0.073	0.079	0.085	0.032	0.045	0.038

Table 5.3: LMSE of the reflectance estimation results by three intrinsic image methods on our dataset. For both single objects and complex scenes, results for white illumination (WL), one illuminant (1L), and two illuminants (2L) are averaged.

Methods	**Single Objects**			**Complex scenes**		
	WL	1L	2L	WL	1L	2L
Barron & Malik	0.063	0.065	0.079	0.019	0.023	0.024
Gehler *et al.*	0.081	0.081	0.090	0.018	0.021	0.020
Serra *et al.*	0.028	0.027	0.032	0.011	0.011	0.011

Table 5.4: LMSE of the shading estimation results by three intrinsic image methods on our dataset. For both single objects and complex scenes, results for white illumination (WL), one illuminant (1L), and two illuminants (2L) are averaged.

5.4 Conclusion

Datasets have played a crucial role as a catalyst in many computer vision fields. To motivate more research into computer vision in more complex lighting situations, such as multi-illuminant scenes and scenes with inter-reflections, new datasets are required. In this chapter, we provided two such data sets.

For research in multi-illuminant estimation, we have proposed a dataset of photographed images consisting of a set of toy scenes under control laboratory conditions and a set of real-world indoor and outdoor images. These scenes present highly complex reflectance, highlights, and lighting conditions such as shadows, inter-reflections and non-uniform lighting. Along with this dataset, we provided a precise pixel-wise illumination ground-truth data for each image.

The second dataset is provided for intrinsic image estimation in complex lighting situations. Collecting intrinsic image ground-truth data is a laborious task. The assumptions on which the ground-truth procedures are based limit their application to simple scenes with a single object taken in the absence of indirect lighting and inter-reflections. We investigated synthetic data for intrinsic image research since the extraction of ground truth

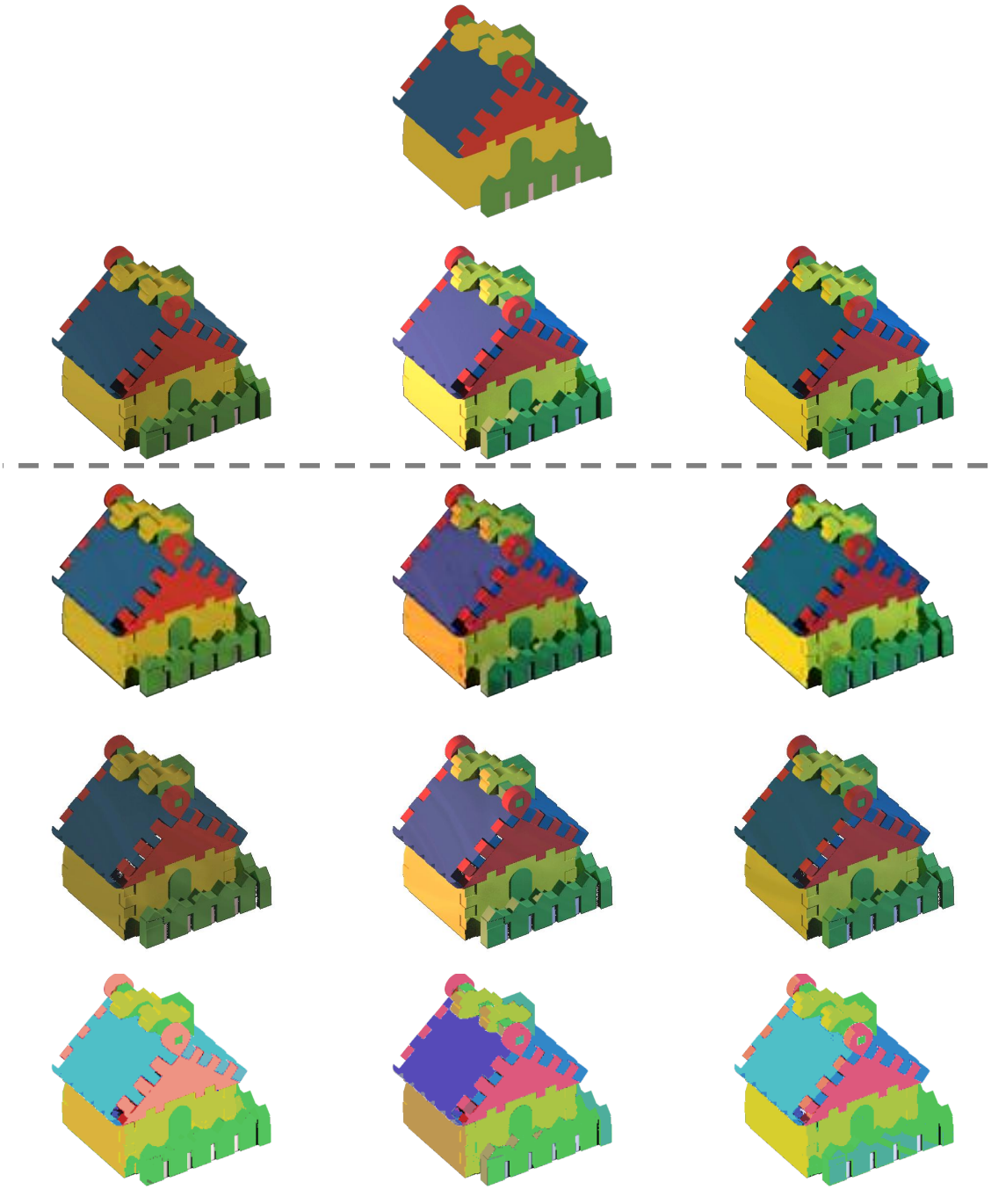


Figure 5.7: An example of reflectance decomposition results for a scene from our dataset. The image on the top row is the ground-truth reflectance. The second row is the scene rendered under three different lighting conditions, namely: white light (WL), single colored illuminant (1L), and two distinct illuminants (2L), respectively from left to right. The rows three to five are the reflectance estimation results obtained by methods of Gehler *et al.* [38], Barron and Malik [9], and Serra *et al.* [87] respectively for each of the rendered images as input. Note that the images are enhanced for visualization.

is straightforward, and it allows for scenes in more realistic situations (e.g, multiple illuminants and inter-reflections).

Chapter 6

Conclusions and Future Directions

This work aims at improving the modeling of the illumination and its interaction with object surfaces. 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 Chapter 3, 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 improved 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.

We have created two datasets which are explained in Chapter 5. Our multi-illuminant

scene dataset captured from both laboratory controlled scenes and real-world indoor and outdoor scenes is more general and complex than existing dataset for multi-illuminant color constancy. This chapter further presents our synthetic scene dataset for intrinsic image decomposition using physically sound 3D modelling and rendering programs. The main advantage of synthetic data is that the extraction of the ground truth is straightforward. This allows us to evaluate intrinsic image algorithms in the presence of complex lighting situations such as multiple illuminants and interreflections, which were previously absent in intrinsic image data sets.

6.2 Future Directions

The object reflectance model in Chapter 3 can be further extended to account for ambient lighting. Also an important improvement over the current framework can be achieved by automatic segmentation of the object from the input image to minimize the manual user interaction. Lastly the Algorithm 1 can potentially be extended to extract any desired number of illuminants using the object pixels' color distribution given that the areas affected by these illuminants are big enough and that the noise ratio allows for making reliable clustering over the object color distribution.

Moreover, our multi-illuminant estimation method of Chapter 4 can strongly benefit from incorporating top-down cues for illuminant estimation such as object recognition and sparseness of the reflectance. Also Gamut mapping approach can easily be embedded into the current framework as a further constraint on the illuminants. And combining various statistical methods can improve the bottom-up local estimations. In addition, the calculation of *pairwise potentials* can be improved using heuristics such as Retinex.

Furthermore, we acknowledge the necessity of expanding our datasets which are introduced in Chapter 4 and Chapter 5 to include more variety of scenes and illumination conditions. The ground-truth calculation procedure for our MIMO dataset requires further improvements to better capture the effect of specularities and inter-reflections. Also our synthetic dataset for intrinsic image estimation can be extended to include specular surfaces.

Finally, 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 removing the effect of complex illumination facilitates other Computer Vision tasks (e.g., segmentation, and object classification). Moreover, the illumination and reflectance modeling could be applied to *image forensics* approaches.

Chapter 7

Appendix

7.1 Further Results

Here we show further examples of results achieved using the MIDR method of Chapter 3.

Figure 7.1: From left to right: Original image, recolored object, changing the primary illuminant, removing and recoloring the secondary illuminant.

Figure 7.3 is a visualization of the two-illuminant MIDR decomposition, recoloring and photo-fusion respectively.

$$\mathbf{f} = m_b^1 \mathbf{c}\, \mathbf{L^1} + m_s^1 \mathbf{l^1} + m_b^2 \mathbf{c}\, \mathbf{L^2}. \tag{7.1}$$

Figure 7.2: Examples of object recoloring while preserving the *bluish* ambient light and shadows.

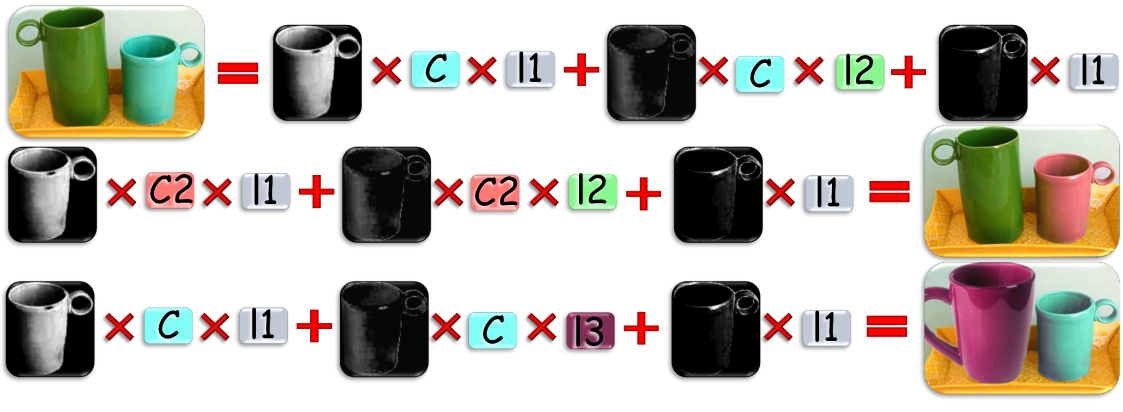


Figure 7.3: The first row is an example of image decomposition using MIDR. In the middle, the object has been recolored automatically simply by changing the object color RGB value while the inter-reflections have been preserved. The bottom row demonstrates an interesting further application of MIDR in improving the results of *photo-fusion* in which after replacing an object in the scene with another, the inter-reflections can be modified to generate more realistic results (Note that the general problem of photo-fusion can be more complex than this).

7.2 The Algorithm Convergence to Grey-world

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

$$E(\mathbf{x}|\mathcal{F}) = \sum_{i \in \mathcal{V}} \phi(\mathbf{x}|\mathcal{F}_i) = \sum_i \left\| \sum_{j \in p_i} \mathbf{f}^j \right\| \left(1 - \mathbf{i}_i{}^{\mathrm{T}} \mathbf{x}\right) \tag{7.2}$$

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

$$E(\mathbf{x}|\mathcal{F}) = \sum_i \left\| \sum_{j \in p_i} \mathbf{f}^j \right\| - \sum_i \left(\sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} \ , \tag{7.3}$$

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

$$\underset{\mathbf{x}}{\mathrm{argmax}} \sum_i \left(\sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} = \underset{\mathbf{x}}{\mathrm{argmax}} \left(\sum_i \sum_{j \in p_i} \mathbf{f}^j \right)^{\mathrm{T}} \mathbf{x} \ , \tag{7.4}$$

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

$$\mathbf{x}^* \propto \sum_i \sum_{j \in p_i} \mathbf{f}^j \tag{7.5}$$

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 be 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.

7.3 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

$$\mathbf{i}_{\mathbf{ab}} = w \cdot \mathbf{a} + (1 - w) \cdot \mathbf{b} \ , \tag{7.6}$$

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 $\hat{f}_{\mathbf{a},g}$, $\hat{f}_{\mathbf{b},g}$ 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 $\hat{f}_{\mathbf{a},g}$ with respect to $\hat{f}_{\mathbf{b},g}$,

$$w = \frac{\hat{f}_{\mathbf{a},g}}{\hat{f}_{\mathbf{a},g} + \hat{f}_{\mathbf{b},g}} \quad . \tag{7.7}$$

Assuming Lambertian reflectance and sharpened sensors, $\hat{f}_{\mathbf{a},g} = \omega_{\mathbf{a}} k_{\mathbf{a}} \rho_g$, where $\omega_{\mathbf{a}}$ 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. 7.7, the ratio of the pixel under both illuminant corresponds to the ratio of their scaling factors $\omega_{\mathbf{a}}$ and $k_{\mathbf{a}}$,

$$w = \frac{\omega_{\mathbf{a}} k_{\mathbf{a}} \rho_g}{(\omega_{\mathbf{b}} k_{\mathbf{b}} + \omega_{\mathbf{a}} k_{\mathbf{a}}) \rho_g} = \frac{\omega_{\mathbf{a}} k_{\mathbf{a}}}{\omega_{\mathbf{b}} k_{\mathbf{b}} + \omega_{\mathbf{a}} k_{\mathbf{a}}} \quad , \tag{7.8}$$

as albedo and neutral illuminant are identical in $\hat{f}_{\mathbf{a},g}$ and $\hat{f}_{\mathbf{b},g}$. 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.

$$w = \begin{cases} 1 & \text{if } \hat{f}_{\mathbf{a},g} / \hat{f}_{\mathbf{b},g} > t_{\mathrm{B}} \\ 0 & \text{if } \hat{f}_{\mathbf{b},g} / \hat{f}_{\mathbf{a},g} > t_{\mathrm{B}} \\ \tilde{w} \hat{f}_{\mathbf{a},g} / \hat{f}_{\mathbf{b},g} & \text{otherwise} \end{cases} \quad , \tag{7.9}$$

where $\tilde{w}$ 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 [88]). 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.

Publications

International Journal

- Shida Beigpour and Christian Riess and Joost van de Weijer, and Elli Angelopoulou. "Multi-illuminant estimation using Conditional Random Field". In *IEEE Transactions on Image Processing (TIP)*, Submitted.

International Conferences

- Shida Beigpour, Marc Serra, Joost van de Weijer, Robert Benavente, Maria Vanrell, Olivier Penacchio, and Dimitris Samaras. "Intrinsic Image Evaluation On Synthetic Complex Scenes". In *Proceedings of 20th International Conference on Image Processing (ICIP'2013)*, Melbourne, Australia, September, 2013 - Submitted.

- Fahad Khan and Shida Beigpour and Joost van de Weijer, and Michael Felsberg. "Painting-91: A Large Scale Database For Digital Painting Processing". In *Proceedings of 20th International Conference on Image Processing (ICIP'2013)*, Melbourne, Australia, September, 2013 - Submitted.

- Shida Beigpour and Joost van de Weijer. "Object Recoloring based on Intrinsic Image Estimation". In *Proceedings of 13th IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, November, 2011.

- Joost van de Weijer and Shida Beigpour. "The Dichromatic Reflection Model: Future Research Directions and Applications". In *Proceedings of International Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP11)*, Portugal, 2011.

- Shida Beigpour and Joost van de Weijer. "Photo-Realistic Color Alteration For Architecture And Design". In *Proceedings of Colour Research for European Advanced Technology Employment (CREATE) Conference*, Norway, Jun, 2010.

International Workshops

- Michael Bleier, Christian Riess, Shida Beigpour, Eva Eibenberger, Elli Angelopoulou, Tobias Troeger, and Andre Kaup. "Color Constancy and Non-Uniform Illumination: Can Existing Algorithms Work?". In *International Workshop on Color and Photometry in Computer Vision (CPCV'2011) in conjunction with ICCV 2011*, Barcelona, Spain, November, 2011.

- Shida Beigpour and Joost van de Weijer. "Object Color Alteration". In *4th CVC workshop on the progress of research and development (CVCR& D)*, Bellaterra, Spain, October, 2009.

Bibliography

[1] http://www.uef.fi/spectral/spectral-database. Last accessed on January 31, 2013.

[2] X. An and F. Pellacini. User-controllable color transfer. *Eurographics*, 2010.

[3] K. 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.

[4] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms – part 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. *Image Processing, IEEE Transactions on*, 11(9):985 – 996, sep 2002.

[7] K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for color research. *Color Research & Application*, 27(3):147–151, 2002.

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

[9] J.T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *European Conference on Computer Vision (ECCV)*, pages 55–70, 2012.

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

[11] S. Beigpour and J. van de Weijer. Object recoloring based on intrinsic image estimation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 327–334, 2011.

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

[13] S. Bianco and R. Schettini. Color constancy using faces. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 65–72. IEEE, 2012.

[14] M. Bleier, C. Riess, S. 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.

[15] 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.

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

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

[20] B. Darling, J.A. Fewerda, R.S. Berns, and T. Chen. Real-time multispectral rendering with complex illumination. In *19th Color and Imaging Conference*, pages 345–351, 2010.

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

[22] M.S. Drew and G.D. Finlayson. Multispectral processing without spectra. *Journal of the Optical Society of America A*, 20(7):1181–1193, 2003.

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

[24] M. Ebner. Color constancy using local color shifts. In Tomás Pajdla and Jiri Matas, editors, *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.

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

[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. *The Journal of the Optical Society of America A.*, 11(5):1553–1563, May 1994.

[30] G.D. Finlayson, S.D. Hordley, and P.M. HubeL. Color by correlation: a simple, unifying framework for color constancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1209 –1221, nov 2001.

[31] 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.

[32] 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.

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

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

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

[36] D. Freedman and P. Kisilev. Object-to-object color transfer: Optimal flows and smsp transformations. In *Computer Vision and Pattern Recognition, CVPR*, 2010.

[37] 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.

[38] P.V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf. Recovering intrinsic images with a global sparsity prior on reflectance. In *Advances in Neural Information Processing Systems (NIPS)*, pages 765–773, 2011.

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

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

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

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

[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. Computational color constancy: Survey and experiments. *Image Processing, IEEE Transactions on*, 20(9):2475 –2489, sept. 2011.

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

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

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

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

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

[50] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand. Light mixture estimation for spatially varying white balance. In *ACM SIGGRAPH 2008 papers*, pages 70:1–70:7, New York, NY, USA, 2008. ACM.

[51] H.W. Jensen. *Realistic image synthesis using photon mapping*. A.K. Peters, Ltd., Natick, MA, USA, 2001.

[52] 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.

[53] G.J. Klinker and S.A. Shafer. A physical approach to color image understanding. *Int. Journal of Computer Vision*, 4:7–38, 1990.

[54] P. Kohli, L. Ladick, and P. HS Torr. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009.

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

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

[57] P. Kubelka. New contributions to the optics of intensely light-scattering materials. part i. *Journal of the Optical Society of America*, 38(5):448–448, 1948.

[58] E. H. Land, John, and J. Mccann. Lightness and the Retinex Theory. *Journal of the Optical Society of America*, pages 1–11, 1971.

[59] E.H. Land. The retinex theory of colour vision. *Scientific American*, 237(6):108–129, 1977.

[60] 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.

[61] 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.

[62] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *PAMI*, 30, 2008.

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

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

[65] 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.

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

[67] D. G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[68] S. P. Mallick, T. Zickler, P. N. Belhumeur, and D. J. Kriegman. Specularity removal in images and videos: A pde approach. pages 550–563, 2006.

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

[70] B. Maxwell, R. Friedhoff, and C. Smith. A Bi-Illuminant Dichromatic Reflection Model for Understanding Images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[71] B.A. Maxwell and S.A. Shafer. Physics-based segmentation of complex objects using multiple hypothesis of image formation. *Computer Vision and Image Understanding*, 65:265–295, 1997.

[72] B.A. Maxwell and S.A. Shafer. Segmentation and interpretation of multicolored objects with highlights. *Computer Vision and Image Understanding*, 77:1–24, 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] I. Omer and M. 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] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[78] F. Pitié, A. C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107, 2007.

[79] E. Prados and O. Faugeras. Shape from shading. *Handbook of mathematical models in computer vision*, pages 375–388, 2006.

[80] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21, 2001.

[81] 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.

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

[83] 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.

[84] C. Rosenberg, T. Minka, and A. Ladsariya. Bayesian color constancy with non-gaussian models. *Advances in Neural Information Processing Systems*, 16, 2003.

[85] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM SIGGRAPH*, 23, 2004.

[86] 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.

[87] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell. Names and shades of color for intrinsic image estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 278–285, 2012.

[88] S.A. Shafer. Using color to separate reflection components. *Journal of Color Research and Application*, 10(4):210–218, 1985.

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

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

[91] Robby T. Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. In *PAMI*, pages 870–877, 2003.

[92] 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.

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

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

[95] 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.

[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, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. In *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pages 1–8. IEEE, 2007.

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

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

[101] E. Vazquez, R. Baldrich, J. van de Weijer, and M. Vanrell. Describing reflectances for color segmentation robust to shadows, highlights, and textures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):917–930, 2011.

[102] J. Vazquez-Corral, M. Vanrell, R. Baldrich, and F. Tous. Color constancy by category correlation. *Image Processing, IEEE Transactions on*, 21(4):1997–2007, 2012.

[103] J. von Kries. Influence of adaptation on the effects produced by luminous stimuli. *Sources of Color Vision, D. L. MacAdam, ed.*, pages 109–119, 1970.

[104] Y. Weiss. Deriving intrinsic images from image sequences. pages 68–75, 2001.

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