UNIVERSITY OF CALIFORNIA, SAN DIEGO

The structure and analysis of a scene: A photometric approach

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Signal & Image Processing)

by

Satya Prakash Mallick

Committee in charge:

Professor David Kriegman, Chair
Professor Nuno Vasconcelos, Co-Chair
Professor Serge Belongie
Professor Samuel Buss
Professor Pamela Cosman
Professor Truong Nguyen

2006

The dissertation of Satya Prakash Mallick is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2006

To my parents, for their love, guidance, encouragement, and blessings.

To my brother, for being a friend and a teacher.

To my wife for her love, support, and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

There are several people who made my experience as a graduate student wonderful. The most important among them is my advisor, Prof. David Kriegman. David has been an exceptional source of ideas and advice. The quality that sets David apart is the *freedom* he gives to his students – the freedom to work on any problem, the freedom to work at any time, the freedom to work from any place, the freedom to collaborate with anyone, and most importantly the freedom to disagree with him. The reason I could work on many different problems during my graduate studies was because of this creative freedom and also because I could count on David when I hit a technical roadblock.

Prof. Serge Belongie changed the way I looked at computer vision. The computer vision course, CSE 252B, that Serge taught changed my understanding of this world in two ways – it sharpened my understanding of vision algorithms, and blurred my ability to differentiate between vision researchers and celebrities. I am also grateful to Serge for recommending me to David, who have together created an excellent research environment at Pixel Lab.

I have been extremely lucky to get a chance tp work with several collaborators who have contributed immensely toward a rich research experience. The most important aspect of research that I learnt from Prof. Todd Zickler was good writing. Todd always has great research ideas, and very clear ways of presenting them. My collaboration with Todd has therefore been exceptionally fruitful. One of the important research problems I worked on during my graduate studies was suggested by Prof. Peter Belhumeur. His comments and out of the box ideas were very refreshing.

Prof. Bridget Carragher and Prof. Clint Potter at the Automated Molecular Imaging (AMI) group (The Scripps Research Institute) gave me the same freedom to chose problems in Cryo-Electron Microscopy (cryo-EM), as I had enjoyed at Pixel Lab. I learnt the biological aspects of cryo-Electron Microscopy from Bridget and Clint, and they always encouraged me work on really challenging problems. Being part of the AMI family (that is how AMI is run!) was a great chance to utilize lessons learnt in computer vision to apply to real world problems. I would like to thank the entire AMI lab – Anchi Cheng, Craig Yoshioka, Denis Fellmann, Gabriel Lander, Jim Pulokas, Joel Quispe, Phil Mercurio, Scott Stagg, and Yuanxin Zhu.

Many thanks to Prof. Nuno Vasconecelos, Prof. Sam Buss, Prof. Pamela Cosman, and Prof. Truong Nguyen for agreeing to be on my committee and providing feedback.

I would also like to thank Prof. Mohan Trivedi who advised and funded me during the first year in graduate school.

"This is an important problem; we are improving the resolution at which truth can be perceived". That was Sameer Agarwal describing the importance of a molecular reconstruction problem we were working on together. Sameer is my friend, philosopher, guide, and chef. I learnt a lot about research, and life in general during our endless discussions over innumerable meals I had at his place. Manmohan Chandraker (Chandu) added a third dimension to those discussions and a vegetarian aspect to those meals. Rakesh Kumar was always available to give company for movies, weekend lunches, a game of racquet ball, or lunch burrito at Cortixan. Sachin Talathi is a dear friend, who I can always count on, and who generously lent me his car when I did not own one.

I would also like to thank past and present members of the Pixel Lab – Neil Alldrin, William Beaver, Kristin Branson, Will Chang, Piotr Dollar, Craig Donner, Neel Joshi, Ben Laxton, Kuang-chih Lee, Jongwoo Lim, Sebastian Magda, Ben Ochoa, Vincent Rabaud, Andrew Rabinovich, and Josh Wills.

My wingmates from IIT, who now remain connected via our wing Yahoo group, ensured that there was no dearth of Internet debate to fill up my free time. *Joga Bonito* at Panduz football helped me keep my sanity while writing this thesis. Finally, I would like to thank my roommates at different points of time in my graduate life – Sanjeev Kumar, Manish Patel, Puneet Sharma, Puneet Gupta, Abhishek Agarwal, Tathagata Mitra, and Sagnik Nandy.

Parts of this thesis are based on the papers I have co-authored with my collaborators.

1. Chapter 2 is based on "Beyond Lambert: Reconstructing Specular Surfaces Using Color" by S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. I developed the theory that a linear combination of color channels can be used to obtain diffuse channels and was responsible for implementation and experiments described in the paper. I also contributed toward writing the paper.

2. Chapter 3 is based on "Color Subspaces as Photometric Invariants" by T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur. I was responsible for developing the theory for obtaining specular-invariant image in an M channel image illuminated by N light colors. I conducted all the experiments described in the paper.

3. Chapter 4 is based on "Specularity Removal in Images and Videos: A PDE ap-

proach" by S. P. Mallick, T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman. I derived most of the Partial Differential Equations for specular highlight removal in images and videos. I was also responsible for implementation and experiments. In addition, I contributed toward writing the paper.

4. Chapter 5 is based on "On Refractive Optical Flow" by S. Agarwal, S. P. Mallick, D. J. Kriegman, and S. J. Belongie. I played a small part in theoretical development, and was responsible for implementation and experiments.

VITA

1980	Born, Cuttack, India.
2001	B. Tech., Indian Institute of Technology, Kharagpur.
2004	M. S., University of California, San Diego.
2006	Ph. D., University of California, San Diego.

PUBLICATIONS

S. P. Mallick, S. Agarwal, D. Kriegman, S. Belongie, C. Potter, and B. Carragher, "Vision in the Small: Reconstructing the Structure of Protein Macromolecules from Cryo-Electron Micrographs", *British Machine Vision Conference*, Volume I, Pages 1-6, September 2006, Edinburgh, United Kingdom.

S. M. Stagg, G. C. Lander, J. Pulokas, D. Fellmann, A. Cheng, J. D. Quispe, S. P. Mallick, R. M. Avila, B. Carragher, and C. S. Potter, "Automated cryoEM data acquisition and analysis of 284,742 particles of GroEL", *Journal of Structural Biology*, Volume 155, Issue 3, September 2006, Pages 470-481.

S. P. Mallick, S. Agarwal, D. Kriegman, S. Belongie, C. Potter, and B. Carragher, "Structure and View Estimation for Tomographic Reconstruction: A Bayesian Approach", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Volume II, Pages 2253-2260, June 2006, New York.

T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur, "Color Subspaces as Photometric Invariants", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Volume II, Pages 2000-2010, June 2006, New York.

S. P. Mallick, T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman, "Specularity Removal in Images and Videos: A PDE approach", *European Conference on Computer Vision (ECCV)*, Volume I, Pages 550-563, May 2006, Graz, Austria.

S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur, "Beyond Lambert: Reconstructing Specular Surfaces Using Color", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Volume II, Pages 619-626, 2005, San Diego, California.

S. P. Mallick, B. Carragher, C. S. Potter, D. J. Kriegman, "ACE: Automated CTF Estimation", *Ultramicroscopy*, Volume 104, Issue 1, August 2005, Pages 8-29.

S. Agarwal, S. P. Mallick, D. J. Kriegman, and S. J. Belongie, "On Refractive Optical Flow", *European Conference on Computer Vision (ECCV)*, Part II, Pages 483-494, 2004, Prague, Czech Republic.

S. P. Mallick, Y . Zhu and D. Kriegman, "Detecting Particles in cryo-EM micrographs using learned features" *Journal of Structural Biology*, Volume 145, Issues 1-2, January 2004, Pages 52-62.

Y. Zhu, B. Carragher, R. M. Glaeser, D. Fellmann, C. Bajaj, M. Bern, F. Mouche, F. Haas, R. J. Hall, D. Kriegman, S. J. Ludtke, S. P. Mallick, P. Penczeki, A. Roseman, F. Sigworth, N. Volkmann and C. Potter, "Automatic particle selection: results of a comparative study" *Journal of Structural Biology* , Volume 145, Issues 1-2, January 2004, Pages 3-14.

S. P. Mallick and M. M. Trivedi, "Parametric Face Modeling and Affect Synthesis", *International Conference on Multimedia and Expo (ICME)*, Pages 225-228, July, 2003, Baltimore, Maryland.

J. McCall, S. P. Mallick, and M. M. Trivedi, "Real-Time Driver Affect Analysis and Tele-viewing System", *IEEE Intelligent Vehicles Symposium (IV)*, Pages 372-377, June, 2003, Columbus, Ohio.

T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur, "Color Subspaces as Photometric Invariants", submitted to *International Journal of Computer Vision*.

ABSTRACT OF THE DISSERTATION

The structure and analysis of a scene: A photometric approach

by

Satya Prakash Mallick
Doctor of Philosophy in Electrical Engineering (Signal & Image Processing)
University of California San Diego, 2006
Professor David Kriegman, Chair
Professor Nuno Vasconcelos, Co-Chair

The field of computer vision has grown enormously in the last decade, and significant progress has been made in many different areas. Several reasonable solutions to technically challenging problems like stereo, optical flow, tracking, detection, recognition, shape from shading, photometric stereo *etc.* have emerged. The principle challenge today is to make these algorithms, designed and tested in laboratory conditions, to work in real life situations. Several assumptions that make vision problems tractable, break down in the real world. The most common among these assumptions is that the reflectance of all objects in an image is Lambertian.

This dissertation focuses on algorithms that can be applied to a large class of non-Lambertian materials whose reflectance can be described as a linear combination of diffuse (matte) and a specular (gloss) term. We propose a transformation of the RGB color space that recovers two purely diffuse components of an RGB color image. We extend our approach to the case of multiple colored illuminants in a scene, and derive a class of photometric invariants. These photometric invariants can then be used as inputs to several computer vision algorithms to improve their performance. This is empirically demonstrated for photometric stereo, shape from shading, stereo, optical flow, passive photometric stereo, and color based segmentation.

While the problem of recovering two purely diffuse components of a color image is well-posed, the problem of completely separating the specular and diffuse components of a color image is an ill-posed one. We propose a unified approach to solve this problem for images and videos of textureless as well as textured scenes. A partial differential equation is derived that uses neighborhood color information of a pixel to iteratively erode

the specular component of color. Finally, we introduce an application called *Dichromatic Editing* – the process of independently editing and recombining the two components of color to achieve a variety of visual effects.

This thesis also ventures into a relatively unexplored and difficult problem of inferring information about a scene where refraction occurs. We present a generalization of the optical flow equation for the case when light rays bend (due to reflection or refraction) as they travel from an object in the scene to the camera. By solving the unknowns in this flow equation, we show that it is possible to recover the path of light transport as it travels from a background plane to the camera through a refractive medium.

1

Introduction

"Where shall I begin, please your Majesty?" he asked. "Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

Alice in Wonderland
Lewis Carroll

A bundle of rays captured by a camera tells many interesting, yet incomplete stories about the interplay between different materials and lights in the scene. With the knowledge of the underlying physics of image formation, the images can be used to recover the geometric structure, material properties, and light sources in the scene.

There are two parts to the photometric enterprise of analyzing a scene. The first part deals with the formation of mathematical models for physical phenomena. Photometry – established as a major field in the eighteenth century after Johann Heinrich Lambert's seminal work called *Photometria* – deals with the study of different physical phenomena like reflection, refraction, dispersion, polarization, extinction, and spectral distribution of light. Centuries of research has produced an excellent understanding of the physics governing these different photometric phenomena.

The second part is the inverse problem of recovering information about a scene based on physical principles. This line of research led by Berthold K.P. Horn in the 1970s is referred to as "Physics based vision". It includes using the mathematical models for the properties of surfaces , illuminants, and sensors to recover information about geometrical shape, reflectance, color, shadows, inter-reflections, refractive indices of surfaces, and path of light transport, as well as properties of the sensors such as a camera's spectral response curve. This dissertation extends the tradition of "Physics based vision".

The entire field of computer vision deals extensively with inference problems, where images serve as observations of a scene, and the objective is to obtain certain information about the scene. For example, in binocular stereo, two images of a scene are used to obtain depth (structural) information about the scene. More often than not, these inference problems are ill-posed, and a unique solution can be resolved only under certain assumptions. At other times, assumptions are not strictly necessary for making the problem well posed, but are used to make the problem more tractable leading to an elegant solution. However, deviation from these assumptions in the real world can limit the applicability of various algorithms. Sometimes, a minor variation of the algorithm is able to handle this deviation. At other times, the assumption is so critical to the development of the algorithm that the algorithm becomes unsuitable for the real world and a new one needs to be devised from scratch.

In computer vision, it is common to assume that objects in an image are Lambertian (matte). However, it is well known that different materials reflect light in widely different ways. The interaction of a material with light is usually modeled as a four dimensional function called the Bidirectional Reflectance Distribution Function (BRDF). Lambertian reflectance is the simplest form of the BRDF, and materials with a Lambertian BRDF follow Lambert's cosine law of reflection defined as

Definition 1.1. Lambert's cosine law: The total radiant power observed from a "Lambertian" surface is directly proportional to the cosine of the angle θ between the observer's line of sight and the surface normal.

The cosine law is explicitly used in algorithms like shape from shading, and photometric stereo to recover the 3D structure of objects. The cosine law is a consequence of the fact that Lambertian materials reflect light equally in all directions. As a result, if the lighting in a scene remains the same, a surface patch appears the same when viewed from two different camera positions. This property of Lambertian surfaces forms the basic assumption in a broad range of vision algorithms like multiview stereo, optical flow, tracking, and recognition.

Objects in the real world rarely follow the Lambertian reflection model. Consequently, in each problem domain where the Lambertian assumption is made, there are a few pieces of work that address the extension to non-Lambertian surfaces. In this thesis we present a unified approach that can potentially improve the performance of *all* algorithms that make the Lambertian assumption. Given that we cannot change the real world around us, a natural question to investigate is whether we can modify the acquired

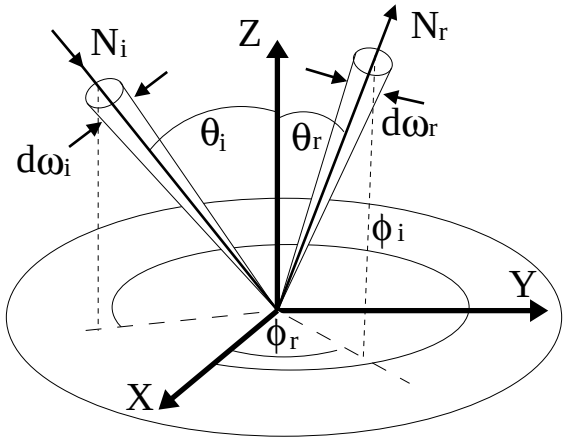


Figure 1.1 Bidirectional Reflectance Distribution Function (BRDF) $f(\theta_i, \phi_i, \theta_r, \phi_r)$, is the ratio of reflected radiance along the direction (θ_r, ϕ_r) to the incident irradiance in a particular direction (θ_i, ϕ_i).

images so that the Lambertian part of reflectance is isolated, and serves as input to all vision algorithms that assume Lambertian reflectance. A significant part of this thesis is devoted to answering the above question, and identifying the limitations of such an approach.

We now define a few basic concepts necessary for understanding the rest of the thesis. In Sec. 1.2 we discuss a few different vision problems, the algorithms used to solve the problems, and the existing literature on handling non-Lambertian surfaces in specific problem domain.

1.1 Background

At an appropriate scale, reflectance at a surface point is described by the bi-directional reflectance distribution function.

Definition 1.2. Bi-directional Reflectance Distribution Function (BRDF) is defined as the ratio of the reflected radiance (watts/meters^2/steradians) in a particular direction (θ_r,ϕ_r) to the incident irradiance (watts/meters^2) from direction (θ_i,ϕ_i). Mathematically, the BRDF, $f(\theta_i, \phi_i, \theta_r, \phi_r)$ is given by

$$f(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{\delta L_r(\theta_r, \phi_r)}{L_i(\theta_i, \phi_i) \cos\theta_i \delta\omega_i} \tag{1.1}$$

where L denotes radiance and ω represents solid angle. The subscripts i and r are used to denoted incident and reflected quantities. The BRDF of a material can also vary with

the wavelength of light incident on it, in which case the BRDF $f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r)$ is a five-dimensional function of wavelength λ and imaging geometry

$$f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) = \frac{\delta L_r(\lambda, \theta_r, \phi_r)}{L_i(\lambda, \theta_i, \phi_i) \cos\theta_i \delta\omega_i} \tag{1.2}$$

A few properties of a BRDF model are enumerated below.

1. Energy Conservation: Any BRDF model assumes that the total energy incident on a surface patch is either reflected or absorbed. A BRDF model does not allow for self-emission. Therefore,

$$\int_\Omega f(\theta_i, \phi_i, \theta_r, \phi_r) \cos\theta_r d\omega_r \leq 1 \quad \forall \quad \theta_i, \phi_i \tag{1.3}$$

2. Helmholtz Reciprocity: The BRDF remains unchanged if the incident and the reflected directions are interchanged.

$$f(\theta_i, \phi_i, \theta_r, \phi_r) = f(\theta_r, \phi_r, \theta_i, \phi_i) \tag{1.4}$$

3. No sub-surface scattering: Sub-surface scattering refers to the phenomenon in which a light ray enters a material, undergoes multiple scattering inside the material, and is subsequently reflected back. A BRDF model does not allow for sub-surface scattering.

In addition to the above properties, for the purpose of mathematical convenience, a BRDF model is sometimes assumed to have the following additional properties

1. Isotropy: An isotropic BRDF is independent of rotation about the surface normal. The BRDF reduces to a 3-parameter function given by

$$f(\theta_i, \phi_i, \theta_r, \phi_r) = f(\theta_i, \theta_r, \phi_i - \phi_r) \tag{1.5}$$

2. Separability: When the BRDF of a material depends on the wavelength of incident light, it is often assumed that the effects of wavelength are independent of the effects of imaging geometry.

$$f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) = f_\lambda(\lambda) f_{\boldsymbol{\theta}}(\theta_i, \phi_i, \theta_r, \phi_r) \tag{1.6}$$

Since the BRDF is a four dimensional function of incident and reflected directions, the estimation of the BRDF of a material requires illuminating a surface patch of the material from all possible directions and measuring the reflected light from all possible directions in a unit hemisphere around the surface. In many vision and graphics applications, instead of accurate estimation of the BRDF, a parametric approximation is used. The Lambertian reflection model (constant BRDF) is the simplest and most widely used approximation. It captures the matte or diffuse component of the reflection. It is common in computer graphics to model glossy materials like plastics using the isotropic Phong reflectance model which assumes that the BRDF is a combination of an ambient term, a diffuse term, and an ad-hoc specular (gloss) term [87]. While the Phong model is a good approximation for smooth plastics, it cannot be used to effectively model metals and rough surfaces. Several other *physics based* BRDF models have emerged that more accurately approximate the true BRDF of different materials [5, 27, 36, 78, 107]. Notable among these is the Torrance-Sparrow reflectance model [107] which assumes that the surface being simulated is composed of a collection of mirror-like micro facets that are oriented in random directions on the surface. Surface BRDF is simulated by assuming a distribution of the micro facets, and the amount by which the micro facets shadow and mask each other. Fresnel reflection is used to model the interaction of light with the micro facets. In recent years non-parametric models for BRDF have emerged. The BRDF is measured directly and suitable basis functions are used for representation [52]. In our work, we assume Shafer's [98] *Dichromatic* reflection model.

1.1.1 Image Formation

According to Eq. 1.2, the radiance reflected in direction (θ_r, ϕ_r) by a surface patch is given by

$$L_r(\lambda, \theta_r, \phi_r) = \int_\Omega f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) L_i(\lambda, \theta_i, \phi_i) \cos\theta_i d\omega_i \tag{1.7}$$

Assuming that there is a single point (sphere with an infinitesimal radius) light source located at an infinite distance from the scene, the above expression can be written as

$$L_r(\lambda, \theta_r, \phi_r) = f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) L_i(\lambda) \cos\theta_i \tag{1.8}$$

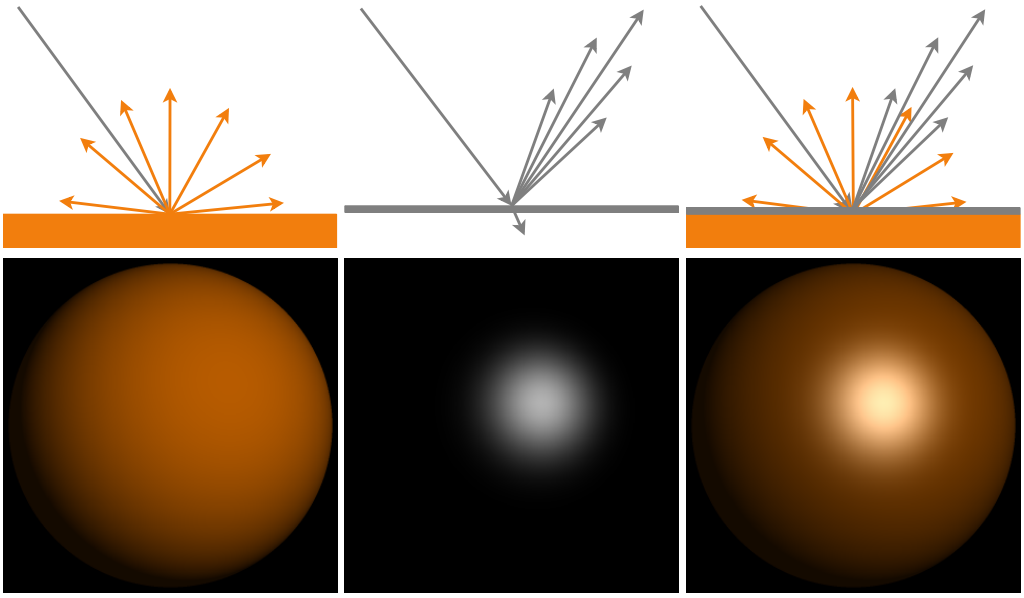


Figure 1.2 Lambertian versus dichromatic reflection: The top row shows the interaction of three different materials with an incident light ray in 1D. A Lambertian surface reflects light in all directions equally. A transparent layer refracts most of the light incident on it, and reflects a small fraction. Dichromatic materials are modeled as a combination of a Lambertian surface covered by a transparent layer. The bottom row shows rendered spheres corresponding to the three material types.

Let $C_k(\lambda)$ represent the camera sensitivity of the k^{th} channel. The image formation equation is given by

$$I_k = \int f(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) L_i(\lambda) C_k(\lambda) \cos\theta_i d\lambda \tag{1.9}$$

where I_k represents the k^{th} image channel.

1.1.2 The Dichromatic Model

The dichromatic model of reflectance is a common special case of the BRDF model, and it was originally developed by Shafer [98] to model dielectrics. It assumes that the BRDF of the surface can be decomposed into two additive components: the interface (specular) reflectance and the body (diffuse) reflectance. Furthermore, it assumes that each of these two components can be factored into a univariate function of wavelength and a multivariate function that depends on the imaging geometry. That is,

$$f(\lambda, \boldsymbol{\theta}) = g_d(\lambda) f_d(\boldsymbol{\theta}) + g_s(\lambda) \tilde{f}_s(\boldsymbol{\theta}). \tag{1.10}$$

where λ is the wavelength of light and $\boldsymbol{\theta} = (\theta_i, \phi_i, \theta_r, \phi_r)$ describes the imaging geometry. Finally, the model assumes that the index of refraction on the surface is constant over the visible spectrum—a valid assumption for many materials—so that $g_s(\lambda)$ is a constant function. This leads to the common expression for the BRDF of a dichromatic surface,

$$f(\lambda, \boldsymbol{\theta}) = g_d(\lambda) f_d(\boldsymbol{\theta}) + f_s(\boldsymbol{\theta}), \tag{1.11}$$

where $f_s(\boldsymbol{\theta}) = g_s \tilde{f}_s(\boldsymbol{\theta})$. The function $g_d(\lambda)$ is often referred to as the *spectral reflectance* and is an intrinsic property of the material.

Figure 1.2 shows an intuitive explanation of the dichromatic model. Many surfaces can be thought to be composed of a diffuse component with a transparent reflecting layer on top. According to Fresnel's law of reflection, The transparent layer allows most of the light to be transmitted, and reflects a small fraction of incident light. The diffuse layer reflects the color of the surface, while the reflection off the transparent layer is the same as the light source color.

Even though it was originally used to describe the reflectance of dielectrics [98], the dichromatic model has been used successfully as an approximation of the reflectance of many different materials. For example, although the reflectance of human skin is more accurately described by a higher dimensional function (the 8-dimensional BSSRDF [115]), the BRDF approximation—and more specifically, the dichromatic BRDF model—has proven useful for a number of applications, including face recognition [19, 32] and pigment-based image analysis and synthesis [109]. The validity of this approximation for skin reflectance is also supported by the spectrophotometry measurements of Marszalec *et al.* [71].

In order to derive an expression for the image of a dichromatic surface, consider an observed surface point $\mathbf{x}$ illuminated from direction $\hat{\mathbf{l}}$, and let $\boldsymbol{\theta} = (\theta_i, \phi_i, \theta_r, \phi_r)$ represent this direction and the viewing direction in a local coordinate system. We assume that the sensor is a linear devise with sensitivity function $C_k(\lambda)$, and that the spectral power distribution (SPD) of the light source is given by $L(\lambda)$. Under these conditions, assuming that the BRDF is given by Eq. 1.11, the sensor response is

$$I_k = \left(D_k f_d(\boldsymbol{\theta}) + S_k f_s(\boldsymbol{\theta})\right) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \tag{1.12}$$

with

$$\begin{aligned} D_k &= \int C_k(\lambda) L(\lambda) g_d(\lambda) d\lambda \\ S_k &= \int C_k(\lambda) L(\lambda) d\lambda. \end{aligned} \tag{1.13}$$

An RGB color vector $\mathbf{I} = [I_R, I_G, I_B]^\top$ from a typical camera consists of three such measurements, each with a different sensor response with support in the visible spectrum.

Note that S_k represents the effective source strength as measured by the k^{th} sensor and is independent of the surface being observed. For notational simplicity, we define $\mathbf{S} = [S_R, S_G, S_B]^\top$ (with a corresponding definition for $\mathbf{D}$), and since scale can be absorbed by f_d and f_s, we assume $\|\mathbf{D}\| = \|\mathbf{S}\| = 1$.

1.2 Lambertian Assumption in Computer Vision

The most common photometric assumption made by computer vision algorithms is that the scene is Lambertian. The Lambertian assumption makes many difficult problems in computer vision tractable. For example, matching pixels (or patterns) across multiple images (or frames of a video) is a central problem in stereo, optical flow, object detection, and recognition. These algorithms assume that appearance of a surface patch does not change when the camera is moved, but the lighting remains constant. This assumption holds only when all surfaces in the scene follow the Lambertian reflectance model.

Several other computer vision algorithms, like shape from shading, photometric stereo, and passive photometric stereo directly exploit Lambert's cosine law to recover 3D structure of a scene. The results of these algorithms are severely affected by the deviation of surface BRDF from the assumed Lambertian model.

Two classes of techniques have emerged for handling non-Lambertian surfaces.

1. *Specular-diffuse separation techniques.* The algorithms take as input one color image, and output a color diffuse image and a monochromatic specular image. These algorithms have the potential to improve the performance of computer vision techniques across the board. However, the problem of specular-diffuse separation is inherently ill-posed as it attempts to recover a 4-channel image (3 diffuse channels and 1 specular channel) from an input 3-channel image.

2. *Problem specific techniques.* These techniques address the issue of specular high-

lights in their problem domain.

Specular diffuse separation is dealt with in detail in chapter 4. An introduction to different computer vision problems, and a review of literature related to handling non-Lambertian surfaces in each case follows.

1.2.1 Vision Algorithms

In this section, the basics of a few vision algorithms are covered. The description of each algorithm is followed by a literature review of techniques to handle non-Lambertian reflectance.

Photometric Stereo

Photometric stereo [121] is the process of recovering 3D shape from a series of images captured from a fixed viewpoint under changing light source direction. The camera is typically assumed to be orthographic, and the light sources are assumed to be point lights located far from the object to be reconstructed. Figure 1.3 shows a typical photometric stereo setup.

Since they rely on the inversion of the image formation process, photometric stereo methods typically require significant knowledge about the reflectance of surfaces in the scene. Usually the scene is assumed to be Lambertian. The i^{th} light source can be represented as a 3×1 vector $\hat{l}_i$, the magnitude of which specifies the effective strength of the source, and the direction specifies the orientation of the line connecting the object to the light source. Given $n > 3$ images $I = [\ i_1\ i_2\ \cdots i_n]$ of a scene point, and known light sources $L = [\ \hat{l}_1\ \hat{l}_2\ \cdots \hat{l}_n]$, the image formation equation can be written as

$$I = \rho L^\top \hat{n} \tag{1.14}$$

where ρ is the surface texture, often referred to as the *albedo*. The surface normal $\hat{n}$ can be recovered by solving the above linear system. Finally, the normals are integrated to recover the surface. The normal at a point (x, y) of a surface $z(x, y)$ is given by $\hat{n} = [-\frac{\partial y}{\partial z}\ \frac{\partial x}{\partial z}\ 1]^\top$. The recovered normals can therefore be integrated to recover the surface $z(x, y)$ up to a scale.

Many photometric stereo techniques assume that surfaces are Lambertian [121], and when this reflectance assumption is not satisfied, the accuracy of the recovered shape is compromised. Figure 1.4 shows an example where photometric stereo is applied to

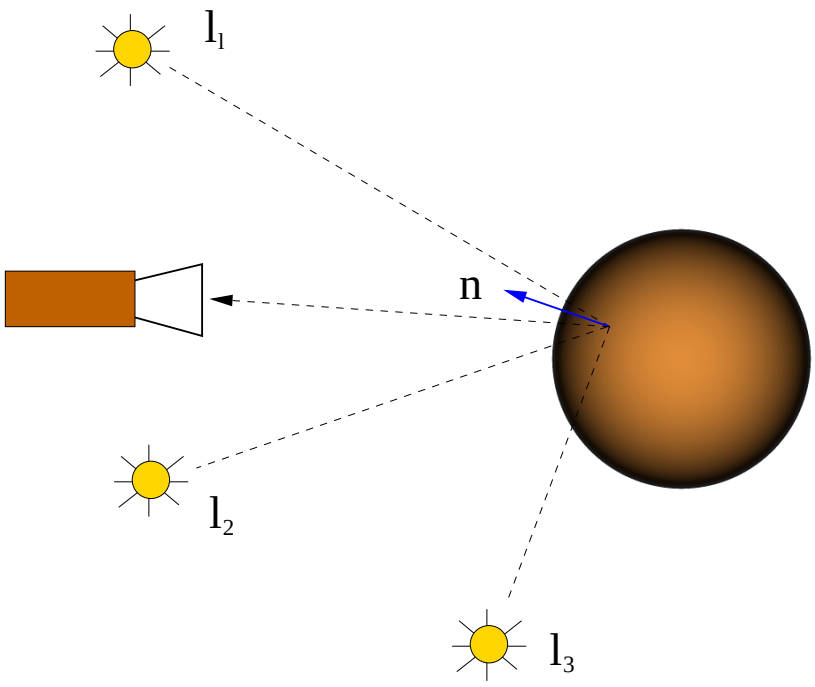


Figure 1.3 A typical photometric stereo setup. Three images of an object are acquired using a static camera by varying the position of the light source.

images of a specular sphere and it recovers a wrong structure. This is because the algorithm wrongly attributes the highlight on the sphere to structure, while in fact it is due to reflectance.

The dramatic degradation in the quality of the recovered structure as the reflectance of an object deviates from the Lambertian model has prompted researchers to address this problem. Coleman and Jain [26] were perhaps the first to present a photometric technique for reconstructing non-Lambertian surfaces without an explicit reflectance model. In their method, the BRDF is assumed to be a linear combination of a Lambertian diffuse component and an undefined specular component with limited angular support. When images with four point-source illuminations are available, specular measurements can be treated as outliers and discarded, provided that the illumination directions are far from one another relative to the angular extent of the specular lobe. (This ensures that the specular reflectance component is zero for three of the four observations of each surface point.) Barsky and Petrou [13] refine this technique by using color information to improve the detection of specular measurements. Like the original work, however, specular measurements are treated as outliers, and the specular component is assumed to have limited angular support.

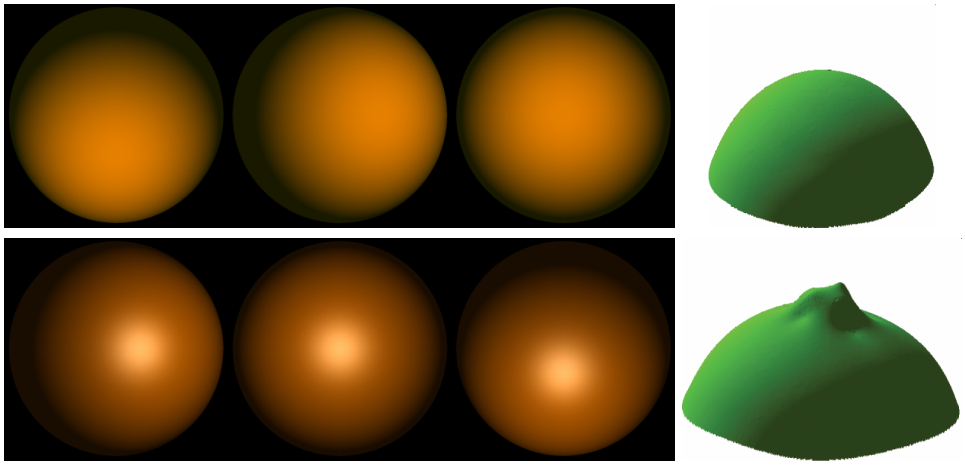


Figure 1.4 Result of deviation from Lambertian assumption in photometric stereo. The top row shows three images of a diffuse sphere taken under three different lighting conditions, and the structure recovered using traditional photometric stereo. The bottom row shows three images of a non-Lambertian sphere under three different lighting conditions, and the recovered structure. The incorrect 3D structure is a result of assuming a Lambertian BRDF.

Researchers have used several different assumptions about reflectance to deal with non-Lambertian surfaces. For example, Silver [99] assumed that the reflectance is given *a priori* by a reference object. This technique was independently rediscovered and extended by Hertzmann and Seitz [38] who modeled surface reflectance using a linear basis of reference objects. Several researchers have used an analytic BRDF model [44, 76, 100].

Another approach to photometric stereo for non-Lambertian surfaces is to assume dichromatic surfaces, and to remove highlights as a pre-processing step. This is the approach taken by Schlüns and Wittig [96], who assume homogeneous dichromatic surfaces, and separate the diffuse and specular components using color histogram analysis techniques similar to Klinker *et al.* [51]. Sato and Ikeuchi [92] take a similar approach, but avoid the restriction to homogeneous surfaces by using a large number of light source directions to compute a distinct color histogram at each point. Because these methods explicitly recover the diffuse and specular components, they have the additional benefit of providing an estimate of the diffuse color $\mathbf{D}$ at each point in addition to recovering the surface shape. Since they are based on conventional dichromatic color separation, however, they are subject to the restrictions discussed in the previous section. Most importantly, they assume that the specular lobe is narrow relative to the surface curvature, an assumption similar to that underlying the four-source method of Coleman and

Jain [26].

Shape from Shading

Shape from shading (SFS) is a method for recovering the 3D structure of a textureless Lambertian surface (having constant albedo) from a single image illuminated by a light source with known direction. Unlike photometric stereo, SFS is an ill-posed problem – we have a single equation relating the brightness of a pixel to the cosine of the angle between the light source direction and the surface normal at a point, and the objective is to recover the unit normal resulting in two unknowns. Solving SFS requires additional constraints (assumptions).

Zhang *et al.* [126] provide a thorough review of various SFS algorithms. They classify SFS algorithms as *Minimization based* – an energy function is minimized to obtain the 3D structure [31, 41, 45, 127], *Propagation based* – shape information is propagated from a few surface points (*e.g.*, singular points) [6], *Local approaches* – assumption is made about the local shape of the surface [83], *Linear approaches* – the solution is computed based on linearization of the reflectance map [84, 108].

Minimization based algorithms are arguably the most popular approaches. As mentioned in [126], the ill-posed problem of SFS becomes solvable by incorporating one or more of the following assumptions in a variational framework

1. *Brightness* constraint requires the reconstructed image R to have the same intensity as the input image I. In other words, it requires the minimization of the following error term

$$E_b = \int \int (I - R) .^2 dx \; dy \tag{1.15}$$

2. *Smoothness* constraint, introduced by Ikeuchi and Horn [45], demands a smooth reconstructed surface. Let p and q represent the surface gradient in the x and y directions. The smoothness constraint requires the minimization of the following error

$$E_s = \int \int \left(p_x^2 + q_y^2\right) dx \; dy \tag{1.16}$$

3. *Integrability* constraint, introduced by Frankot and Chellappa [31], requires the recovered surface normal to be integrable; which is equivalent to minimizing the

following error

$$E_i = \int\int (p_x - q_y)^2 \, dx \; dy \tag{1.17}$$

4. *Intensity gradient* constraint, introduced by Zheng and Chellappa [127], requires the intensity gradient of the reconstructed image and input image to be the same. This corresponds to minimizing the following constraints

$$E_g = \int\int \left((R_x - I_x)^2 + (R_y - I_y)^2\right) dx \; dy \tag{1.18}$$

5. *Unit normal* constraint requires the recovered normals $\hat{n}$ to be unit magnitude

$$E_n = \int\int \left(\|\overrightarrow{n}\|^2 - 1\right) dx \; dy \tag{1.19}$$

Recently it was shown by Prados and Faugeras [89] that if the attenuation of light with distance is modeled, and the light source is co-located with the camera center, SFS becomes a completely well-posed problem. Using the same modeling as Prados and Faugeras [89], and modeling the diffuse reflectance using a more realistic Oren and Nayar [78] reflectance model, Ahmed and Farag [2] propose a new SFS algorithm that resolves the convex-concave ambiguity.

There is a sparse body of literature dealing with SFS for non-Lambertian surfaces. Bakshi and Yang [8] modified a heuristic based SFS algorithm called *Shading Logic* [112] to handle the Torrance and Sparrow [107] reflectance model. Tian and Tsui [105] assume the surfaces to be dichromatic (instead of Lambertian). The specular reflection is assumed to lie along the direction of perfect mirror reflection. Hue information is used to segment the image into perfect diffuse and hybrid (specular + diffuse) reflectance. Finally, they extend the Zheng and Chellappa [127] SFS algorithm to account for the new reflectance model. Ragheb and Hancock [90] assume the surface reflectance at a pixel to be either diffuse or specular. Reflectance of specular pixels is modeled using the Torrance and Sparrow [107] reflectance model. A Bayesian framework is set up by assuming different probability distributions for diffuse and specular pixels. Using the iterated conditional modes [14] algorithm, the *a posteriori* probability of a pixel being diffuse or specular is computed. Once an assignment is made, the surface normals are reconstructed using the smoothness assumption.

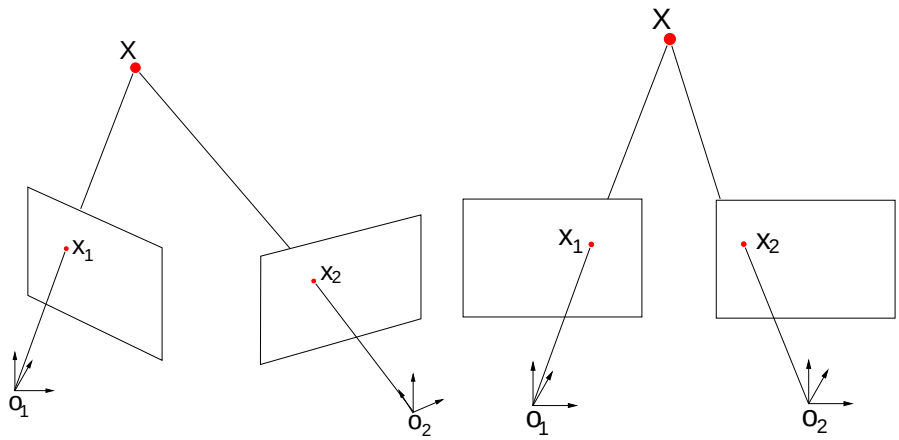


Figure 1.5 (a) A simplified stereo set up. The points x_1 and x_2 are images of the same 3D point X and are called *corresponding points*. (b) Geometrical interpretation of rectification. The images acquired in a stereo setup can be transformed so that corresponding points lie on the same scanline in the two images. This process is called *Rectification*. Geometrically two rectified images are equivalent to two images acquired using a stereo rig with their images planes aligned, and separated by a distance.

Stereo

The problem of stereo vision has fascinated many researchers, and is still one of the most active research areas in computer vision. Motivated partly by human vision, a stereo vision system aims to recover the 3D depth of a scene from two images obtained from two different viewpoints. Figure 1.5 (a) shows a simplified stereo setup in which a single scene point X projects to x_1 and x_2 (called *corresponding points*) on the image planes of two pin hole cameras with optical centers at o_1 and o_2. When the locations x_1, x_2, o_1, and o_2, and the camera focal lengths are all known, in theory, the depth of the 3D point X can be found by intersecting the vectors $o_1\bar{x}_1$ and $o_2\bar{x}_2$. Therefore, the fundamental problem in stereo vision is to find dense correspondence among pixels of the two images. Fortunately, it can be shown that the geometry of image formation imposes a constraint on corresponding points, such that the point corresponding to a particular pixel in one image lies on a line in the other image. This constraint, called the *Epipolar constraint*, allows the images to be geometrically transformed (rectified), so that corresponding points lie along the same scanline of the two images. Figure 1.5 (b) shows a geometrical interpretation of stereo rectification. In turn, the distance $x_1 - x_2$ between coordinates of corresponding points in rectified images, called the *disparity*, is proportional to the inverse depth of the 3D point X [20]. Figure 1.6 shows a pair of

rectified images, and the disparity map.

Dense correspondence between images is inherently under-constrained [10], and therefore requires assumptions about reflectance of scene surfaces. Most stereo algorithms assume that the appearance of an image feature or a patch does not change across images. This simplifying assumption implicitly assumes all surfaces in the scene are Lambertian. The most common approach to handle non-Lambertian specular highlights is to treat them as outliers. These outliers are either explicitly detected and removed [23], or handled implicitly by modifying the matching function to a robust one.

A natural extension of binocular stereo is multiview stereo. Seitz *et al.* [97] present a comprehensive evaluation of multiview stereo algorithms. In addition to more precise depth estimation due to availability of more data, some multiview stereo algorithms also attempt to address the problem of specular highlights [15, 48, 57, 60, 122]. A further extension of multiview stereo is to use multiple views under different (possibly known) lighting conditions [29, 65, 129]. Some of these methods, like the Helmholtz stereopsis [129] can be used to extract depth and normal information about surfaces with arbitrary BRDF.

A small number of binocular stereo algorithms propose methods to handle specular highlights. Bhat and Nayar [15] assume a lower bound on surface roughness and propose an optimal stereo configuration that maximizes the accuracy of depth estimation even in the presence of specular highlights. They acknowledge that it is extremely difficult to estimate surface roughness in a practical situation, and therefore, they propose a roughness independent trinocular configuration such that accurate depth at each pixel can be measured using at least two of the three cameras. Zickler *et al.* [131] extend Helmholtz stereopsis [129] and show that "the fundamental Helmholtz stereopsis constraint defines a nonlinear partial differential equation, which can be solved using only two images". Binocular Helmholtz stereopsis can handle arbitrary BRDFs.

Unlike the two stereo algorithms mentioned above, a few stereo algorithms do not require special camera and/or light configurations. Kim *et al.* [50] propose a new measure of patch similarity that does not depend on pixel intensities, but instead maximizes mutual information for correspondence. Empirically it is shown to improve performance for non-Lambertian surfaces. In [124], specular highlights are treated as outliers by first separating the specular and diffuse components of color, and then assigning lower weights to pixels containing a large specular component in the matching process. Yoon and Kweon [123], assume the dichromatic model for surface reflectance and propose a two band specular free image ($I - \min(I_R, I_G, I_B)$) that is used as input

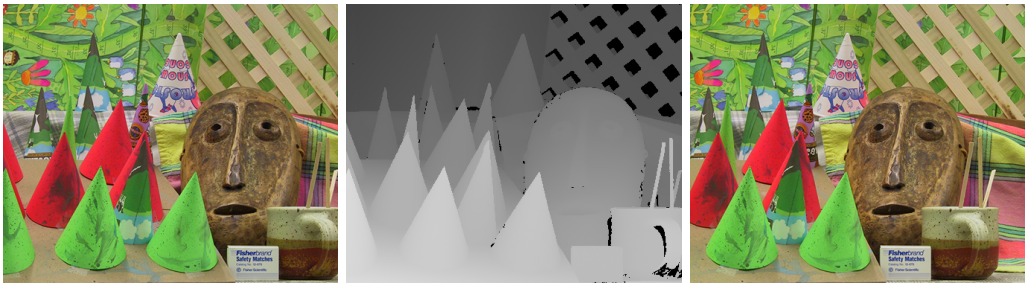


Figure 1.6 In a typical stereo setup, two images of a scene (left and right) are acquired from two different viewpoints. The output of a stereo vision algorithm is a depth map (center) with intensity proportional to the depth of the scene as measured from one of the cameras. *The above images were obtained from the Middlebury Stereo Page [1].*

to a standard vision algorithm.

Optical Flow

3D motions of objects relative to the camera appear as 2D motions in the image plane. This image motion is called *Optical Flow*. In other words [43], "The optical flow is a velocity field in the image that transforms one image into the next image in a sequence."

The problem of locally recovering optical flow is an ill-posed one. The following assumptions make the computation of optical flow feasible.

1. *Brightness Constancy* constraint [42] states that the brightness of an image patch $I(x, y, t)$ remains constant as it moves to location $(x + \delta x, y + \delta y)$ over time δt. Therefore,

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{1.20}$$

$$= I(x, y, t) + \delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} + \delta t \frac{\partial I}{\partial t} + \epsilon \tag{1.21}$$

   where ϵ contains higher order terms. Cancelling the term $I(x, y, t)$ from both sides of the equation, and dividing through by δt we obtain

$$\frac{\delta x}{\delta t}\frac{\partial I}{\partial x} + \frac{\delta y}{\delta t}\frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} + O(\delta t) = 0 \tag{1.22}$$

As $\partial t \rightarrow 0$, we obtain the *Optical Flow Equation*

$$\begin{aligned} &\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \\ &\Rightarrow \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \end{aligned} \tag{1.23}$$

where $(u, v) = (\frac{dx}{dt}, \frac{dy}{dt})$ is the optical flow vector. Note that if the flow is due to the motion of the camera while the objects remain stationary, the brightness constancy constraint implicitly assumes that the objects in the scene are Lambertian. On the other hand, if the flow is due to the motion of the objects, the objects are assumed to be locally planar in addition to being Lambertian.

2. *Smoothness Constraint* The flow Eq 1.23 provides only one contraint on the flow vector (u, v), and additional constraints are required to solve it. Smoothness of the velocity field [42, 62] is usually used as the additional constraint required to solve the optical flow equation. Smoothness of the flow field can be expressed as a minimization of the gradient of the optical flow velocity [42]

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \text{ and } \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2$$

Horn and Schunck [42] use a variational framework in which an error function consisting of a brightness constancy term and a smoothness term is minimized. Lucas and Kanade [62] assume that the flows of all pixels in a small neighborhood around a pixel are the same. This allows them to set up a system of linear equations which can be solved to obtain the optical flow locally at each pixel.

3. *Gradient Constancy* constraint [111] assumes that the gradient of image patch $I(x, y, t)$ remains constant as it moves to a new location $(x + \delta x, y + \delta y)$ after time δt. It can be written as

$$\nabla I(x, y, t) = \nabla I(x + \delta x, y + \delta y, t + \delta t) \tag{1.24}$$

This constraint was introduced to relax the brightness constancy assumption which frequently breaks down in natural scenes.

Black and Anandan [17] introduced robust estimation of the optical flow by rejecting outliers using robust error functions (for example, the truncated quadratic) instead of the quadratic error function used in least squares estimation. Traditional least

squares based optical flow estimation algorithms aggregate information over a spatial neighborhood and result in incorrect estimates at motion boundaries. Robust estimation of optical flow alleviates the "problems of oversmoothing and noise sensitivity typically associated with the least-squares formulations."

Yet another recent development in optical flow literature is the use of color images [4, 116, 118]. Barron and Klette [11] present a survey of color based optical flow methods. The optical flow equation can be extended to each channel separately

$$\begin{aligned}
\frac{\partial R}{\partial x}u + \frac{\partial R}{\partial y}v + \frac{\partial R}{\partial t} &= 0 \\
\frac{\partial G}{\partial x}u + \frac{\partial G}{\partial y}v + \frac{\partial G}{\partial t} &= 0 \\
\frac{\partial B}{\partial x}u + \frac{\partial B}{\partial y}v + \frac{\partial B}{\partial t} &= 0
\end{aligned} \tag{1.25}$$

The flow vector (u, v) can be calculated from the over determined system in Eq. 1.25. However, due to the high correlation between the three color channels, the smoothness constraint is used in addition to Eq. 1.25. Almost all color based optical flow algorithms use color information to produce photometric invariant images that are independent of shading (for example, the two channel normalized RGB image $\left(\frac{R}{R+G+B}, \frac{G}{R+G+B}\right)$) and specular highlights (for example, the hue component of the HSV color space). The optical flow calculated using these photometric invariant images as inputs is a better representation of the underlying scene motion.

Photometric/Geometric Stereo

In recent years there have been a few attempts to recover the 3D structure of a Lambertian object using a video of the moving object in front of a stationary camera [58, 125]. As the object moves in front of the camera, "its appearance changes in two fundamental ways: geometrically and photometrically" [125]. The basic idea is to recover 3D shape by utilizing geometrical constraints (optical flow) for heavily textured regions and photometric constraints (shading) for lightly textured regions. Notice that in these approaches, both normal information and depth information are estimated.

Zhang *et al.* [125] use standard structure from motions algorithms to first estimate the camera motion and an initial estimate of lighting using a small set of tracked features. They then device an iterative scheme to alternately estimate the shape and lighting. The surface normals are first estimated at each pixel, and a rough surface constructed by integrating the normals. This is followed by refinement of the structure by

using a dense optical flow algorithm devised to take into account varying illumination. The optical flow problem is formulated such that "surface positions, normals, motion, and illumination are all coupled together into the same minimization problem".

Lim *et al.* [58] track a few feature points across all frames of a video, and reconstruct a piecewise linear 3D shape of the object. An iterative scheme based on the current reconstruction, and available photometric information is used to refine the 3D shape. At each iteration, normals are estimated up to a *Generalized Bas Relief* (GBR) ambiguity, and integrated to obtain a new surface reconstruction. The GBR is resolved using the tracked feature points.

Both the above algorithms assume that the surfaces in the scene are Lambertian. To the best of our knowledge, we do not know of any method extending the above approaches to non-Lambertian surfaces.

1.3 Thesis Overview

This section provides a brief overview of the rest of the thesis.

1.3.1 The SUV Color Space

The ubiquity of the Lambertian assumption in computer vision notwithstanding, materials in the real world rarely follow the convenient Lambertian reflection model. The reflectance of a wide variety of materials like plastics, paper, plant leaves, ceramics, fruits and vegetables, wood, leather, and human skin *etc.* can be approximated by the dichromatic reflectance model that models reflectance as a linear combination of a diffuse (Lambertian) term and a specular (glossy) term.

Chapter 2 of the thesis introduces a novel color space, that gives access to two purely diffuse channels of a color image of a dichromatic surface. These two diffuse channels preserve shading, and can potentially improve the performance of all vision algorithms that assume the scene to be Lambertian. We show results on photometric stereo, shape from shading, stereo, optical flow, and passive photometric stereo to demonstrate the usefulness of this novel color space transformation.

1.3.2 Complex Illumination and Photometric Invariants

In many real world scenarios, the illumination of a scene is complex. For example, in an office environment, there are two sources of illumination – the light bulb inside the

office, and the sunlight through the windows. In Chapter 3 of this thesis we address the problem of complex illumination. More specifically, we show that it is possible to transform a k channel image, into a $k - m$ channel illumination invariant image, that is independent of the effects of m color illuminants in a scene. We also generalize the concept of *hue* for the case of non-white illuminants.

1.3.3 Removing Specular Highlights

Using the transformation described in chapter 2, an RGB color image can transformed into another three channel image, two channels of which are diffuse, while the third channel contains the entire specular component, and an unknown amount of diffuse component. While this representation is useful for certain tasks, several other vision and graphics problems demand a complete separation of the diffuse and specular components of an RGB color image. For example, in the movie industry, it is common to use polarizers on the lights and the camera, to obtain diffuse texture maps of human faces.

In chapter 4 of this thesis, we propose a unified approach to remove specular highlights from images and videos. We show that under some mild assumptions, it is possible to evolve a partial differential equation (PDE) to iteratively remove specular highlights from a color image. Once separated, the diffuse and specular components can be independently processed and recombined to obtain dramatic visual effects.

1.3.4 Refractive Optical Flow

A class of materials that are rarely dealt with in computer vision are transparent materials. Part of the reason for this is that refraction is extremely difficult to handle, even though the laws governing refraction are pretty simple. Refraction through a material, governed by Snell's law, not only depends on refractive indices of materials, it also depends on the geometry of the refracting material, as well as the number of refractive surfaces involved.

In chapter 5, we generalize the optical flow equation to the case of refraction and reflection, and present a novel solution to the problem of *Environment Matting*. In a typical Environment Matting setup, a refractive object is matted into a new background. The background is assumed to be planar, and the objective is to find the path of light transport, as it travels from the background into the camera. We show that it is possible to use only the differential motion of a background plane to solve the problem.

1.4 Contributions

This thesis makes the following major contributions

- Introduces a novel color space transformation that gives access to two purely diffuse components of color images.
- Experimentally demonstrates that the two diffuse components can be used to improve the performance of a wide range of computer vision algorithms including photometric stereo, shape from shading, binocular stereo, optical flow, and passive photometric stereo.
- Extends the above transformation to handle complex illumination and experimentally demonstrates performance improvements of optical flow, stereo, and segmentation.
- Proposes a novel unified framework for solving the ill-posed problem of removing specular highlights for images and videos.
- Introduces the concept of *Dichromatic Editing* – the process of independently editing and then combining the two separated dichromatic layers to achieve dramatic visual effects.
- Proposes a novel solution to Environment Matting.
- Introduces a generalization of the optical flow equation – *Refractive Optical Flow* – valid in the presence of refraction (and other transformations).

1.5 Other Works

This section describes work undertaken during the course of my graduate study that is not part of this dissertation. In addition to the material covered in this dissertation, I have worked on two other problem areas.

1.5.1 3D Modeling and Animation of Human Faces

This work was motivated by the application of low bandwidth virtual Tele-presence using avatars. The idea was to generate a 3D model of a person's face and animate different emotions on the face at a remote observer's end, thus eliminating the need to send raw video. A color based algorithm was used to segment skin pixels, and features (corners)

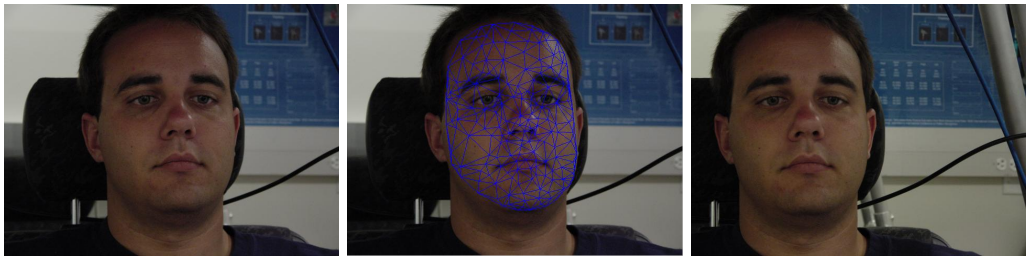


Figure 1.7 3D modeling: The left and right images serve as input to the algorithm. A 3D mesh is obtained by deforming a pre-defined 3D mesh according to depth estimates of feature points extracted from the input images. The image in the center shows the deformed 3D mesh projected back onto the image plane.

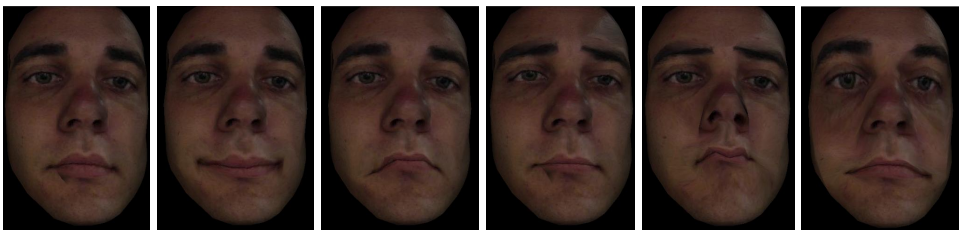


Figure 1.8 Animating emotions: Different emotions can be animated by warping the 3D mesh according to pre-defined transformations.

in skin regions were automatically detected and matched in the two views. The depth estimates of these feature points were used to warp a predefined 3D mesh model. See Fig. 1.7. Six different emotions were modeled by warping the 3D mesh according to pre-defined transformations. See Fig. 1.8.

1.5.2 Cryo-Electron Microscopy

Single particle reconstruction using Cryo-Electron Microscopy (cryo-EM) is an emerging technique in structural biology for estimating the 3D structure (density) of protein macromolecules. Unlike tomography where a large number of images of a specimen can be acquired, the number of images of an individual molecule (particle) is limited because of radiation damage. Instead, the specimen consists of identical copies of the same protein macro-molecule embedded in vitreous ice at random and unknown 3D orientations. Because the images are extremely noisy, thousands to hundreds-of-thousands of projections are needed to achieve the desired resolution of 5 Å. Along with differences of the imaging modality compared to photographs, single particle reconstruction provides a

unique set of challenges to existing computer vision algorithms. We briefly describe our contributions in areas of particle detection, contrast transfer function (CTF) estimation, and initial 3-D model construction.

Particle Detection

One of the first steps in single particle reconstruction using cryo-EM is locating protein macromolecules (particles) in acquired images. Since the reconstruction process typically requires a few tens of thousands of images, the task of manually locating particles in images is extremely tedious and time consuming. A new learning-based approach [68] was presented for particle detection using the Adaboost learning algorithm. The approach builds directly on the successful detectors developed for the domain of face detection [114]. It is a discriminative algorithm which learns important features of the particles appearance using a set of training examples of the particles and a set of images that do not contain particles. The algorithm is fast (10 seconds on a 1.3 GHz Pentium M processor), is generic, and is not limited to any particular shape or size of the particle to be detected. The method was evaluated on a publicly available dataset of 82 cryo-EM images of keyhole lympet hemocyanin (KLH). Figure 1.9 shows the result of detecting two different views of KLH on two typical images. From 998 automatically extracted particle images, the 3D structure of KLH has been reconstructed at a resolution of 23.2 Å which is the same resolution as obtained using particles manually selected by a trained user. In [128] , it was shown to be one of the most effective methods for particle detection in a benchmark dataset of KLH. See Fig. 1.10.

Automated CTF Estimation

A critical step in the processing and analysis of cryo-EM images involves the estimation of the factors that modulate the image and which must be corrected in order to generate an accurate 3D reconstruction of the specimen. Principal among these is the contrast transfer function (CTF) of the microscope. The effect of the CTF is to introduce spatial frequency dependent oscillations into the Fourier space representation of the image. The theory of contrast transfer in the electron microscope [34,35] provides a parametric form for the CTF, the envelope function, and the background noise. In our own work [67], we introduce a completely automated algorithm for estimating the parameters of the CTF of a transmission electron microscope. Once estimated, the micrographs can be corrected in order to generate an accurate 3-D reconstruction of the specimen.

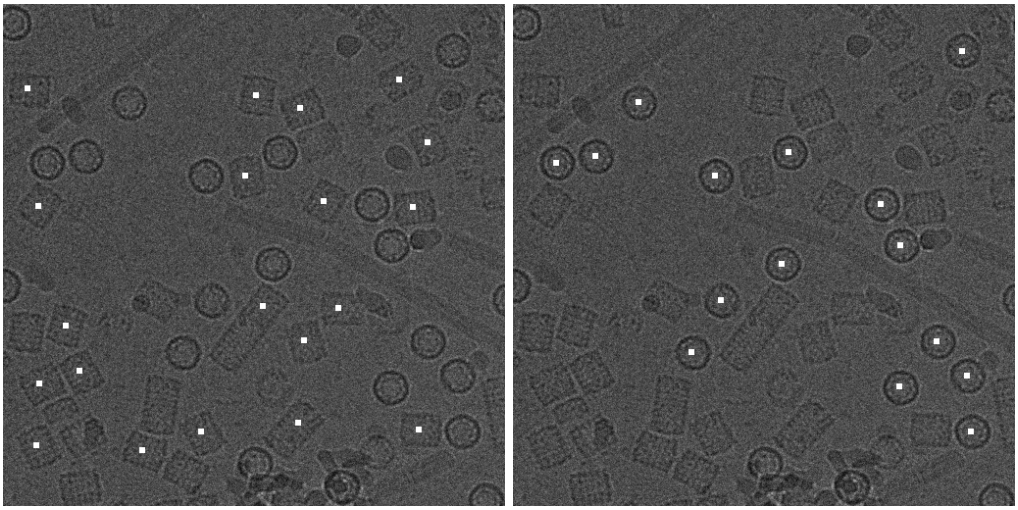


Figure 1.9 Particle detection: Rectangular and circular views of KLH detected in two typical micrographs. The locations of detected particles are shown using white marks.

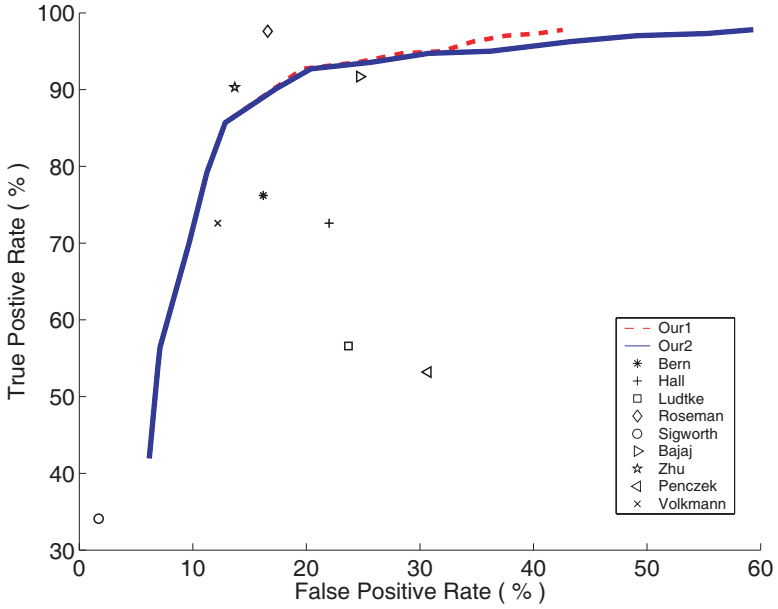


Figure 1.10 Receiver operating characteristic (ROC) of the our particle detector. Our1 is the ROC with circular particles removed and Our2 is the ROC obtained without removing the circular particles before detection of the rectangular particles.

Molecular Reconstruction

In [66] we address the problem of reconstructing the density of a scene from multiple projection images produced by modalities such as x-ray, electron microscopy, *etc.* where an image value is related to the integral of the scene density along a 3D line segment between a radiation source and a point on the image plane. While computed tomography (CT) addresses this problem when the absolute orientation of the image plane and radiation source directions are known, this work addresses the problem when the orientations are unknown – it is akin to the structure-from-motion (SFM) problem when the extrinsic camera parameters are unknown. We study the problem within the context of reconstructing the density of protein macro-molecules in cryo-EM, where images are very noisy and existing techniques use several thousands of images. In a non-degenerate configuration, the viewing planes corresponding to two projections intersect in a line in 3D. Using the geometry of the imaging setup, it is possible to determine the projections of this 3D line on the two image planes. In turn, the problem can be formulated as a type of orthographic structure from motion from line correspondences where the line correspondences between two views are unreliable due to image noise. We formulate the task as the problem of denoising a correspondence matrix and present a Bayesian solution to it. Subsequently, the absolute orientation of each projection is determined followed by density reconstruction. We show results on cryo-EM images of proteins and compare our results to that of Electron Micrograph Analysis (EMAN) – a widely used reconstruction tool used by the cryo-EM community. See Fig. 1.11.

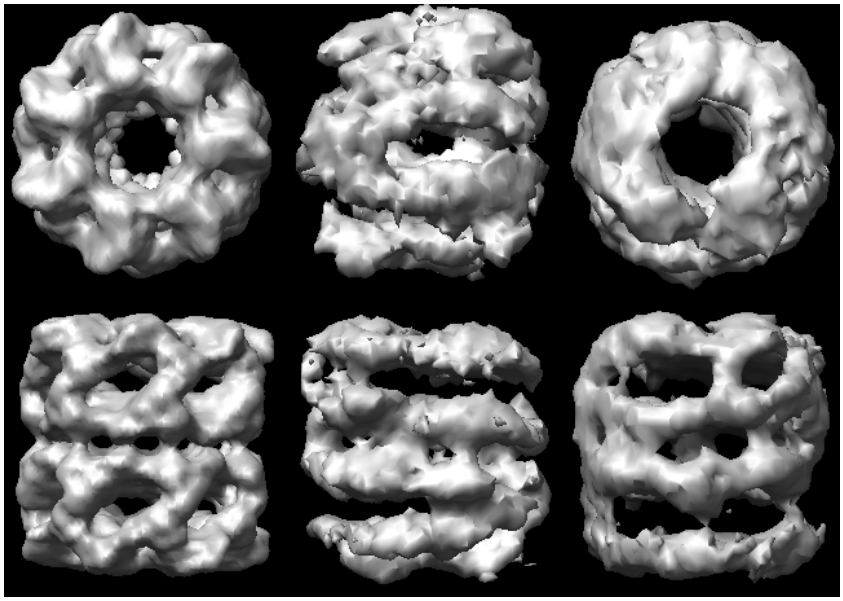


Figure 1.11 The first column shows the top and side views of a macro-molecule called GroEL produced from a 11.5 Å reconstruction [64] in a publicly available Molecular Structure Database. The middle column shows the initial model estimated using EMAN [63] – A widely used tool in cryo-EM. The right column shows the initial model estimated using our method. The same dataset was used to generate the two initial models.

2

The SUV Color Space

"The least questioned assumptions are often the most questionable."

Paul Broca

Scene analysis is complicated by the presence of highlights due to specular reflection, since these highlights are a complex function of viewpoint, illumination and surface microstructure. Typically, in order to obtain meaningful information about a scene from its images, complex reflection effects such as highlights are either roughly approximated or completely ignored by vision systems. For example, most stereo and structure-from-motion techniques ignore specular highlights altogether, and instead model surfaces as being Lambertian; even state-of-the-art recognition systems are based on relatively simple (e.g., spatially-invariant, parametric) models of reflectance. See, Ref. [19].

Even when the illumination of a scene can be controlled, the coupling of reflectance and shape makes the recovery of information (*i.e.,* 3D shape) a difficult task. Recently, a small number of surface reconstruction techniques have overcome these difficulties by effectively decoupling shape and reflectance in images. These techniques rely on the careful acquisition of images to exploit physical properties such as reflectance isotropy [61], reciprocity [65,130] and the constancy of radiance in free space [53,65]. By reducing or eliminating the restrictions on surface reflectance, these techniques provide accurate reconstructions for a much broader class of surfaces.

This chapter introduces a color space transformation as a means to remove specular reflection effects. This color space transformation allows us to distill the image, removing the specular effects, and leaving only the much simpler – ideally Lambertian – diffuse effects. (See Fig. 2.1.) The technique can be applied to any *dichromatic surface* (*i.e.,* a surface whose reflectance can be represented using Shafer's dichromatic

(a) (b)

Figure 2.1 Removal of specular reflectance effects. A rotation of RGB color space (a) provides an image that is void of specular reflection effects (b) and provides a more direct measurement of scene shape.

model [98]) for which the body spectral reflectance varies significantly over the visible spectrum (*i.e.,* not 'white'). The color transformation is applied to the problem of photometric stereo, shape from shading, binocular stereo, optical flow, and passive photometric stereo. We show that by eliminating specular effects, this transformation enables well-known Lambertian algorithms to be applied non-Lambertian surfaces.

2.1 A Data-dependent Color Space

In this section we introduce a color space transformation that gives access to two purely diffuse channels of a three channel color image. The two diffuse channels can be used to improve the performance of algorithms that make the Lambertian assumption.

Suppose we linearly transform RGB color space by rotating the coordinate axes, and as shown in the left of Fig. 2.2, suppose this rotation is such that one of the axes becomes aligned with the direction of the effective source color $\mathbf{S}$. This transformation defines a new color space, which we refer to as the SUV color space. It can be defined according to

$$\mathbf{I}_{SUV} = [R]\mathbf{I} \tag{2.1}$$

using any $[R] \in SO(3)$ that satisfies $\mathbf{S} = [R][1,\ 0,\ 0]^\top$. Here, we choose $[R] = [R_G(-\theta_\mathbf{S})][R_B(\phi_\mathbf{S})]^\top$ where $[R_k(\theta)]$ is a right-handed rotation about the k-axis by angle θ, and $(\theta_\mathbf{S}, \phi_\mathbf{S})$ are the elevation and azimuthal angles of the source vector $\mathbf{S}$ in the RGB

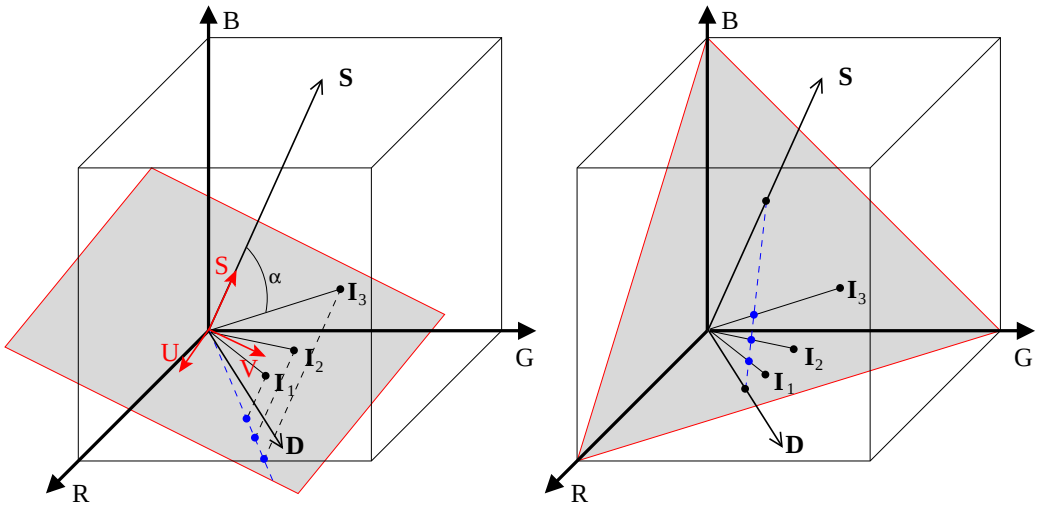


Figure 2.2 Linear and non-linear transformations of RGB color space. Three observations of the same material yield color vectors $\mathbf{I}_1 \ldots \mathbf{I}_3$ in the dichromatic plane spanned by the source and diffuse colors $\mathbf{S}$ and $\mathbf{D}$. Left: The proposed SUV color space is a rotation of the RGB space. One axis is aligned with the source color, and two of three transformed channels (UV) are independent of specular reflectance. Diffuse shading information is preserved and can be used to recover shape. Right: Central projection used to compute r-g chromaticity values and HSV-type color spaces does not preserve shading information.

coordinate system. From Eq. 1.12 it follows that

$$\mathbf{I}_{SUV} = \left(\bar{\mathbf{D}} f_d(\boldsymbol{\theta}) + \bar{\mathbf{S}} f_s(\boldsymbol{\theta})\right) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}, \tag{2.2}$$

where $\bar{\mathbf{D}} = [R]\mathbf{D}$ and $\bar{\mathbf{S}} = [R]\mathbf{S} = [1,\ 0,\ 0]^\top$. $\mathbf{D}$ and $\mathbf{S}$ are defined in Eq. 1.13.

This SUV space is a *data-dependent* color space because it depends on the effective source color in the image. It has two important properties. First, it separates the diffuse and specular reflection effects. The first channel (the S channel) encodes the entire specular component and an unknown fraction of the diffuse component, while the remaining two channels (U and V) are independent of $f_s(\boldsymbol{\theta})$ and are functions of only the diffuse reflectance.

The second important property is that, since the transformation is a rotation, shading information is preserved. This is clear from Eq. 2.2; if $\mathbf{r}_i^\top$ denotes the i^{th} row of $[R]$, the two diffuse channels are

$$\begin{aligned} I_U &= \mathbf{r}_2^\top \mathbf{D} f_d(\boldsymbol{\theta}) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}} \quad \text{and} \\ I_V &= \mathbf{r}_3^\top \mathbf{D} f_d(\boldsymbol{\theta}) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}. \end{aligned} \tag{2.3}$$

The S-channel is given by

$$I_S = \mathbf{r}_1^\top \mathbf{D} f_d(\boldsymbol{\theta}) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}} + f_s(\boldsymbol{\theta}) \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}. \tag{2.4}$$

It contains all of the specular component in addition to an unknown portion of the diffuse component. Also note that while the S-channel of an RGB color vector is positive, the U and V channels can be negative.

Assuming Lambertian diffuse reflectance, $f_d(\boldsymbol{\theta})$ is a constant function of $\boldsymbol{\theta}$. In this case, the two-channel color vector

$$\mathbf{J} = \left[I_U,\ I_V \right]^\top \tag{2.5}$$

provides direct information about the normal vector on the surface, with the coefficients $\rho_U = \mathbf{r}_2^\top \mathbf{D} f_d$ and $\rho_V = \mathbf{r}_3^\top \mathbf{D} f_d$ behaving as Lambertian albedos. Note that the norm of the two channel image $\mathbf{J}$ given by

$$\|\mathbf{J}\| = \left(I_U^2 + \ I_V^2 \right)^{\frac{1}{2}} = f_d \left((\mathbf{r}_1^\top \mathbf{D})^2 + (\mathbf{r}_1^\top \mathbf{D})^2 \right)^{\frac{1}{2}} \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}. \tag{2.6}$$

is also purely diffuse and preserves shading information. In addition, since it contains information from both diffuse channels, it has improved signal to noise ratio, and is therefore suitable for use in vision algorithms that use grayscale images as input.

Figure 2.2 compares this linear, data-dependent color space with conventional non-linear color spaces. Non-linear color spaces such as r-g chromaticity and hue-saturation-value (HSV) are computed by central projection. Each RGB pixel corresponds to a vector in the RGB cube, and these are intersected with the plane $R + G + B = c$ for some c. For example, hue and saturation correspond to the distance and polar angle of these intersection points relative to the cube diagonal, and chromaticity coordinates are derived from the intersection of these color vectors with the plane $R + G + B = 1$. Non-linear color spaces such as these are useful for recognition, for example, since they remove Lambertian shading and shadow information. (All positive scalar multiples of $\mathbf{I}$ map to the same chromaticity coordinates and the same hue.) Since they do not preserve photometric information, however, they are generally unsuitable for recovering shape.

In contrast, the SUV color space *does* preserve shading information, and by providing two channels that are independent of the complex, specular component of reflectance, it can be a useful tool for scene analysis.

In the last few years there has been a burst of activity in defining color space

transformations that exploit knowledge of the illuminant color to provide more direct access to the diffuse information in an image. While motivated by different applications, the transformations discussed below all share the same idea of linearly combining the three color channels of an RGB image to obtain one or two "diffuse" channels.

Tan and Ikeuchi [102] obtain a one-channel diffuse image through the transformation

$$I_d = \frac{3\max_k(I_k/S_k) - \sum_k(I_k/S_k)}{3\tilde{\lambda} - 1}, \tag{2.7}$$

where $k \in \{1, 2, 3\}$, and the bounded quantity $1/3 < \tilde{\lambda} \leq 1$ is chosen arbitrarily. This transformation yields a positive monochromatic diffuse image, which can be seen by expanding Eq. 2.7 using Eq. 1.12 and assuming (for argument's sake) that $I_1/S_1 > \max(I_2/S_2,\ I_3/S_3)$. In this case,

$$I_d = \frac{2I_1/S_1 - I_2/S_2 - I_3/S_3}{3\tilde{\lambda} - 1} = \frac{(2D_1/S_1 - D_2/S_2 - D_3/S_3)\, f_d \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}}{3\tilde{\lambda} - 1}. \tag{2.8}$$

Since this expression is independent of f_s and is directly related to $\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}$, the positive image I_d is specular-free and depends directly on diffuse shading information.

An alternative transformation is proposed by Park [82], who isolates *two* predominantly diffuse channels while retaining a similarity to HSI color space. The transformation is composed of a linear transformation $\mathbf{L}_p$ and rotation $\mathbf{R}_p$, and is written

$$\mathbf{I}_p = \mathbf{R}_p \mathbf{L}_p \mathbf{I}, \quad \text{with} \quad \mathbf{R}_p \mathbf{L}_p \mathbf{S} = [\ 0\ 0\ 2\ ]^\top. \tag{2.9}$$

The matrices $\mathbf{R}_p$ and $\mathbf{L}_p$ are chosen such that the third color axis is aligned with the illumination color. As a result, that channel contains the majority of the specular component, leaving the other two channels to be predominantly diffuse.

Similar to Park's transformation, one of the transformed axes in SUV space is aligned with the illuminant color. Unlike Park's transformation, however, this channel includes *the complete* specular component, leaving the remaining two channels to be *purely diffuse*.

Each of the three transformations described in this section exploits knowledge of the illuminant to provide a partial dichromatic separation. Unlike Eq. 2.7, the SUV color space is obtained by a linear transformation that yields two "diffuse" channels, and unlike Eq. 2.9, these two "diffuse" channels are in fact completely free of specularity.

2.1.1 Practical Considerations

For shape recovery we are generally interested in the two-channel diffuse signal **J** in Eq. 2.5. The quality of this signal depends on the spectral characteristics of the surface in addition to sensor noise.

Spectral Reflectance

When the surface is 'white', the spectral reflectance is a constant function of wavelength. In this case, $g_d(\lambda) = g_d$, and since

$$D_k = g_d \int C_k(\lambda) L(\lambda) d\lambda = g_d S_k,$$

it follows that the color vector **I** and the source color **S** are collinear in color space. For these surfaces, the UV channels of the image of the scene point are zero; and as a result, they provide no information about the surface, regardless of the illuminant and sensors that we choose. (This is the same restriction experienced by Klinker *et al.* [51]; when the diffuse surface color and source color are the same, there is no way to distinguish between the two reflection components.) In order to recover surface shape for these materials without restricting the BRDF, alternative (and more complex) reconstruction methods can be used. See, for example, Refs. [53, 130].

Sensor Noise

Assuming independent, additive Gaussian noise with zero mean and variance σ^2 in each of the three channels of color vector **I**, and assuming $\|\mathbf{I}\| \leq 1$, the signal-to-noise ratio (denoted SNR(**I**)) is $10 \log_{10}(1/\sigma)$ dB. The magnitude of the diffuse color vector **J** is related to that of the original color vector by $\|\mathbf{J}\| = \|\mathbf{I}\| \sin\alpha$, and since the noise is the same in both cases, it follows that

$$\mathrm{SNR}(\mathbf{J}) = \mathrm{SNR}(\mathbf{I}) + 10 \log_{10}(\sin\alpha). \tag{2.10}$$

This relationship is shown in Fig. 2.3, and it suggests that when the angle between the image and the source color is less than 10°, the two-channel diffuse signal suffers severe degradation. The effects of this degradation can be mitigated by using multiple exposures to collect high dynamic range (HDR) images. When exposures are chosen judiciously [33], we can increase the SNR of the original image and therefore that of the two-channel diffuse image as well.

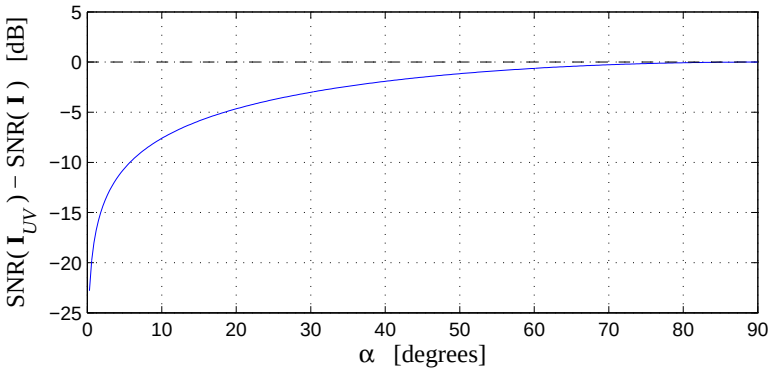


Figure 2.3 The signal-to-noise ratio (SNR) of the two-channel diffuse image ($\mathbf{J}$) relative to that of the original image ($\mathbf{I}$) as a function of α, the angle between $\mathbf{I}$ and the source color $\mathbf{S}$ in RGB color space. (See Fig. 2.2.)

2.1.2 Computing Diffuse Images

In order to compute the two-channel diffuse image, it is necessary to know the effective source color, $\mathbf{S}$. Under controlled conditions, the source color can be calibrated, and in an uncontrolled setting it can be estimated using a number of established techniques.

Tominga and Wandell [106] demonstrate a method for estimating the source color based on the fact that color vectors from a homogeneous surface span the dichromatic plane. They determine the source color by intersecting multiple dichromatic planes (from different dichromatic surfaces) in color space. Similarly, Lee [55] finds the source color by intersecting lines in 2D chromaticity space. More recently, Finlayson and Schaefer [30] use the fact that the chromaticity of the majority of illuminants lie along a known curve in chromaticity space and recover the source color from an image of a single homogeneous dichromatic surface.

Figure 2.1 shows a diffuse image that was computed using the source color determined by intersecting lines in chromaticity space [55]. (Three homogeneous surfaces were manually segmented for this purpose.) The monochrome diffuse image was created by transforming the input image to the SUV color space and computing $\|\mathbf{J}\|$ at each pixel. Comparing this to the original image, we see that the specular effects have been removed as expected. Note that the dichromatic model is violated when saturation occurs in one or more color channels of the input images, and this causes errors at points of extreme specularity. This is seen in Fig. 2.1 as violations at saturated pixels of the image are readily identified. Additional examples computed with calibrated sources are shown in Sect. 2.2.1.

2.2 Applications

2.2.1 Photometric Stereo

The two-channel diffuse image of the previous section is derived without making any assumptions about the nature of specular reflectance, and yet it is independent of specular reflection effects for dichromatic surfaces. By assuming that the diffuse component of the reflectance is Lambertian, we can use these images along with established techniques for photometric stereo to obtain estimates of surface shape.

Assuming the light source color is known, the color vector $\mathbf{I}$ at each pixel can be transformed to SUV space using Eq. 2.2. The two diffuse UV channels preserve shading information, and the two-channel image $\mathbf{J}$ is independent of the specular component. Assuming that the diffuse reflection component is Lambertian, standard photometric stereo techniques can be used for surface reconstruction. Here, we discuss a modified version[1] of the color photometric stereo method of Barsky and Petrou [12].

Let $\mathbf{J}^1$, $\mathbf{J}^2$ and $\mathbf{J}^3$ be three 2-channel color vectors produced by observing a single point under three different light source directions $\hat{\mathbf{l}}^1$, $\hat{\mathbf{l}}^2$ and $\hat{\mathbf{l}}^3$. As in Eq. 2.3, it follows that

$$\mathbf{J}^k = \left[I_U^k, I_V^k\right]^\top = (\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}^k)\boldsymbol{\rho}, \tag{2.11}$$

where $\boldsymbol{\rho}$ is a 2-channel UV albedo. A *shading vector* is defined according to $\mathbf{F} = \left[f^1, f^2, f^3\right]^\top = [\hat{\mathbf{l}}^1 \;\; \hat{\mathbf{l}}^2 \;\; \hat{\mathbf{l}}^3]^\top \hat{\mathbf{n}}$, and the shading information from all three observations is combined in an *intensity matrix* that satisfies

$$[J] = \begin{bmatrix} J_1^1 & J_2^1 \\ J_1^2 & J_2^2 \\ J_1^3 & J_2^3 \end{bmatrix} = \begin{bmatrix} f^1\rho_U & f^1\rho_V \\ f^2\rho_U & f^2\rho_V \\ f^3\rho_U & f^3\rho_V \end{bmatrix} = \mathbf{F}\boldsymbol{\rho}^\top. \tag{2.12}$$

The least squares estimate of the shading vector $\mathbf{F}$ can easily be computed from the intensity matrix; it is the principal eigenvector of $[J][J]^\top$. Once the shading vector is known, the surface normal is found by solving the matrix equation $\mathbf{F} = [\hat{\mathbf{l}}^1 \;\; \hat{\mathbf{l}}^2 \;\; \hat{\mathbf{l}}^3]^\top \hat{\mathbf{n}}$. Once the normals are estimated for every pixel in the image, the normal field is integrated to produce a surface. In the implementation here, we use an iterative least squares

[1] The method was originally designed for color images with three channels. Here we use the same method for two-channel images.

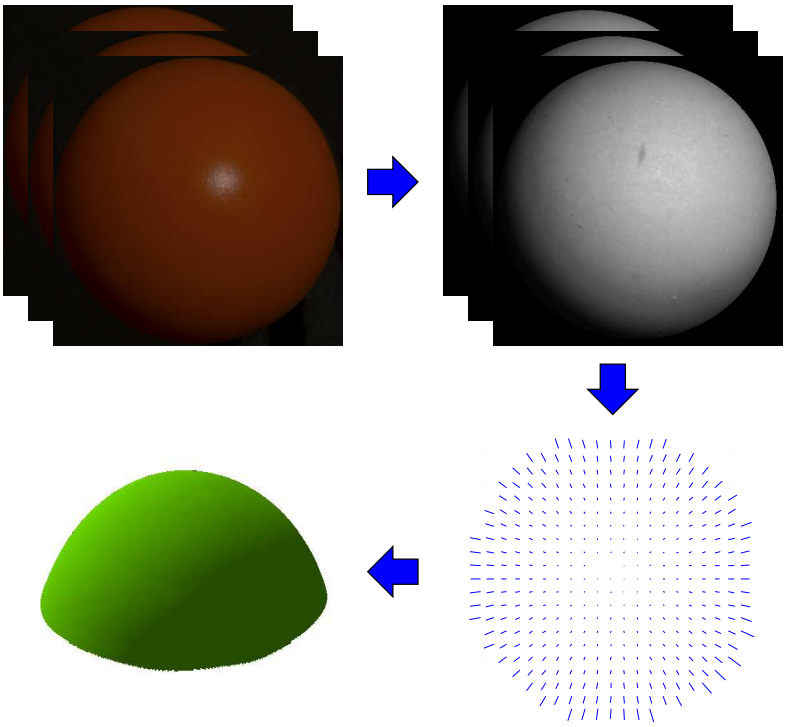


Figure 2.4 Photometric stereo using SUV color space. Three or more RGB images are acquired under known illumination conditions, and these are transformed to SUV space using the known source color. The UV channels represent diffuse images of the object, and these are used with standard photometric stereo techniques to estimate the surface normal at each pixel. The normals are integrated to recover the surface.

optimization that minimizes the re-construction error [40].

This reconstruction procedure is outlined in Fig. 2.4, and it can be applied without change to any number of images larger than three.

Experimental Results

To validate the method quantitatively, we used objects of known shape with varying material properties. The objects are shown in Fig. 2.7, and they consist of a set of spheres with increasing specularity. The incident-plane BRDFs are shown in Fig. 2.5.

For each sphere, a set of four high dynamic range (HDR) images were captured from a fixed viewpoint and four known illumination directions. The source color was calibrated by imaging a Macbeth color checker, and it was used to compute the SUV images as described in Sect. 2.1. The second column of Fig. 2.7 confirms that the UV channels of these images depend largely on the diffuse reflectance. These monochrome diffuse images show the magnitude of the two diffuse channels, computed using $\|\mathbf{J}\| = (I_U^2 + I_V^2)^{\frac{1}{2}}$ at each pixel.

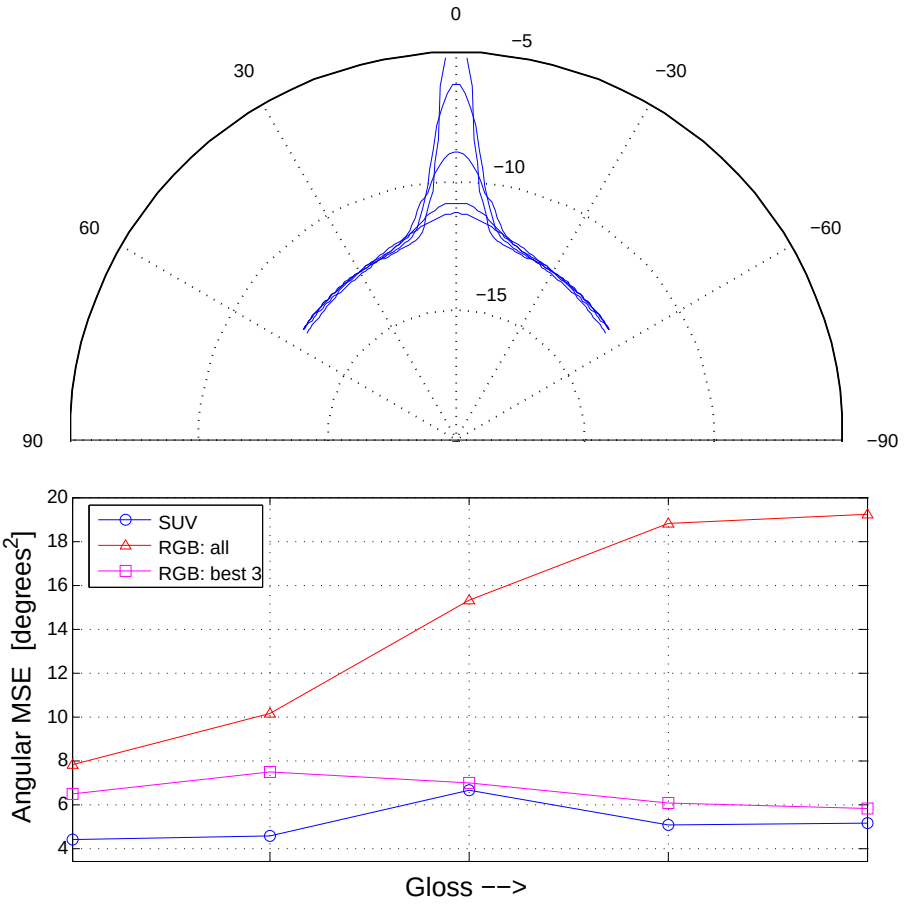


Figure 2.5 Comparison of photometric stereo using two different color spaces. Top: Relative BRDFs (in decibels) of the five red spheres of Fig. 2.7 as a function of half-angle. Bottom: Mean-square angular error in the recovered surface normals as a function of increasing specularity using both the SUV color space and existing RGB methods.

Using the diffuse UV images, the surface normals of each sphere were estimated using the photometric stereo method described in Sect. 2.2.1. As a means of comparison, we implemented two different RGB-based photometric techniques. The first method uses all four RGB images and assumes Lambertian reflectance [12]. The second method assumes Lambertian+specular reflectance and reconstructs the surface by choosing the three 'least specular' RGB measurements at each pixel [13, 26].

The results are shown in Figs. 2.5 and 2.7. The recovered surfaces, including cross-sections overlaid on the true shape, are displayed in Fig. 2.7. More quantitative results are shown in Fig. 2.5, with the bottom of that figure displaying the angular difference between the true and estimated surface normals as a function of increasing specularity. (This plot shows the mean-square error computed over a window encom-

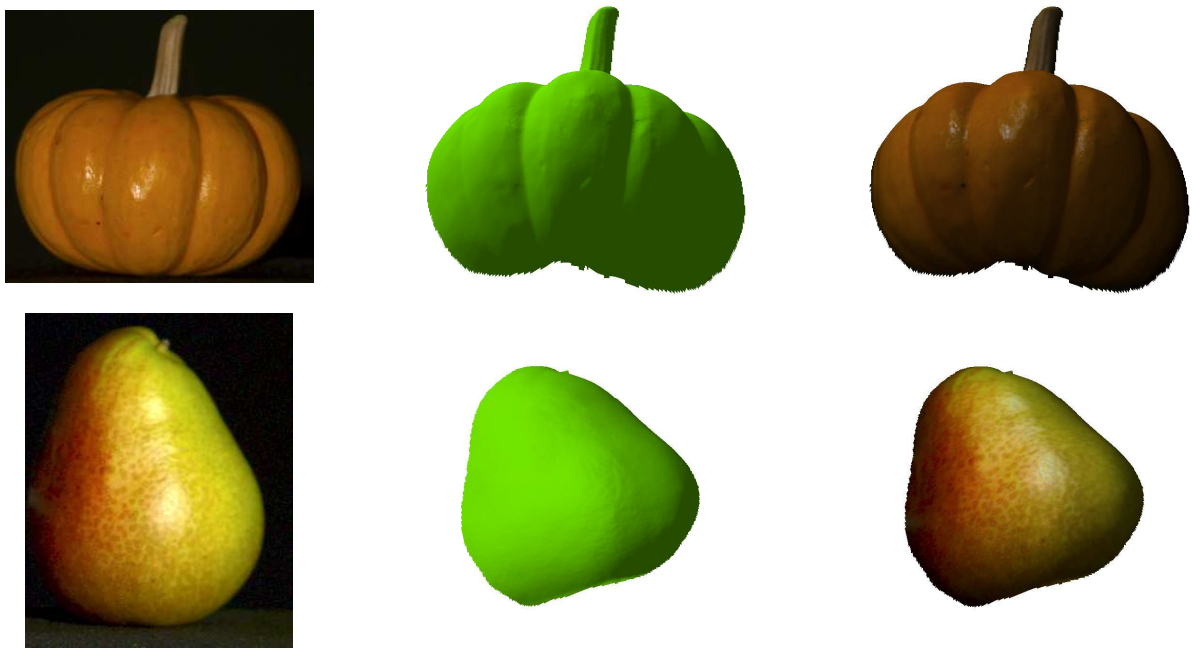


Figure 2.6 SUV-based photometric stereo applied to natural surfaces. Left: Input RGB images show significant specular reflectance and texture. By transforming the images to SUV space, the specular effects are removed, enabling accurate recovery of shape. Middle, Right: The surfaces recovered by integrating the estimated surface normals.

passing the visible specularities in the four images.) These results demonstrate that the SUV-based reconstruction is largely independent of the specular reflectance, whereas both the four-image and three-image RGB methods are affected by it. The four-image method assumes Lambertian reflectance and its performance degrades monotonically as gloss increases; and while the three-image RGB method performs well for the high-gloss (narrow specular lobe) spheres, it performs less well when the angular support of the specular lobe is large relative to the separation of the light source directions.

Figure 2.6 shows the results of applying our SUV-based photometric stereo method to two natural objects (a pear and a pumpkin.) For each object, four HDR images were captured and transformed to SUV color space, and the two-channel diffuse images were used to estimate the surface normals. For each object, we show an input image and the surfaces recovered by integrating these surface normals. Since the transformation to SUV space is purely local, the method requires no spatial coherence in the image, and it performs well for surfaces with arbitrary texture as shown by the pear example. This is not true for alternative photometric stereo techniques that use color-based highlight removal [51, 96], since these methods generally require spatial coherence.

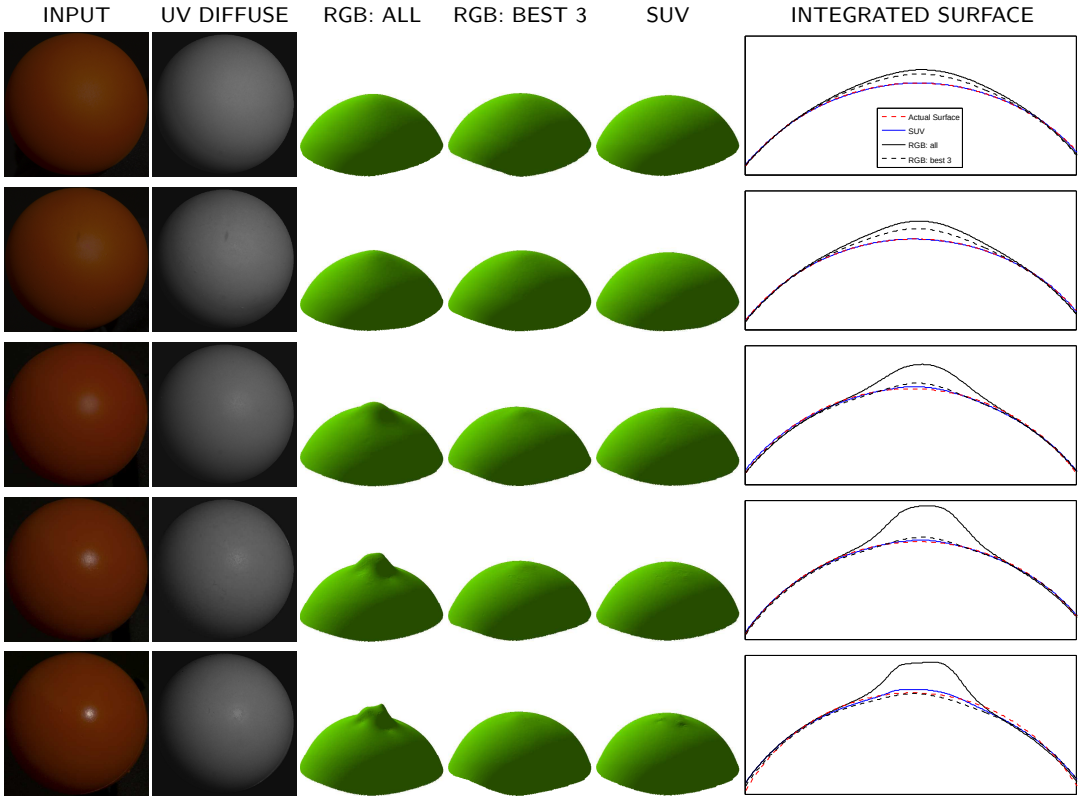


Figure 2.7 Comparison of photometric stereo methods using two different color spaces. Five red spheres with increasing specular reflectance are each observed under four illumination directions, and these images are used to recover the surface. From left to right, each row shows: i) an input RGB image, ii) the corresponding diffuse image given by the linear transformation to SUV space, iii) surfaces integrated from the surface normals estimated by three photometric stereo methods, and iv) cross-sections of the surfaces overlaid on the true shape.

2.2.2 Shape from Shading

Shape from shading algorithms use a single grayscale image to recover the 3D shape of an object. As mentioned earlier, the norm $\|\mathbf{J}\|$ of the 2-channel diffuse image $\mathbf{J}$ represents a surface with an effective albedo $f_d \left((\mathbf{r}_1^\top \mathbf{D})^2 + (\mathbf{r}_1^\top \mathbf{D})^2 \right)^{\frac{1}{2}}$, and it suggests that the specular invariant can be used directly as input to Lambertian-based shape from shading algorithms.

The benefit of this approach is demonstrated in Fig. 2.8, where we assess the performance of Zheng and Chellappa's shape from shading algorithm [127] for both a conventional grayscale image I_g and a single-color invariant image $\|J\|$. The top row of the figure shows grayscale and specular invariant images computed from an RGB image of a pear, and the middle row shows the surfaces that are recovered by applying the

same algorithm in the two cases. The solid blue profile in the bottom graph shows that specular reflections cause severe artifacts when the algorithm is applied to the grayscale image. In contrast, as shown by the dashed red profile, one can obtain vastly improved results using the same algorithm by computing the specular invariant as a pre-processing step.

2.2.3 Stereo

The vast majority of stereo algorithms are based either explicitly or implicitly on the assumption that surfaces are Lambertian. Since specular reflections violate this assumption, stereo reconstructions of specular surfaces are often inaccurate. In cases of significant specular reflections, we can improve the accuracy of existing stereo algorithms by computing the specular invariant of Eq. 2.6 as a pre-process. This is demonstrated by Fig. 2.9, which compares binocular stereo results obtained using conventional grayscale images to those obtained using specular invariant images. The grayscale and specular invariant images are computed from a rectified RGB stereo pair (top of Fig. 2.9), and they are used as input to the binocular stereo algorithms of Birchfield and Tomasi [16] and Boykov *et al.* [22]. As seen by the noisy disparity maps on the left of Fig. 2.9, these stereo algorithms break down in the presence of strong specularities. When the same algorithms are applied to specular invariant images, however, specular reflections are effectively ignored, and the results are greatly improved.

2.2.4 Optical Flow

Motion estimation through the computation of optical flow is another example of an application that can benefit from specular invariance. Recovering dense optical flow relies on the constant-brightness assumption, which is violated when an observer moves relative to a static, specular scene. As demonstrated by the results in Fig. 2.10, optical flow in the presence of specular reflections in a complex illumination environment can be improved by computing a specular invariant as a pre-processing step.

In Fig. 2.10, an RGB image sequence is captured by a camera translating horizontally relative to a static scene. The sequence is used to compute a conventional grayscale sequence $I_g(t)$ and a specular invariant sequence computed using Eq. 2.6. These two videos are used as input to Black and Anandan's algorithm for robust optical flow [17]. The left of Fig. 2.10 shows a single color image from the sequence, and the right shows the recovered flows in the indicated window. Since the camera undergoes pure

translation, the 'ground-truth' flow lies along parallel horizontal lines. Incorrect flow (shown in red) is recovered using the conventional grayscale image; and as expected, the flow is severely corrupted by specular highlights. In contrast, the flow computed from the specular invariant image (shown in blue) is close to the ground truth and is largely unaffected by these non-Lambertian effects.

2.2.5 Photometric/Geometric Reconstruction

More generally, the specular invariant can be used to improve the performance of a broad class of Lambertian-based reconstruction systems in the presence of specular, non-Lambertian surfaces. This includes, for example, methods that combine both geometric and photometric constraints to obtain accurate surface shape [47,58,125]. To provide an example, we use the passive photometric stereo algorithm described by Lim *et al.* [58]. This method begins with an approximate, piece-wise planar reconstruction obtained by tracking a small number of features across a video sequence under (possibly varying) directional illumination. Then, an iterative method based on uncalibrated Lambertian photometric stereo simultaneously refines the reconstruction and estimates the unknown illumination directions.

Figure 2.11 compares the results obtained from an image sequence that consists of a moderately specular cylinder moving under fixed illumination and viewpoint. The shape is estimated by applying the same algorithm to both the conventional grayscale sequence ($I_g(t)$) and the specular invariant sequence ($J_{\text{inv}(2)}$) computed from the same RGB data. The right-most surface in Fig. 2.11 shows that the reconstruction obtained using the specular invariant is nearly cylindrical, while that computed from the conventional grayscale sequence is severely corrupted by specular reflections.

2.3 Conclusion

This chapter presents a data-dependent rotation of RGB color space that separates the diffuse reflection effects from the more complex, specular reflection effects in images of dichromatic surfaces. Since it is linear, this transformation preserves diffuse shading information, and images in this transformed space can be used to obtain photometric reconstructions that are independent of the specular reflectance. The method is evaluated both qualitatively and quantitatively, on a variety of vision algorithms, and it is shown to perform well for both painted surfaces with varying specular reflectance and natural surfaces with and without texture.

This chapter is based on "Beyond Lambert: Reconstructing Specular Surfaces Using Color" by S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. I developed the theory that a linear combination of color channels can be used to obtain diffuse channels and was responsible for implementation and experiments described in the paper. I also contributed toward writing the paper.

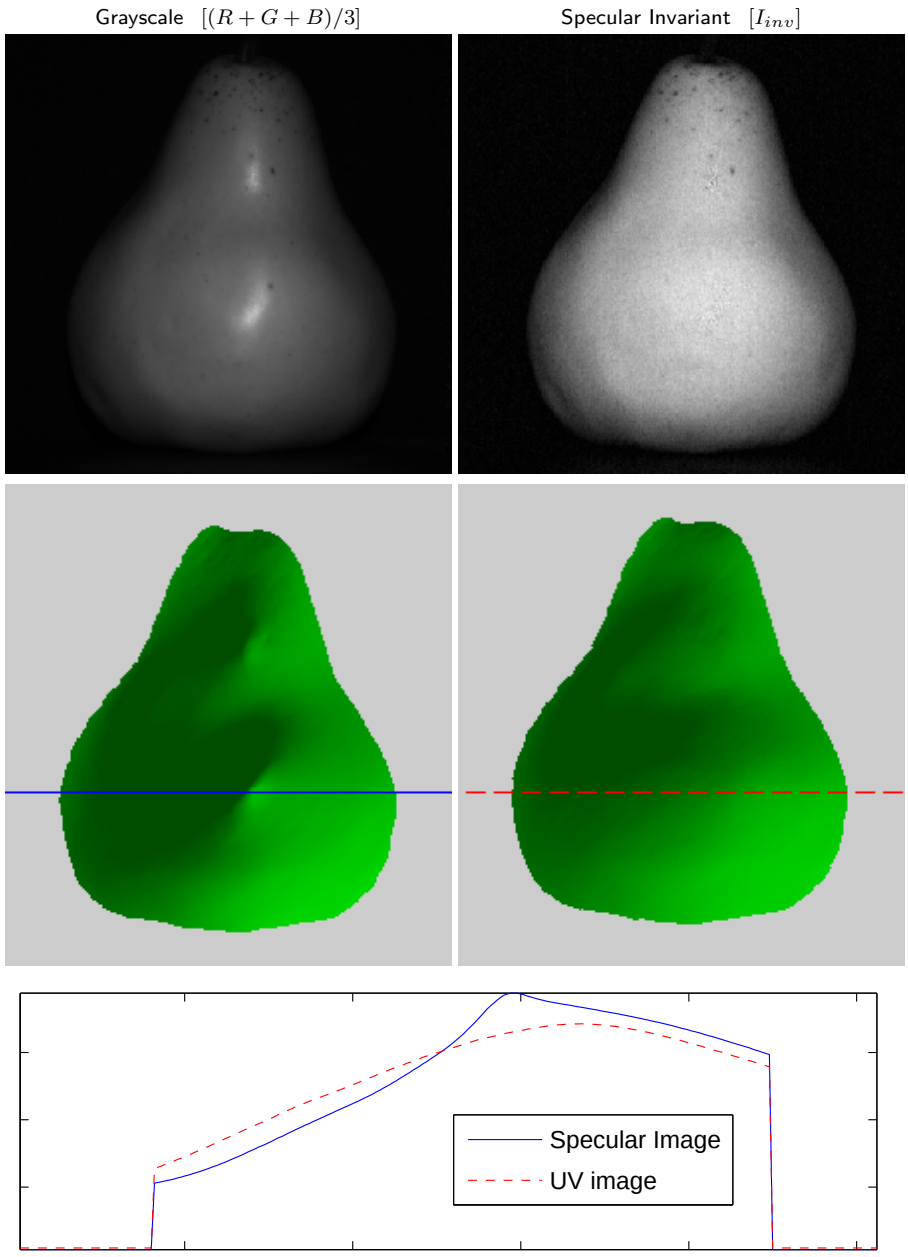


Figure 2.8 Shape from shading comparison. An RGB image of a pear is used to compute conventional grayscale (top-left) and specular invariant (top-right) images, and these are input to Zheng and Chellappa's shape from shading algorithm [127]. Middle row: surfaces recovered in both cases. Bottom row: cross-sections of the recovered surfaces along the indicated horizontal lines.

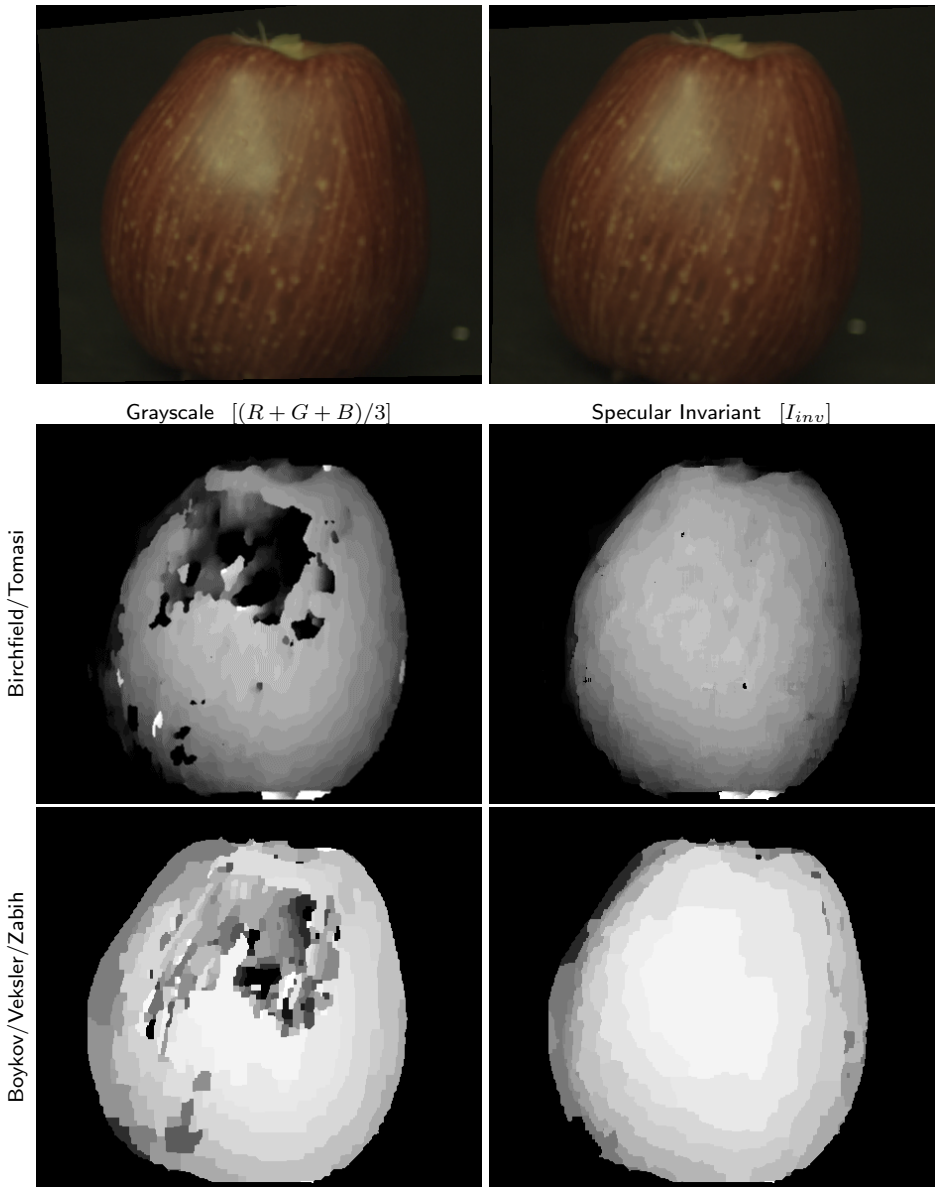


Figure 2.9 Stereo comparison. Both conventional grayscale images and specular invariant images (Eq. 2.6) are computed from a rectified stereo pair (top) and these are used as input to existing binocular stereo algorithms. Middle row: disparity maps obtained from the grayscale (left) and specular invariant (right) images using the method of Birchfield and Tomasi [16]. Bottom row: those obtained using the method of Boykov, Veksler and Zabih [22].

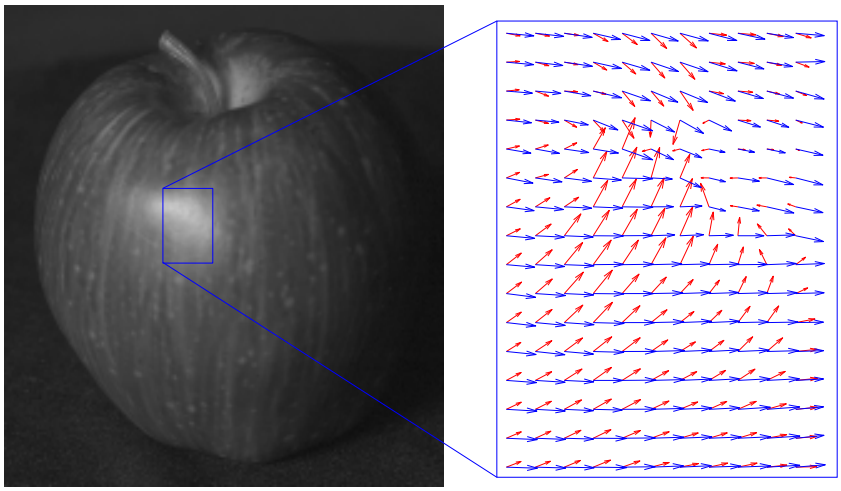


Figure 2.10 Optical flow comparison. An RGB image sequence is captured by a camera translating left relative to a specular apple. Both conventional grayscale and specular invariant images are computed from this RGB sequence, and these are used as input to Black and Anandan's robust optical flow algorithm [17]. Left: Single frame from the grayscale sequence. Right: flows obtained in the two cases. Red flow is computed from the grayscale sequence and is severely corrupted by specular reflection. Blue flow is computed from the specular invariant sequence and is much closer to ground truth, which is horizontal and to the right.

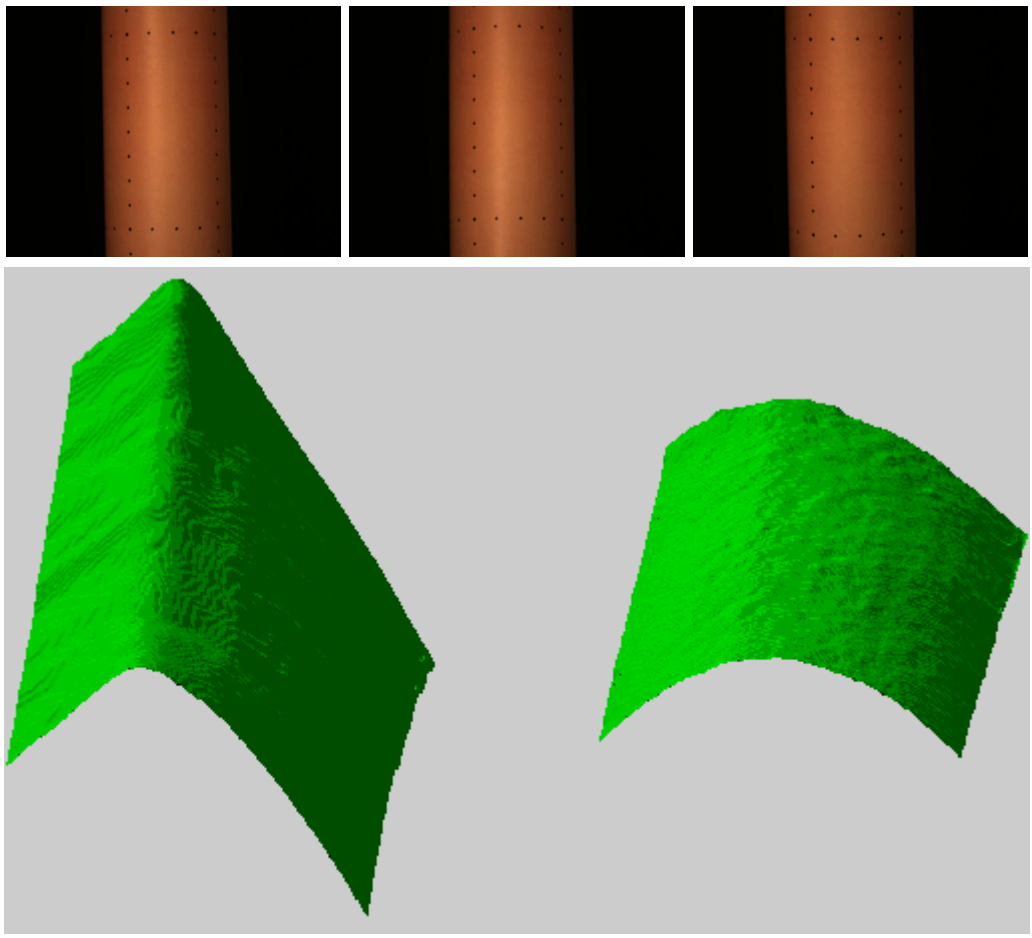


Figure 2.11 Comparison of shape from combined photometric and geometric constraints. (Top) Three RGB frames of a specular cylinder moving under fixed view and illumination. (Botton) Result of simultaneous tracking and photometric reconstruction (as described by Lim *et al.* [58]) using both the conventional grayscale (left) and specular invariant (right) sequences.

3

Complex Illumination

"Sometimes the questions are complicated and the answers are simple."

Theodor Seuss Geisel

An image is the product of the shape, reflectance and illumination in a scene. For many visual tasks, we require only a subset of this information, and we wish to extract it in a manner that is insensitive to variations in the remaining 'confounding' scene properties. For 3D reconstruction, for example, we seek accurate estimates of shape, and we design systems that are insensitive to variations in reflectance and illumination.

One practical approach to these problems is to compute a function of the input images that is invariant to confounding scene properties but is discriminative with respect to desired scene information. Such functions yield so-called *invariants*, and a number of examples are described in the literature. Perhaps the simplest example for a Lambertian scene is a normalized-RGB image. The normalized RGB color vector at each pixel depends on the spectral reflectance of the corresponding surface patch but not its orientation, which makes it useful for material-based segmentation.

Like normalized-RGB, many existing invariants seek to isolate information about the material properties in a scene and are therefore designed to be invariant to local illumination and viewing geometry. In contrast, we consider a class of invariants that deliberately preserve geometry information in a way that is invariant to specular reflections. These invariants give direct access to surface shape in the form of diffuse shading effects, and since diffuse shading is often well approximated by the Lambertian model, they satisfy the 'constant-brightness assumption' underlying most approaches to stereo reconstruction and structure-from-motion. In addition, these invariants provide access to surface normal information, which can be recovered using Lambertian-based photometric reconstruction methods.

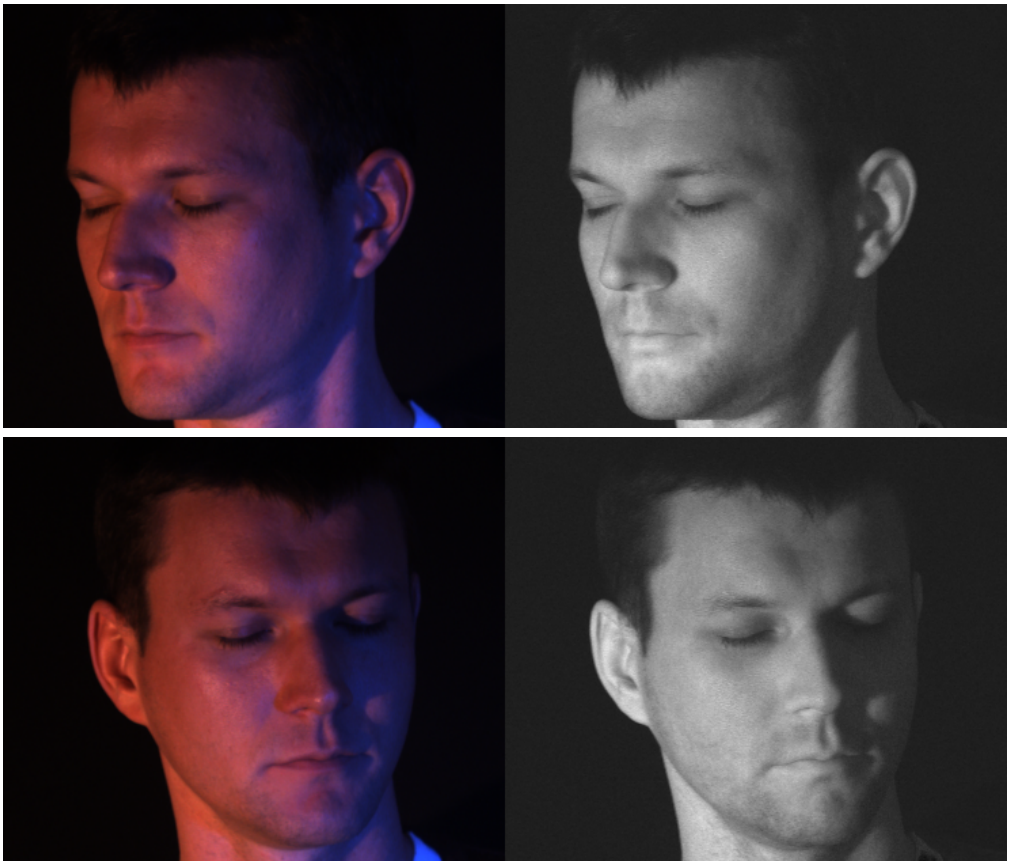


Figure 3.1 Left: Two frames of an RGB video of a scene with mixed illumination. A blue light on the right and a yellow light on the left induce complex specular effects. Right: Projecting these images onto the one-dimensional subspace orthogonal to the source color vectors in RGB space, yields an invariant to specular reflections that preserves diffuse shading and spectral reflectance information. (The complete video is included as supplemental material.)

In the previous chapter we showed that when surfaces are well-described by the dichromatic model [98], a specular-free image can be computed by projecting the RGB color vector at each image point onto the two-dimensional subspace orthogonal to the illuminant color. Inspired by these results, we:

1. Derive a general class of specular invariants based on color subspaces. These invariants can be applied to dichromatic surfaces under *mixed* illumination environments with multiple colored lights (see Fig. 3.1.)

2. Derive a second class of functions that, in addition to specular reflections, are invariant to diffuse shading. They depend only on the spectral reflectance of a dichromatic surface.

3. Demonstrate that these invariants can enhance many existing Lambertian-based vision techniques, vastly expanding their domain of applicability.

3.1 Related Work

A number of photometric invariants have been proposed for Lambertian scenes without specularity. Normalized-RGB, r-g chromaticity, and hue/saturation images are all examples of representations that are independent of 'diffuse shading' (the geometric relation between a surface normal and the illumination direction) and depend only on the spectral reflectance of the surface and the spectral power distribution (SPD) of the illuminant. Additional invariants to either local geometry or spectral reflectance can be computed when multiple images of a scene are available (*e.g.*, [119]), or when the reflectance of the surface is spatially coherent (*e.g.*. [77]); and an invariant to both local geometry and illuminant SPD can be computed from a single image under appropriate imaging conditions [39].

Invariants for scenes with more general reflectance functions are developed by Narasimhan *et al.* [74]. They describe a general model of reflectance consisting of a product of a "material" term (Lambertian albedo, Fresnel coefficient, etc.) and a "geometry" term that encodes the relationship between the surface normal, light-source, and viewing direction. Invariants to both of these terms can be computed from either multiple observations of a single point under variable view or illumination, or from one observation of a spatially-coherent scene. The geometry invariant is of particular interest, since it can be used directly for material-based segmentation [74].

3.1.1 Invariants for Dichromatic Surfaces

A substantial body of work is devoted to exploiting the dichromatic model of reflection [98] in order to separate diffuse and specular reflection components, which are independent of specular and diffuse reflection effects, respectively. According to the dichromatic model, the observation of a surface point is written

$$I_k = \sigma_d D_k + \sigma_s S_k, \tag{3.1}$$

where σ_d and σ_s are geometric scale factors that depend on surface shape and material properties and

$$D_k = \int E(\lambda)R(\lambda)C_k(\lambda)d\lambda \tag{3.2}$$

$$S_k = \int E(\lambda)C_k(\lambda)d\lambda. \tag{3.3}$$

Here, $E(\lambda)$ is the SPD of the incident illumination, $R(\lambda)$ is the spectral reflectance of the surface, and $C_k(\lambda)$ is the spectral sensitivity of a linear sensor. A typical RGB camera yields three such observations, and in this case we write $\mathbf{I}_{RGB} = \{I_k\}_{k=R,G,B}$ and define $\mathbf{D} = \{D_k\}_{k=R,G,B}$ and $\mathbf{S} = \{S_k\}_{k=R,G,B}$ to be the *diffuse color* and *specular color*, respectively. These are conventionally assumed to be vectors of unit length.

There is practical utility in separating the diffuse and specular components in an image. Since diffuse reflections are typically well-represented by the Lambertian model, this separation allows the application of powerful Lambertian-based vision algorithms to a broad class of non-Lambertian scenes. Unfortunately, computing such a separation is ill-posed. It traditionally requires additional constraints such as texture-less surfaces [51], knowledge about (e.g., segmentation of) diffuse colors, constraints on the neighborhood of a pixel [69, 102] or specific parametric models for specular reflectance [91].

When the source color is known and constant over a scene, one can compute invariants to specular reflections that are based on transformations of RGB color space and do not require explicit specular/diffuse separation. As described in the previous chapter, Tan and Ikeuchi [102] obtain such a specular invariant using a non-linear combination of the RGB values $\mathbf{I}_{RGB}$ at a pixel and the RGB color vector of the source $\mathbf{S}$. Park [82] defines an alternative, linear transformation providing two color channels that, while not pure invariants, are highly insensitive to specular reflections. In this transformation, one of the coordinate axes of color space is aligned with the source color $\mathbf{S}$, leaving the remaining two channels to be predominantly diffuse.

In the previous chapter, we defined the *SUV color space* using $\mathbf{I}_{SUV} = \mathbf{R}\mathbf{I}_{RGB}$, where $\mathbf{R}$ is any rotation of $\mathbb{R}^3$ satisfying $\mathbf{RS} = (1,\ 0,\ 0)$. Like Park's transformation, this rotation aligns one of the color axes (the S-axis) with the source color $\mathbf{S}$, but unlike Park's transformation, the components along the remaining two axes are indeed invariant to specular reflections. This is easily seen by applying the transformation to Eq. 3.1 and verifying that the U and V components are given by

$$I_U = \sigma_d \mathbf{r}_2^\top \mathbf{D}, \quad I_V = \sigma_d \mathbf{r}_3^\top \mathbf{D}, \tag{3.4}$$

where $\mathbf{r}_2^\top$ and $\mathbf{r}_3^\top$ denote the 2nd and 3rd rows of the rotation matrix $\mathbf{R}$. Since they are independent of the specular geometric scale factor σ_s, these components constitute a specular invariant. An important feature of this transformation is that it preserves and isolates the diffuse shading information (σ_d).

3.2 Color Subspaces

The SUV color transformation can be viewed as a projection of RGB color vectors onto the two-dimensional subspace orthogonal to the source color $\mathbf{S}$. (See left of Fig. 3.2.) This interpretation provides the main motivation for this chapter, and in this section we show that: 1) it can be generalized to mixed illuminants and hyper-spectral images; and 2) it leads naturally to a notion of generalized hue.

The SUV color transformation is based on Eq. 3.1, which in turn is premised on the assumption that the illuminant SPD is constant over the incident hemisphere of a surface point (*i.e.,* that the illuminant 'color' is the same in all directions.) Notationally, if $L(\boldsymbol{\omega}_i, \lambda)$ represents the incident radiance at a surface point, where $\boldsymbol{\omega}_i = (\theta_i, \phi_i) \in \Omega$ parameterizes the hemisphere of incident directions, the model requires that this input radiance field can be factored (with a slight abuse of notation) as $L(\boldsymbol{\omega})E(\lambda)$. To relate this to the terms in Eq. 3.1, we let $f(\boldsymbol{\theta}, \lambda)$ with $\boldsymbol{\theta} = (\theta_i, \phi_i, \theta_o, \phi_o, \lambda)$ denote the BRDF of the surface, and we write the image formation equation as

$$I_k = \int_\lambda \int_\Omega f(\boldsymbol{\theta}, \lambda) L(\boldsymbol{\omega}_i, \lambda) C_k(\lambda) \cos\theta_i \; d\boldsymbol{\omega}_i d\lambda.$$

According to the dichromatic model, the BRDF of the surface can be decomposed according to

$$f(\boldsymbol{\theta}, \lambda) = f_d(\boldsymbol{\theta}) S(\lambda) + k_s f_s(\boldsymbol{\theta}), \tag{3.5}$$

where k_s is a constant, and this yields the expressions

$$\begin{aligned}
\sigma_d &= \int_\Omega f_d(\boldsymbol{\theta}) L(\boldsymbol{\omega}_i) \cos\theta_i \; d\boldsymbol{\omega}_i \\
\sigma_s &= k_s \int_\Omega f_s(\boldsymbol{\theta}) L(\boldsymbol{\omega}_i) \cos\theta_i \; d\boldsymbol{\omega}_i \\
D_k &= \int S(\lambda) E(\lambda) C_k(\lambda) \; d\lambda \\
S_k &= \int E(\lambda) C_k(\lambda) \; d\lambda.
\end{aligned}$$

To generalize the model, we consider a mixed illumination environment whose spectral content can be written in terms of a finite linear basis:

$$L(\boldsymbol{\omega}_i, \lambda) = \sum_{j=1}^{N} L_j(\boldsymbol{\omega}_i) E_j(\lambda). \tag{3.6}$$

An example with $N = 2$ is an office environment where the illumination can be described as a solid angle of daylight in the direction of the window and a distinct solid angle of fluorescent light in the direction of the ceiling. When the input radiance field can be decomposed in this manner, the BRDF decomposition of Eq. 3.5 yields

$$I_k = \sum_{j=1}^{N} \sigma_d^{(j)} D_k^{(j)} + \sigma_s^{(j)} S_k^{(j)}, \tag{3.7}$$

with

$$\begin{aligned}
\sigma_d^{(j)} &= \int_{\Omega} f_d(\boldsymbol{\theta}) L_j(\boldsymbol{\omega}_i) \cos\theta_i \; d\boldsymbol{\omega}_i \\
\sigma_s^{(j)} &= k_s \int_{\Omega} f_s(\boldsymbol{\theta}) L_j(\boldsymbol{\omega}_i) \cos\theta_i \; d\boldsymbol{\omega}_i \\
D_k^{(j)} &= \int S(\lambda) E_j(\lambda) C_k(\lambda) \; d\lambda \\
S_k^{(j)} &= \int E_j(\lambda) C_k(\lambda) \; d\lambda.
\end{aligned}$$

Equation 3.7 suggests the existence of a specular invariant that is analogous to the two-dimensional UV subspace of the SUV color space. In the SUV formulation, the illuminant color is assumed constant over the input hemisphere (which corresponds to $N = 1$ in Eq. 3.7) and the specular invariant subspace computed from a three-channel RGB image is two-dimensional. In general, given an M-channel (possibly hyper-spectral) image and an N-dimensional spectral basis $\{E_j(\lambda)\}_{j=1\ldots N}$ for the incident illumination, there exists a subspace of dimension $(M-N)$ that is independent of all $\sigma_s^{(j)}$ and therefore invariant to specular reflections. Letting $\{\mathbf{r}_l\}_{l=1\ldots(M-N)}$ represent an orthonormal basis for this specular invariant subspace, the l^{th} component (or 'channel') of the specular invariant image is given by

$$J_l = \sum_{j=1}^{N} \sigma_d^{(j)} \mathbf{r}_l^{\top} \mathbf{D}^{(j)}. \tag{3.8}$$

A specular invariant image with $(M - N)$ channels defined by this equation can be

treated as an image with $(M-N)$ 'colors', but these 'colors' can assume negative values. In some cases it is more convenient to use a grayscale specular invariant given by

$$J_{\text{inv(M−N)}} = \left(\sum_{l=1}^{M-N} I_l^2 \right)^{\frac{1}{2}}, \tag{3.9}$$

where the subscript $J_{\text{inv(u)}}$ is used to indicate that the grayscale invariant is derived from a u-dimensional invariant subspace.

Since the vast majority of cameras record three (RGB) channels, the most interesting case to consider is $N = 2$. An example is shown in Fig. 3.1, where light comes from two sources with different SPDs. These SPDs induce two source color vectors $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ in RGB space, and by projecting the RGB color vectors of the input image onto the one-dimensional subspace orthogonal to these vectors, we create an image that is void of specular reflection effects.

An essential feature of the specular invariants of Eqs. 3.8 and 3.9 is that they *preserve diffuse reflection effects* encoded in the geometric scale factors $\sigma_d^{(j)}$. For many surfaces, the diffuse component is well approximated by the Lambertian model, meaning that the term $f_d(\boldsymbol{\theta})$ in Eq. 3.7 is a constant function of $\boldsymbol{\theta}$ and the geometric scale factors $\sigma_d^{(j)}$ do not change with viewpoint. This implies that the specular invariant images defined by Eqs. 3.8 and 3.9 often: 1) satisfy the 'constant-brightness assumption' underlying most stereo and structure-from-motion systems; and 2) provide access to surface normal information through Lambertian-based photometric reconstruction methods such as shape-from-shading. As a result, by computing these invariants as a pre-processing step, we can expand the domain of applicability of many Lambertian-based algorithms to include a much broader class of specular, non-Lambertian surfaces as seen under complex lighting.

Applications are explored in Sect. 3.3. Next, we define a second class of invariants that can be computed from the color subspaces defined above.

3.2.1 Generalized Hue

In an M-channel image of a scene, illuminated by N different colored lights, the *generalized hue* is defined as the surface of an $(M-N-1)$ dimensional unit sphere parameterized by angles $\boldsymbol{\Psi}$ embedded in the $(M-N)$ dimensional diffuse space. The generalized hue is independent of both shading and specularity.

Returning to the case of uniform source color ($N = 1$) in Eq. 3.6), we derive a

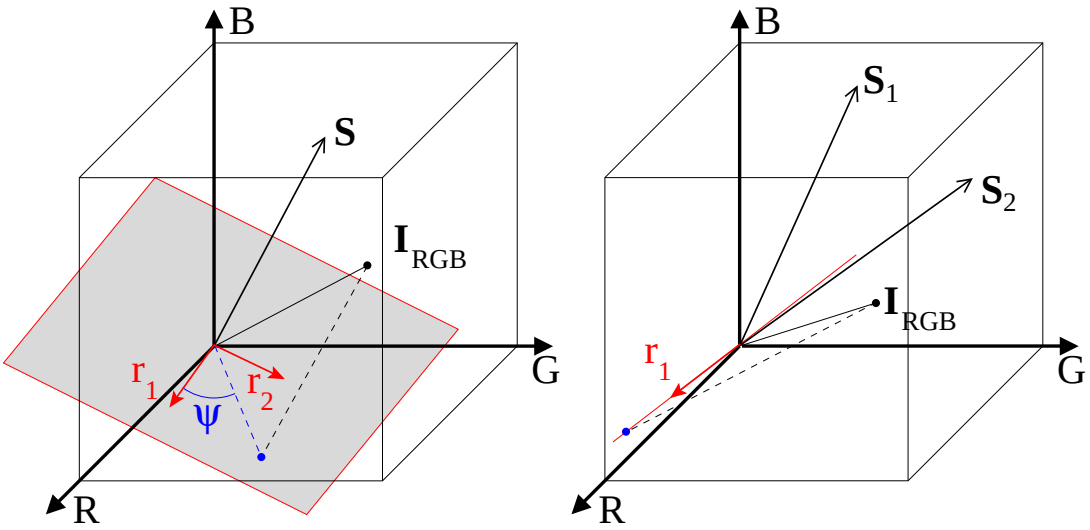


Figure 3.2 Left: Projecting RGB color vector $\mathbf{I}_{RGB}$ onto the 2D subspace orthogonal to the source color $\mathbf{S}$ results in a specular invariant that preserves diffuse shading. The ratio between the channels in this subspace represents *generalized hue* (ψ), which provides a second invariant depending only on spectral reflectance. Right: For two source colors, the specular invariant subspace is one-dimensional. By projecting RGB color vectors onto this line, a specular invariant can still be computed (see Fig. 3.1.)

second invariant by taking the ratio between specular invariant channels in Eq. 3.8. The result,

$$J_1/J_2 = \mathbf{r}_1^\top \mathbf{D}/\mathbf{r}_2^\top \mathbf{D},$$

is independent of both the diffuse and specular geometric scale factors σ_d and σ_d. As shown in Fig. 3.2, it is instructive to interpret this ratio as an angle and define

$$\psi = \tan^{-1}\left(J_1/J_2\right) = \tan^{-1}\left(\mathbf{r}_1^\top \mathbf{D}/\mathbf{r}_2^\top \mathbf{D}\right). \tag{3.10}$$

We refer to ψ as the generalized hue, since it reduces to the standard definition of hue when the source color $\mathbf{S}$ is white.

Examples of generalized hue images are shown in Fig. 3.3 for a specular globe under two different source colors. In each case, the known source color is used to compute a two-channel subspace image according to Eq. 3.8, and the ratio between the two channels is used to compute ψ. Since it depends only on $\mathbf{D}$, each country on the globe appears 'flat' and is free of both specular reflections and diffuse shading.

It is interesting to note that this isolation of spectral reflectance can be generalized to mixed illumination and hyper-spectral images, albeit at the expense of requiring either multiple images, spatial coherence, or both. As an illustration, consider an illumi-

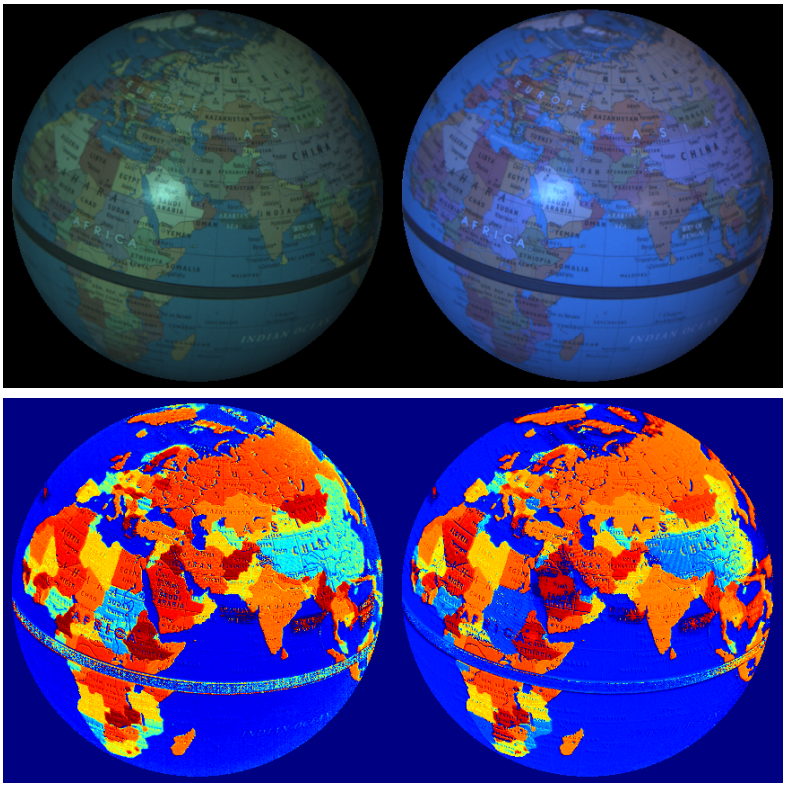


Figure 3.3 Generalized hue images (bottom), each computed from a single RGB image (top) of a globe. Generalized hue is invariant to both specularities and diffuse shading, and is discriminative only with respect to the spectral reflectance of the surface.

nation environment with two different source colors ($N = 2$ in Eq. 3.6), and suppose we acquire two RGB images with altered illumination directions. (The SPDs of the sources remain the same.) In this case, the specular invariant subspace is one-dimensional, and the specular invariant images J and $\bar{J}$ of a surface point under the two lighting configurations are given by (see Eq. 3.8)

$$\begin{aligned} J &= \sigma_d^{(1)} \mathbf{r}^\top \mathbf{D}^{(1)} + \sigma_d^{(2)} \mathbf{r}^\top \mathbf{D}^{(2)} \\ \bar{J} &= \bar{\sigma}_d^{(1)} \mathbf{r}^\top \mathbf{D}^{(1)} + \bar{\sigma}_d^{(2)} \mathbf{r}^\top \mathbf{D}^{(2)}. \end{aligned}$$

Here, we have suppressed the subscript corresponding to the invariant channel index, since there is only a single channel.

Now, suppose we have identified two additional image points (p_2 and p_3) that correspond to surface points having the same geometric configurations (*e.g.*, the same surface normal under distant lighting) but distinct spectral reflectance. This yields a total of six specular invariant observations—three points under two illuminations—that

can be assembled into a 3×2 observation matrix. As shown by Narasimhan *et al.* [74], such a matrix can be factored as

$$\begin{bmatrix} J_{p_1} & \bar{J}_{p_1} \\ J_{p_2} & \bar{J}_{p_2} \\ J_{p_3} & \bar{J}_{p_3} \end{bmatrix} = \begin{bmatrix} \mathbf{r}^\top \mathbf{D}_{p_1}^{(1)} & \mathbf{r}^\top \mathbf{D}_{p_1}^{(2)} \\ \mathbf{r}^\top \mathbf{D}_{p_2}^{(1)} & \mathbf{r}^\top \mathbf{D}_{p_2}^{(2)} \\ \mathbf{r}^\top \mathbf{D}_{p_3}^{(1)} & \mathbf{r}^\top \mathbf{D}_{p_3}^{(2)} \end{bmatrix} \begin{bmatrix} \sigma_d^{(1)} & \bar{\sigma}_d^{(1)} \\ \sigma_d^{(2)} & \bar{\sigma}_d^{(2)} \end{bmatrix},$$

from which it follows that the ratio of determinants of its two 2×2 sub-matrices satisfies

$$\frac{J_{p_1}\bar{J}_{p_2} - \bar{J}_{p_1}J_{p_2}}{J_{p_2}\bar{J}_{p_3} - \bar{J}_{p_2}J_{p_3}} = \frac{\mathbf{D}_{p_1}^{(1)\top}\mathbf{D}_{p_2}^{(2)} - \mathbf{D}_{p_1}^{(2)\top}\mathbf{D}_{p_2}^{(1)}}{\mathbf{D}_{p_2}^{(1)\top}\mathbf{D}_{p_3}^{(2)} - \mathbf{D}_{p_3}^{(2)\top}\mathbf{D}_{p_2}^{(1)}}, \tag{3.11}$$

which depends only on the spectral reflectance of the surface points and is invariant to the geometric scale factors.

While this example relies on the identification of three surface points (p_1, p_2, p_3), a similar invariant can be computed using a single surface point and multiple specular invariant channels collected from a hyper-spectral image. Determinant-based invariants of this form have been well studied by Narasimhan *et al.* [74], who apply them directly to RGB images and obtain invariants for relatively general reflectance functions under monochromatic environments. The analysis presented here can be viewed simultaneously as an extension of their approach to mixed-illumination environments as well as a specialization to the dichromatic model. The latter is an important difference, because it enables the distinction between specular and diffuse material properties in the resulting invariants.

3.2.2 Practical Limitations

The utility of the proposed invariants relies on the angular separation between the diffuse and source colors ($\mathbf{D}^{(j)}$ and $\mathbf{S}^{(j)}$) in color space. When this separation is small, the signal-to-noise ratio (SNR) in the subspace image can be prohibitively low. This is evident, for example, in the generalized hue image of the globe in the bottom-right of Fig. 3.3, where the hue variation within the People's Republic of China is seen to be large. In practice, this can be improved using high dynamic range images. Additionally, surface points with low SNR can be detected by monitoring the angle between the source colors $\mathbf{S}^{(j)}$ and the input color vectors $\mathbf{I}_{RGB}$, and this information can be incorporated into any robust vision algorithm (see, *e.g.*, [113]).

It is also important to note that in order to compute the invariants described

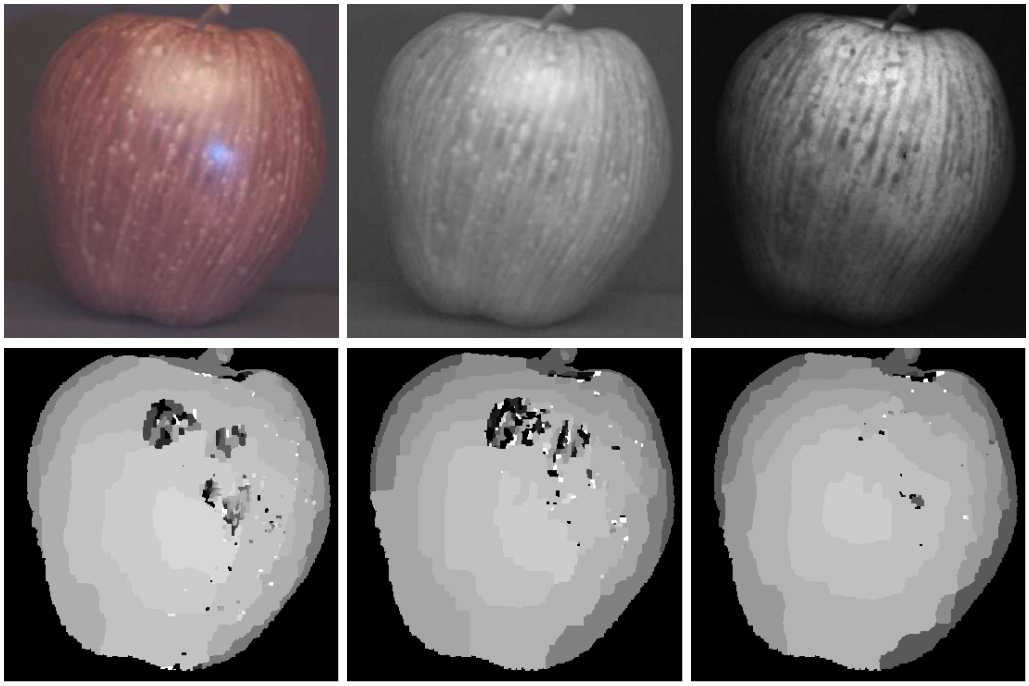


Figure 3.4 Stereo reconstruction under mixed illumination. Top left: One image of an input stereo pair with blue and yellow illumination. Top center: Single-color invariant image $J_{\mathrm{inv}(2)}$ from Eqs. 3.8 and 3.9 with $\mathbf{S}$ in the direction of the blue source. Top right: Two-color invariant $J_{\mathrm{inv}(1)}$ obtained by projecting to the 1D subspace orthogonal to both sources. Bottom row: depth map obtained using the stereo algorithm of Boykov *et al.* [22] in each case.

in the previous sections, we require knowledge of the source colors $\mathbf{S}^{(j)}$. In a controlled setting, these colors can be measured by imaging a 'white' surface under the given illuminants; and estimates of the illuminant colors in an uncontrolled setting can be obtained using existing methods (*e.g.*, [9,103]). From a practical standpoint, it is difficult to provide a meaningful quantitative evaluation of the sensitivity of these invariants to errors in the source colors since it depends on the spectral reflectance and illuminant SPD of a particular scene. We leave a statistical evaluation for future work, and instead, in the next section we assess the practical utility of these invariants by evaluating their performance in a very broad range of applications.

3.3 Applications and Evaluation

This section demonstrates the utility of the proposed invariants using RGB images for a number of vision algorithms and compares the results to those obtained

using standard grayscale images $I_g = (R + G + B)/3$. For RGB images, when the illumination is a mixture of two known colors, the two-color specular invariant J_1 from Eq. 3.8 is grayscale and is equal to $J_{\text{inv}(1)}$ from Eq. 3.9. On the other hand, a single-color specular invariant computed from an RGB image includes two diffuse channels $\{J_1, J_2\}$, which can be combined into a grayscale invariant $J_{\text{inv}(2)}$ using Eq. 3.9. (This is equivalent to the representation $\|J\|$ of Eq. 2.6.) The results in this section show that $J_{\text{inv}(2)}$ is effective in situations where the source color is uniform, and that it performs much better than I_g. In situations where the scene illumination is a mixture of two colors, however, $J_{\text{inv}(2)}$ is not invariant to all specular reflections, and significantly better results can be obtained using the invariant $J_{\text{inv}(1)}$, which is derived from the one-dimensional subspace orthogonal to both source colors.

This section also includes an application of generalized hue to the problem of material-based segmentation.

3.3.1 Stereo

Specular highlights arising due to complex illumination can severely limit the accuracy of stereo algorithms that assume simple illumination and/or Lambertian surfaces. To the best of our knowledge, there are no algorithms for stereo reconstruction that can work for non-Lambertian objects under complex illumination. In cases of significant specular reflections and complex illumination conditions, we can improve the accuracy of existing stereo algorithms by computing the specular invariant as a pre-process. This is demonstrated in Fig. 3.4, which compares binocular stereo results obtained using conventional grayscale images I_g, the single-illuminant (2D subspace) invariant $J_{\text{inv}(2)}$, and the two-color (1D subspace) invariant $J_{\text{inv}(1)}$. In this figure, the grayscale and invariant images are computed from a rectified RGB stereo pair (top of Fig. 3.4) and are used as input to the binocular stereo algorithm of Boykov *et al.* [22]. The original RGB image includes two specular highlights caused by blue and yellow illuminants. The blue highlight is largely eliminated in the single-color invariant $J_{\text{inv}(2)}$, while image $J_{\text{inv}(1)}$ is invariant to specular reflections of both colors. As expected, the results from the grayscale and single-color invariant images are poor in specular regions, and the depth map obtained using $J_{\text{inv}(1)}$ is significantly improved.

3.3.2 Optical Flow

As shown in chapter 2, traditional optical flow algorithms applied to non-Lambertian objects yield incorrect results. These problems are amplified under complex illumination. As demonstrated by the results in Fig. 3.5, optical flow in the presence of specular reflections in a complex illumination environment can be improved by computing a specular invariant as a pre-processing step. In Fig. 3.5, an RGB image sequence is captured by a camera translating horizontally relative to a static scene. The sequence is used to compute a conventional grayscale sequence $I_g(t)$, a single-color invariant sequence $J_{\text{inv}(2)}(t)$, and a two-color invariant sequence $J_{\text{inv}(1)}(t)$. These three videos are used as input to Black and Anandan's algorithm for robust optical flow [17]. The left of Fig. 3.5 shows a single image from each sequence, and the right shows the recovered flows in the indicated window. Since the camera undergoes pure translation, the 'ground-truth' flow lies along parallel horizontal lines. The flows recovered using the conventional grayscale and single-color invariant sequences are shown in green and blue, respectively; and as expected, these flows are severely corrupted by specular highlights. In contrast, the flow computed from the mixed-illuminant invariant (shown in red) is close to the ground truth and is largely unaffected by these non-Lambertian effects.

3.3.3 Material-based Segmentation

Sections 3.3.1 and 3.3.2 demonstrate the utility of the first class of specular invariants under mixed illumination for stereo and optical flow. In this section, we demonstrate an application of the second class of invariants, which is independent of both the specular reflections and diffuse shading in an image. Potential applications of this invariant include segmentation, tracking, and recognition. Here, we consider the application of generalized hue to the problem of material-based segmentation.

Figure 3.6 shows an RGB image of a dichromatic scene under uniform source color ($N = 1$) as well as a number of pseudo-colored representations related to the invariants presented in Sect. 3.2. The top row shows conventional grayscale and specular invariant images, and in the latter, the specular effects (most notably on the green apples, the pumpkin, and the red pepper) are largely eliminated. The bottom-right of Fig. 3.6 shows the generalized hue image given by Eq. 3.10, which is invariant to diffuse shading in addition to specular reflections, and therefore depends only on the spectral reflectance. The fact that the generalized hue within each region is relatively constant suggests that it is a useful representation for segmentation. The same is not true for the conventional

hue image (shown on the bottom-left) because the illuminant is not white.

3.4 Conclusion

This chapter presents two classes of photometric invariants that are derived from color subspaces. They are efficiently computed from a single image of a dichromatic scene and are valid in cases of mixed (*i.e.,* spectrally-varying) illumination environments. The invariants are computed point-wise and therefore place no restriction on scene texture.

Computation of these invariants requires that the source color(s) be known *a priori*, but in the future, we plan to investigate methods that exploit these representations to recover this information from the data. For scenes such as that in Fig. 3.6, for example, it is possible that the entropy of the generalized hue image provides an indicator of the accuracy of the estimated source color.

The practical utility of these invariants is demonstrated by their ability to improve the performance of a wide variety of vision algorithms, including those for stereo and motion estimation. As a result, they provide a means for extending the applicability of existing Lambertian-based algorithms to a more general class of non-Lambertian scenes.

This chapter is based on "Color Subspaces as Photometric Invariants" by T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur. I was responsible for developing the theory for obtaining specular-invariant image in an M channel image illuminated by N light colors. I conducted all the experiments described in the paper.

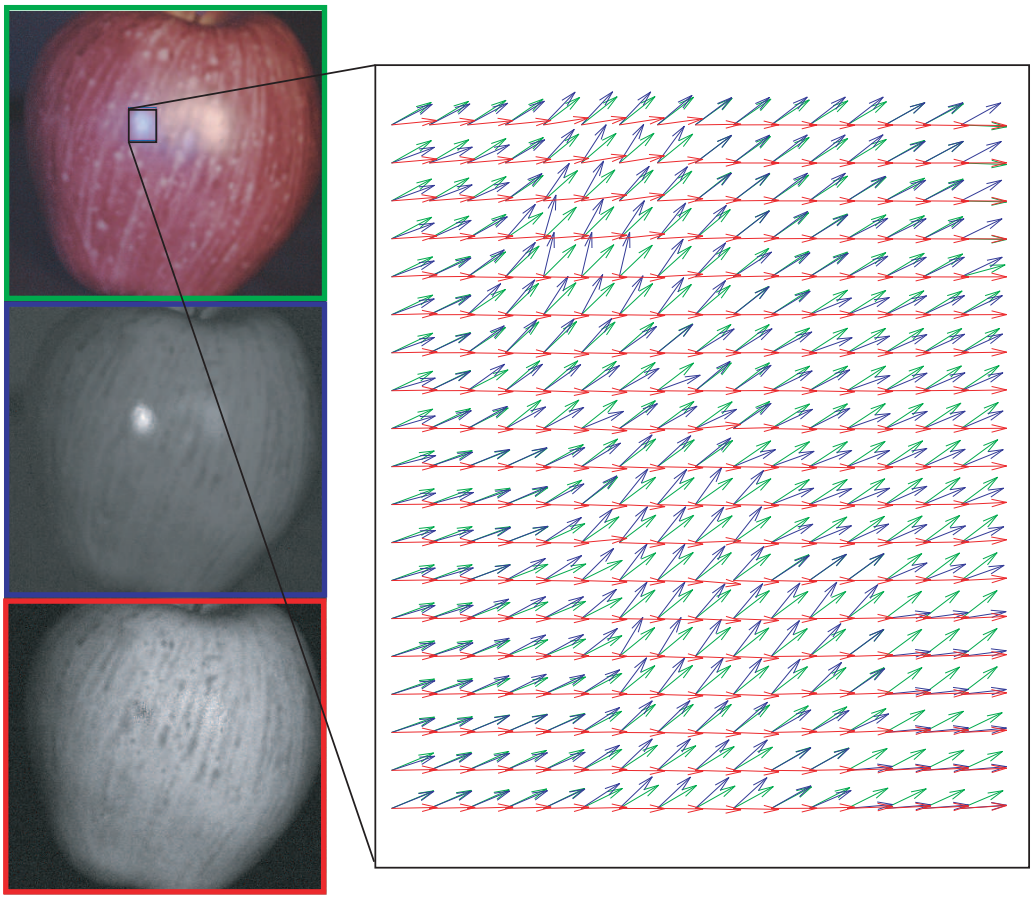


Figure 3.5 Optical flow comparison. An RGB image sequence (top left) is captured by a camera translating left relative to a specular apple under yellow and blue illumination. Derived conventional grayscale $I_g(t)$, yellow-invariant $J_{\text{inv}(2)}(t)$ (left middle), and two-color invariant $J_{\text{inv}(1)}(t)$ (left bottom) sequences are computed and used as input to Black and Anandan's robust optical flow algorithm [17]. Right: flows obtained in the three cases. Green and blue flows are from grayscale and yellow-invariant sequences, respectively, and both are corrupted by specular reflections. Red flow is computed from the two-color invariant and is much closer to ground truth, which is horizontal and to the right.

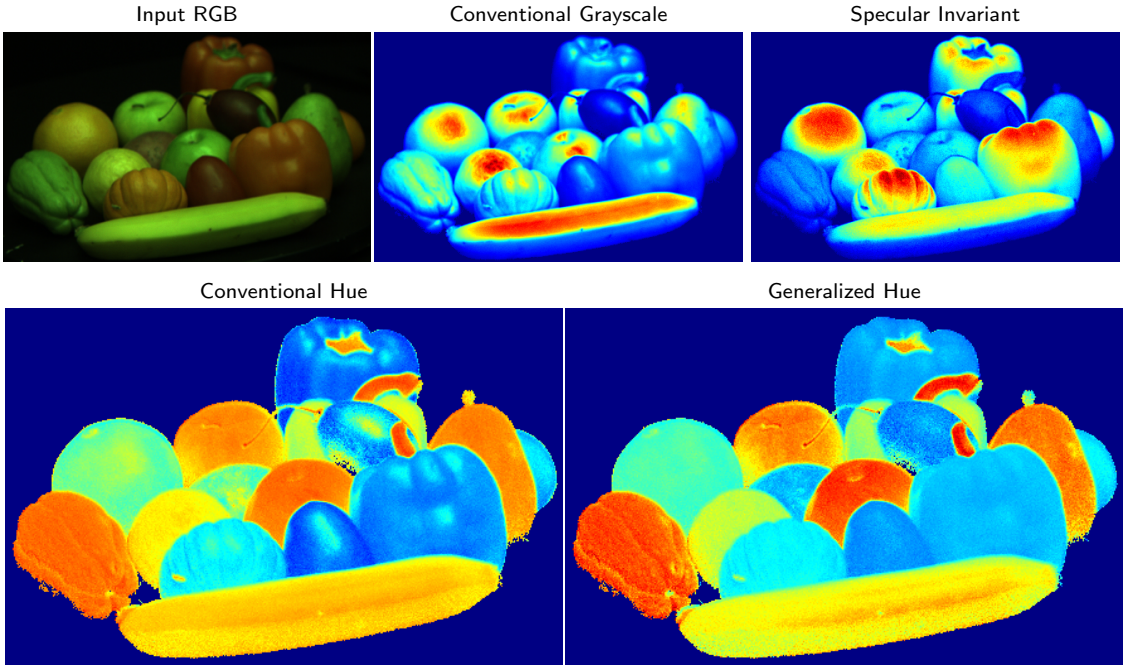


Figure 3.6 Generalized hue for material-based segmentation. Each panel shows a pseudo-colored representation that is computed from the RGB image on the top-left. The generalized hue image on the bottom-right is useful for segmentation because it depends only on the spectral reflectance of the surfaces. The same is not true for a conventional hue image (bottom-left) unless the illuminant is white.

4

Removing Specular Highlights

"There are two kinds of light - the glow that illuminates, and the glare that obscures."

James Thurber

The reflectance of a wide variety of materials (including plastics, plant leaves, cloth, wood and human skin) can be described as a linear combination of specular and diffuse components. When this description is accurate, there are benefits to decomposing an image in this way. The diffuse reflectance component is often well-described by the Lambertian model, and by isolating this component, powerful Lambertian-based tools for tracking, reconstruction and recognition can be applied more successfully to real-world, non-Lambertian scenes. There is also evidence that specular reflectance plays a role in human perception, and there is a set of computer vision algorithms that rely solely on this component (e.g., [18, 37, 80]). Finally, in addition to image-analysis applications, specular/diffuse separation is important in image-based 3-D modeling, where (specular-free) diffuse texture maps are often desired, and in photo-editing, where the two components can be independently processed and recombined.

The methods presented in the previous chapters separate part of the diffuse information from mixed signal (*e.g.*, the U and V channels of the SUV color space). Here we address the problem of complete separation of reflection components in images of general, possibly textured, scenes. We continue to restrict our attention to surfaces that are well-represented by Shafer's dichromatic reflectance model [98], in which the spectral distribution of the specular component is similar to that of the illuminant while that of the diffuse component depends heavily on the material properties of the surface.

Given a single image, specular/diffuse separation requires that color information be shared between image points, and significant research has been devoted to developing

Figure 4.1 An image of a textured surface (left) is separated into its diffuse (middle) and specular (right) components. *Images should be viewed on a monitor or in very high-quality color print.*

global approaches that operate by explicitly segmenting an image into distinct regions of homogeneous diffuse color. When one or more such regions can be identified, the diffuse and specular colors can be estimated, for example, through color histogram analysis [51]. Global approaches have proven useful in color constancy applications where only the illuminant color is desired [30,55,106], but since many natural surfaces exhibit significant diffuse texture, explicitly separating reflectance components using these methods is often difficult to achieve.

We have also seen the development of methods that share color information exclusively through *local* interactions between image points. These methods do not rely on explicit segmentation and can therefore be applied to a much broader class of scenes, including those with texture. Nayar et al. [75], for example, use an iterative scheme that shares color information between each specular point and its diffuse neighbors. The selection of neighboring pixels is governed by both color and polarization constraints that are obtained from multiple exposures. More recently, Tan and Ikeutchi [102] have demonstrated impressive results on single images of textured scenes by sharing diffuse color information between each pixel and one of its four neighbors.

This chapter presents a framework for dichromatic separation based purely on local interactions. Unlike previous approaches, the method is developed in the continuous domain, with local interactions governed by partial differential equations (PDEs). This process selectively shares color information between nearby image points through multi-scale erosion [3, 21, 24] with structuring sets that vary over the image plane. We derive a family of PDEs that are appropriate for differing conditions, including images of both textured and untextured surfaces. We also show how this framework extends naturally

to videos, where motion information is available as an additional cue.

On the practical front, we present results on high-quality images acquired in the laboratory (Figs. 4.1 and 4.5), and show that they compare favorably to ground-truth determined using cross polarization (Fig. 4.6). We also present results on 8-bit images downloaded from the Internet (Fig. 4.7), suggesting robustness to artifacts caused by low dynamic range, JPEG compression, and lack of knowledge of the illuminant color. Finally, we show results on videos (Fig. 4.8) for which explicit optical flow is not necessarily available.

4.1 Color-based Specular/Diffuse Separation

The dichromatic model suggests the possibility of decomposing an image into its specular and diffuse components based on color information. Beginning with a three-channel RGB image, the objective is to recover an RGB diffuse image and a monochromatic specular layer. This is an ill-posed problem, even when the illuminant color is known. Given a single image, most methods operate by aggregating color information spatially across the image plane. As mentioned in the introduction, we can differentiate between methods that are global and local in nature.

Klinker et al. [51] show that when the diffuse color is the same at each point on an object's surface, the color histogram of its image forms a T-shaped distribution, with the diffuse and specular pixels forming linear clusters. They use this information to estimate a single "global" diffuse color, and in principle, this approach can be extended to cases in which an image is segmented into several regions of homogeneous diffuse color. Results can be improved by exploiting knowledge of the illuminant color through transformations of color space [7, 104], but these methods also require an explicit segmentation of the scene into large regions of constant diffuse color. In recent work, Tan and Ikeutchi [101] avoid explicit segmentation by representing all of the diffuse colors in a scene by a global, low-dimensional, linear basis.

In addition to the global approaches mentioned above, there has been considerable interest in separating reflection components through purely local interactions. The advantage of this approach is that it admits highly textured scenes that do not contain piecewise constant diffuse colors. In most local methods, the illuminant color is assumed to be known *a priori*, which is not a severe restriction since it can often be estimated using established (global) methods (e.g., [30, 55, 106]). Tan and Ikeutchi [102] iteratively reduce the specular component of a pixel by considering one of its neighbors that puta-

tively has a related diffuse component. Ping *et al.* [88] allow a user to specify a closed curve surrounding a specular region and then minimize an objective function based on local variations in diffuse chromaticity and specular intensity. One of the earliest local methods is that of Nayar *et al.* [75], which also uses polarization as an additional cue to enable the recovery of a spatially-varying source color.

The goal of this chapter is to formalize the notion of "local interactions" for specular/diffuse separation, and thereby develop a general framework for achieving separation through local interactions in both images and videos. We consider images to be continuous signals of two (or, in the case of video, three) dimensions. This enables a flexible and precise description of a neighborhood, which leads naturally to a family of PDEs that govern local interactions.

Note that in this discussion of color-based methods, we have omitted a number of other methods that rely on additional cues, such as variable lighting [7, 59, 81, 92] variable polarization [75, 110, 120], and parametric reflectance [90].

4.2 Specularity Removal and Differential Morphology

This section derives a family of non-linear PDEs for completing the partial specular/diffuse separation provided by a transformation to SUV color space. Intuitively, these PDEs define a series of local interactions in which color information is shared along curves (or surfaces) of constant "hue."

We begin by re-parameterizing SUV color space using a combination of cylindrical and spherical coordinates. As depicted in Fig. 4.2, suppressing the spatial dependence for notational simplicity, we define

$$\rho = \sqrt{I_U^2 + I_V^2}, \quad \theta = \tan^{-1}\left(\frac{I_U}{I_V}\right), \quad \phi = \tan^{-1}\left(\frac{I_S}{\rho}\right). \tag{4.1}$$

This parameterization has the following properties:

1. Since they depend only on the diffuse UV channels, both ρ and θ are independent of the specular reflectance component.

2. Since the illuminant color is aligned with the S-axis, the angle θ parameterizes the pencil of dichromatic planes in an image. The *generalized hue* θ reduces to the standard definition of hue in the special case of a white illuminant. It depends on the direction of the diffuse color vector but not the magnitude of the diffuse component.

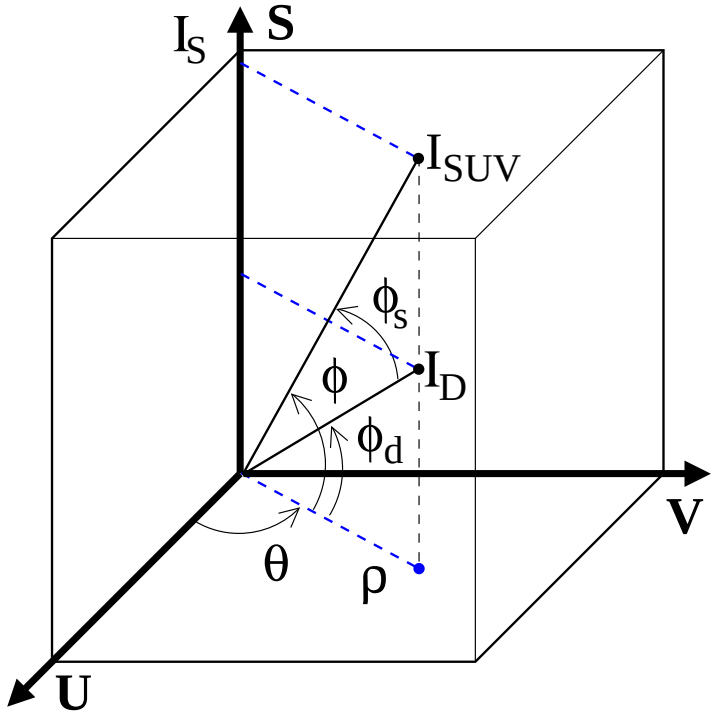


Figure 4.2 A color in the SUV color space is parameterized by (ρ, θ, ϕ). ρ and θ are independent of specularity, and θ generalizes the notion of "hue" for arbitrarily colored illuminants. The problem of computing a specular/diffuse separation is reduced to finding ϕ_d, the diffuse part of ϕ.

3. The ρ-value represents diffuse shading, since it is directly related to $\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}$, and therefore, the magnitude of the diffuse component.

4. The ϕ-value is a linear combination of specular and diffuse components, and we can write $\phi = \phi_s + \phi_d$, where ϕ_s and ϕ_d are the specular and diffuse contributions to ϕ.

According to these properties, the problem of computing a specular/diffuse separation is reduced to one of estimating $\phi_d(x, y)$, the diffuse contribution to ϕ at each image point. Once the scalar function $\phi_d(x, y)$ is known, the RGB specular and diffuse components follow directly from inverting the transformations in Eqs. 4.1 and 2.1, with ϕ replaced by ϕ_s and ϕ_d, respectively.

4.2.1 Multi-scale Erosion

Our goal is to compute a specular/diffuse separation through estimation of the scalar function $\phi_d(x, y)$. As stated in the introduction, we seek to accomplish this goal through

purely local interactions. This section describes how this can be accomplished by evolving a PDE that iteratively "erodes" the specular contribution to ϕ and converges to an estimate of ϕ_d at each point. The erosion process is guided locally by the diffuse color information provided by ρ and θ, and is formulated in the continuous domain using one of a family of non-linear PDEs that define multi-scale erosions [3, 21, 24]. The theory presented in this section is related to the formulation of multi-scale erosion presented by Brockett and Maragos [24].

The multi-scale erosion $\varepsilon(\mathbf{x}, t)$ of a bivariate function $f\colon \mathbb{R}^2 \to \mathbb{R}$ by structuring set $B \subseteq \mathbb{R}^2$ at scale t is defined as

$$\varepsilon(\mathbf{x}, t) = (f \ominus tB)(\mathbf{x}) \stackrel{\triangle}{=} \inf\{f(\mathbf{x} + \Delta\mathbf{x})\colon \Delta\mathbf{x} \in tB\} \ ,$$

where the set B is compact, and $tB \triangleq \{tb\colon b \in B\}$. Intuitively, $\varepsilon(\mathbf{x}, t)$ evaluated at a particular value of t corresponds to an erosion of the function $f(\mathbf{x})$, where the function value at $\mathbf{x} = (x, y)$ has been replaced by the minimum of all function values in the "neighborhood" tB, which is a scaled replica of structuring set B. A multi-scale erosion is computed by considering the PDE

$$\frac{\partial \varepsilon}{\partial t}(\mathbf{x}, t) = \lim_{\Delta t \to 0} \frac{\varepsilon(\mathbf{x}, t + \Delta t) - \varepsilon(\mathbf{x}, t)}{\Delta t}. \tag{4.2}$$

When the structuring set is both compact and convex, the multi-scale erosion has a semigroup structure, allowing one to write [24]

$$\frac{\partial \varepsilon}{\partial t}(\mathbf{x}, t) = \lim_{\Delta t \to 0} \frac{\inf\{\nabla \varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t B\}}{\Delta t}, \tag{4.3}$$

where $\nabla\varepsilon$ is the two-dimensional spatial gradient of ε evaluated at t. Finally, as shown in Ref. [24], in the special case where B is disk-shaped, Eq. 4.3 becomes

$$\varepsilon_t = -\|\nabla \varepsilon\|. \tag{4.4}$$

Equation 4.4 is an example of a PDE that can be used for specular/diffuse separation, albeit in the special case when the scene consists of a texture-less surface with uniform diffuse color. To see this, suppose we are given an input image with corresponding functions $\rho(\mathbf{x}), \theta(\mathbf{x})$ and $\phi(\mathbf{x})$, and suppose we define $\varepsilon(\mathbf{x}, 0) = \phi(\mathbf{x})$. The solution to Eq. 4.4 evaluated at scale t corresponds to the erosion of ϕ by a disk-shaped structuring set, meaning that the value of ϕ at each image point is replaced by the

	Textureless	Texture 1	Texture 2	Video 1	Video 2
Set	2D Disk	2D Line	2D Ellipse	3D Disk	3D Line
Direction	Isotropic	Iso-cont. of θ	Iso-cont. of θ	Iso-surf. of θ	Optic Flow
$\mathbf{M}$	$\mathbf{I}_{2\times 2}$	$\mathbf{I}_{2\times 2} - \nabla\hat{\theta}\ \nabla\hat{\theta}^\top$	$\mathbf{A}\mathbf{A}^\top$	$\mathbf{I}_{3\times 3} - \nabla\hat{\theta}\nabla\hat{\theta}^\top$	$\mathbf{F}\mathbf{F}^\top/\|\mathbf{F}\|^2$

Figure 4.3 Summary of five cases from left to right: (1) image with uniform diffuse color, (2-3) textured image, (4) video, and (5) video with known optical flow. Rows depict: the structuring set used, the direction/surface of erosion, and the matrix $\mathbf{M}$ in the multi-scale erosion equation (Eq. 4.5.) $\mathbf{I}_{n\times n}$ is the identity matrix, and $\nabla\hat{\theta}$, $\mathbf{A}$ and $\mathbf{F}$ are as defined in Sect. 4.2.

minimum value within a disk-shaped neighborhood of radius t. Since $\phi_d(\mathbf{x}) \leq \phi(\mathbf{x})$, it follows that when the image contains at least one image point that is purely diffuse (that is, for which $\phi_s = 0$) then $\varepsilon(\mathbf{x}, t)$ evaluated at t will converge to $\phi_d(\mathbf{x})$ as t is made sufficiently large.

In the next three sub-sections, we develop more sophisticated PDEs for cases of multiple regions of uniform diffuse color, complex diffuse texture, and video. In all of these, the basic idea is the same: the value of ϕ_d at each image point is estimated by eroding the initial function ϕ. By changing the structuring set, however, the process can be controlled so that region boundaries and diffuse texture are preserved during the process. In particular, we show that the PDE governing the evolution of ϕ for three different cases – texture-less images, textured images, and video – can all be written in the form

$$\varepsilon_t = -g(\rho, \nabla\rho)\left(\nabla\varepsilon^\top \mathbf{M} \nabla\varepsilon\right)^{1/2}, \tag{4.5}$$

where $\mathbf{M}$ is a different matrix for each case. $g(\rho, \nabla\rho)$ is called the stopping function and is defined in the following section. Figure 4.3 shows a summary of the different cases described in this chapter.

4.2.2 Texture-less Surfaces: Isotropic Erosion

Equation 4.4 describes a process in which the specular component of ϕ is eroded equally in all directions. This is desirable in cases of homogeneous diffuse color, but if regions of distinct color exist, there is a possibility that "color bleeding" may occur. To prevent this, we introduce a "stopping function" analogous to that used in anisotropic diffusion [85]. In fact, a stopping function is useful for attenuating the erosion process

in two different cases.

1. If a region of the surface is "white" (i.e., it reflects all wavelengths equally) or if the surface is the same color as the light source, the diffuse component of color cannot be isolated using the dichromatic model. Since $\rho = 0$ in this case, no diffuse color information is available, and erosion should be arrested.

2. Information about ϕ should not be shared across boundaries between regions of distinct color. Since these boundaries often coincide with large values of $\|\rho\|$, erosion should be attenuated when $\|\rho\|$ is large.

   One possible stopping function that meets these guidelines is

$$g(\rho, \nabla\rho) = \left(\frac{1 - e^{-\rho}}{1 + e^{-\rho}}\right) \frac{e^{-(\|\nabla\rho\| - \tau)}}{1 + e^{-(\|\nabla\rho\| - \tau)}}, \tag{4.6}$$

where τ is a threshold on $\|\nabla\rho\|$, above which erosion is heavily attenuated. The stopping function is the same for all cases. Incorporating this into Eq. 4.4 yields

$$\varepsilon_t = -g(\rho, \nabla\rho)\|\nabla\varepsilon\| = -g(\rho, \nabla\rho)\left(\nabla\varepsilon^\top \mathbf{I}_{2\times 2} \nabla\varepsilon\right)^{1/2}. \tag{4.7}$$

The erosion process defined by this equation can be used for the specular/diffuse separation of images containing large regions of uniform diffuse color.

4.2.3 Textured Surfaces: Anisotropic Erosion

An example of a scene that does not contain regions of uniform diffuse color is shown in Fig. 4.4. In this case, eroding the function ϕ isotropically would blur the diffuse texture. Instead, we need to erode ϕ anisotropically, only sharing information between neighboring image points for which ϕ_d is likely to be equal. Of course, we have no information about the diffuse color *a priori*, so it is impossible to know the correct direction (if it even exists) with certainty. As depicted in Fig. 4.4, however, since θ is independent of both specularity and shading information, the directions tangent to the iso-contours of $\theta(\mathbf{x})$ provide a good choice. In the absence of any additional information, they provide a good local predictor for the direction in which ϕ_d is constant.

We define

$$\nabla\hat{\theta} = \begin{cases} \nabla\theta/\|\nabla\theta\| & \|\nabla\theta\| > 0 \\ 0 & \|\nabla\theta\| = 0, \end{cases} \tag{4.8}$$

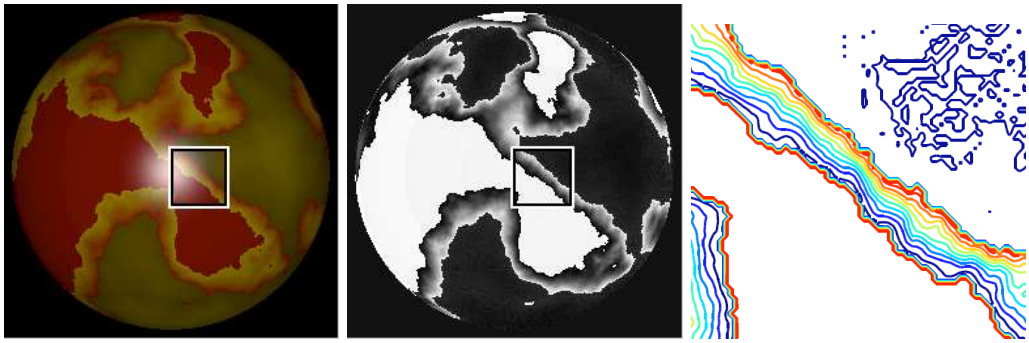


Figure 4.4 Left: a rendered RGB image of a textured sphere. Middle: The value of θ at each pixel. Notice that θ is constant in regions of constant diffuse color and is independent of specularity as well as shading. Right: Blown-up view of the iso-contours of θ in the region indicated in the middle frame. White indicates regions of constant θ. In textured images, erosion of the specular component occurs along iso-contours of θ, which ensures that diffuse texture is preserved while the specularity is removed.

where $\nabla(\cdot)$ refers to the spatial gradient as before, and we denote the direction orthogonal to $\nabla\theta$ by $\mathbf{V}$.[1] The multi-scale erosion of ϕ with the spatially-varying, linear structuring sets $\mathbf{V}(\mathbf{x})$ is derived analogous to the isotropic (disk-shaped) case discussed previously.

$$\begin{aligned}\varepsilon_t &= \lim_{\Delta t \to 0} \frac{\inf\{\nabla\varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t\mathbf{V}\}}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{-\Delta t|\nabla\varepsilon^\top \mathbf{V}|}{\Delta t} = -|\nabla\varepsilon^\top \mathbf{V}| \ . \end{aligned} \tag{4.9}$$

Using the fact that $\mathbf{V} = [\ \hat{\theta}_y \ \ -\hat{\theta}_x\ ]^\top$ (or $[\hat{\theta}_y \ \ -\hat{\theta}_x\ ]^\top$), and including the stopping function, we obtain

$$\varepsilon_t = -g(\rho, \nabla\rho)\left[\nabla\varepsilon^\top \left(\mathbf{I_{2\times 2}} - \nabla\hat{\theta}\nabla\hat{\theta}^\top\right)\nabla\varepsilon\right]^{1/2}. \tag{4.10}$$

Using similar arguments to that in the isotropic case, it can be shown that $\varepsilon(\mathbf{x}, t)$ evaluated at sufficiently large t will be equal to $\phi_d(\mathbf{x})$ (and will yield a correct specular/diffuse separation) if the iso-contour of θ passing through each point $\mathbf{x}$: 1) contains only points for which ϕ_d is constant; and 2) contains at least one point at which a purely diffuse observation ($\phi_s = 0$) is available. Note that in regions where the diffuse color is constant (i.e., $\nabla\hat{\theta} = [\ 0 \ \ 0\ ]^\top$), this equation reduces to Eq. 4.7, and the erosion becomes isotropic as desired.

In practice, the transition from linear to disk-shaped structuring sets in Eq. 4.10

[1] Since θ is periodic, a definition of distance is necessary for its gradient to be correctly computed. We define the distance between two angles θ_1 and θ_2 as $\min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|)$.

is controlled by a threshold on $\|\nabla\theta\|$. This discontinuous transition can be avoided by employing an elliptical structuring set with a minor axis aligned with the direction of $\nabla\theta$ and with an eccentricity that varies smoothly with $\|\nabla\theta\|$. To derive a PDE for the corresponding multi-scale erosion, we let E denote an elliptical structuring set, and we describe this set by the lengths of its major and minor axes (λ_1, λ_2) and the angle between its major axis and the x-axis (ψ). Points $\mathbf{x}$ on the boundary of E satisfy $\mathbf{x}^\top\mathbf{Q}\mathbf{x} = 1$ where $\mathbf{Q} = \boldsymbol{\mathcal{R}}(-\psi)\boldsymbol{\Lambda}^{-2}\boldsymbol{\mathcal{R}}(\psi)$, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2)$ and $\boldsymbol{\mathcal{R}}(\psi)$ is a clockwise rotation of the plane. As before, the multi-scale erosion defined by this set satisfies

$$\varepsilon_t = \lim_{\Delta t \to 0} \frac{\inf\{\nabla\varepsilon^\top \Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t E\}}{\Delta t}. \tag{4.11}$$

To simplify the right-hand side of this equation, we define the transformation $\mathbf{x} = \mathbf{A}\mathbf{x}'$, with $\mathbf{A} = \boldsymbol{\mathcal{R}}(-\psi)\boldsymbol{\Lambda}\boldsymbol{\mathcal{R}}(\psi)'$. The spatial gradient of ε with respect to $\mathbf{x}'$ is given by the chain rule: $\nabla\varepsilon' = \mathbf{A}^\top\nabla\varepsilon$. The transformation $\mathbf{A}$ maps the set E to the unit disk (since $\mathbf{x}^\top\mathbf{Q}\mathbf{x} = \mathbf{x}^\top\mathbf{A}^\top\mathbf{Q}\mathbf{A}\mathbf{x} = \mathbf{x}'^\top\mathbf{x}' = 1$), and as a result, we can write $\inf\{\nabla\varepsilon^\top\Delta\mathbf{x} : \Delta\mathbf{x} \in \Delta t E\} = \inf\{\nabla\varepsilon'^\top\Delta\mathbf{x}' : \Delta\mathbf{x}' \in \Delta t B\}$. Substituting this into Eq. 4.11 and comparing with Eq. 4.4, we obtain $\varepsilon_t = -\|\nabla\varepsilon'\| = -\left(\nabla\varepsilon^\top\mathbf{A}\mathbf{A}^\top\nabla\varepsilon\right)^{1/2}$. Finally, adding the stopping function yields

$$\varepsilon_t = -g(\rho, \nabla\rho)\left(\nabla\varepsilon^\top\mathbf{A}\mathbf{A}^\top\nabla\varepsilon\right)^{1/2}. \tag{4.12}$$

4.2.4 Videos: Anisotropic Erosion in Three Dimensions

Thus far, we have dealt exclusively with still images, but the framework extends naturally to video, which can be treated as a 3D volume $\mathbf{I}(x, y, z)$ in which time is the third dimension z. As in the case of textured images, the direction of $\nabla\theta$ is assumed to be a good local predictor for the direction (in 3D space-time) of maximum diffuse color change. We would like to preserve the component of $\nabla\phi$ along this direction during the erosion process, which is accomplished by restricting the erosion of ϕ to the iso-surfaces of θ. In the absence of additional information, there is no preferred direction within an iso-surface of θ, so a natural choice of structuring set is a circular disk contained within its tangent plane.

To compute the multi-scale erosion equation, we note that the structuring set described above consists of a disk (denoted C) whose surface normal is aligned with $\nabla\theta$. Then, the maximum projection of $\nabla\phi$ onto the plane that contains this disk is given by

$\left(\|\nabla\phi\|^2 - \|\nabla\hat{\theta}^\top\nabla\phi\|^2\right)^{1/2}$, and the evolution equation can be simply written as

$$\begin{aligned}\varepsilon_t &= \lim_{\Delta t\to 0}\frac{\inf\{\nabla\varepsilon^\top\Delta\mathbf{x} : \Delta\mathbf{x}\in\Delta t C\}}{\Delta t} = \lim_{\Delta t\to 0}\frac{-\Delta t\left(\|\nabla\varepsilon\|^2 - \|\nabla\hat{\theta}^\top\nabla\varepsilon\|^2\right)^{1/2}}{\Delta t}\\ &= -\left(\|\nabla\varepsilon\|^2 - \|\nabla\hat{\theta}^\top\nabla\varepsilon\|^2\right)^{1/2}.\end{aligned}$$

After some algebraic manipulations, and taking into account the stopping function, we obtain

$$\varepsilon_t = -g(\rho,\nabla\rho)\left[\nabla\varepsilon^\top(\mathbf{I_{3\times 3}} - \nabla\hat{\theta}\nabla\hat{\theta}^\top)\nabla\varepsilon\right]^{1/2}. \tag{4.13}$$

Note that the erosion equation for textured and texture-less surfaces are special cases of the erosion equation for videos.

As mentioned earlier, if some *a priori* information is known, better structuring sets can be designed. An interesting example is when optical flow estimates are available at each location in a video. Even though the video contains specularities which violate the brightness-constancy constraint, accurate optical flow can be estimated using UV or by using θ, which is independent of both shading and specularity. We let $[\ u(x,y,z)\ v(x,y,z)\ ]^\top$ represent the estimated optical flow at location (x,y,z) in the video, so that space-time points (x,y,z) and $(x+u, y+v, z+1)$ correspond to projections of the same surface element. It follows that ϕ_d can be estimated by eroding ϕ along the direction $\mathbf{F} = [u\ v\ 1]^\top$. Using the expression for erosion by a linear set derived in Eq. 4.10 we obtain

$$\varepsilon_t = -g(\rho,\nabla\rho)\left|\frac{\mathbf{F}^\top}{\|\mathbf{F}\|}\nabla\varepsilon\right| = -g(\rho,\nabla\rho)\left(\nabla\varepsilon^\top\frac{\mathbf{F}\mathbf{F}^\top}{\|\mathbf{F}\|^2}\nabla\varepsilon\right)^{1/2}. \tag{4.14}$$

4.3 Results

The methods were evaluated using images and videos acquired in the laboratory as well as those downloaded from the Internet. Using a known (or approximately known) illuminant color, each image is transformed into SUV space, and the functions ρ, θ and ϕ are computed. Specular/diffuse separation is achieved by numerically integrating the appropriate multi-scale erosion PDE with initial condition $\varepsilon(\mathbf{x},0) = \phi(\mathbf{x})$. The process

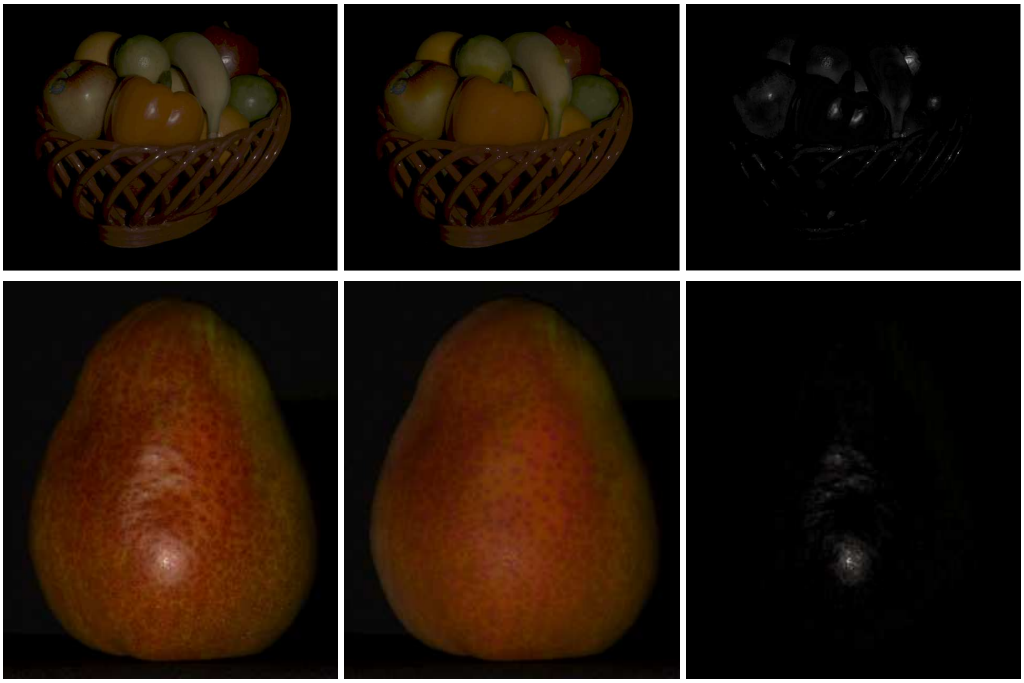


Figure 4.5 Input image (left) with recovered diffuse (center) and specular components (right). Anisotropic multi-scale erosion (Eq. 4.10) is used in both cases, since it naturally handles both textured and untextured surfaces. In the lower example, texture that is nearly completely masked by specularity in the input image is apparent once the specularity is removed. *Images should be viewed on a monitor or in very high-quality print.*

is complete when the maximum change in ε is below a selected threshold, and this yields an estimate of $\phi_d(\mathbf{x})$, which completely defines the specular and diffuse components.

It is important to note that the non-linear PDEs that govern erosion are defined at points where the partial derivatives exist. Even if this is satisfied by the initial data, however, at finite scales the multi-scale erosion generally develops discontinuities in the derivatives referred to as *shocks*. Shocks can be dealt with (as we do here) by replacing standard derivatives by morphological derivatives [24]. They can also be handled using viscosity solutions [28].

Fig. 4.5 shows two 12-bit images acquired in a controlled setting (with known illuminant color) along with the recovered specular and diffuse components. Both results were obtained using the anisotropic erosion defined in Eq. 4.10. The method correctly handles both regions of uniform color (e.g., the orange pepper) and regions with significant texture (e.g., the pear). Looking closely at the pear, we notice that diffuse texture that is barely visible in the input image is revealed when the specularity is removed.

Figure 4.6 Comparison to ground truth. Left: Input image. Middle: Ground truth diffuse component obtained by using linear polarizers. Right: Diffuse component recovered using anisotropic multi-scale erosion. *Images should be viewed on a monitor or in very high-quality print.*

Results from an image of a human face captured under laboratory conditions are shown in Fig. 4.1. Again, diffuse texture is preserved, while the specular component is reduced. Pixels on the forehead between the eyebrows are saturated, and this results in artifacts. The stopping function ensures that these pixels are implicitly identified and treated as outliers during the erosion process. While left here for illustrative purposes, these artifacts can be reduced by inpainting the diffuse and/or specular components in a post-process. (This is done, for example, by Ping *et al.* [88].)

Fig. 4.6 compares the result of our algorithm with the ground truth obtained using polarization filters on the light source and camera. The polarizer in front of the light source is fixed while the polarizer in front of the camera is rotated to an orientation that produces an image with minimal specularity. The result of our algorithm is very close to the ground truth on both the textured surfaces (i.e., the vase and pear) and the untextured surfaces (i.e., the sphere).

Additional still-image results are shown in Fig. 4.7. These images were downloaded from the Internet, so they exhibit low dynamic range (8-bit) and are corrupted by JPEG compression. Since illuminant color was not known, it was assumed to be white, and the gamma was assumed to be 2.2. Despite these sources of noise, the multi-scale erosion defined in Eq. 4.10 still succeeds in separating the diffuse and specular components. An animation of the erosion process has been uploaded as supplemental material.

In addition to still images, we also evaluated the method on video sequences, some frames of which are shown in Fig. 4.8. In both cases, erosion is performed along iso-surfaces of θ using Eq. 4.13, and in both cases, texture is preserved while the specularity is removed. Complete videos are uploaded as supplementary material.

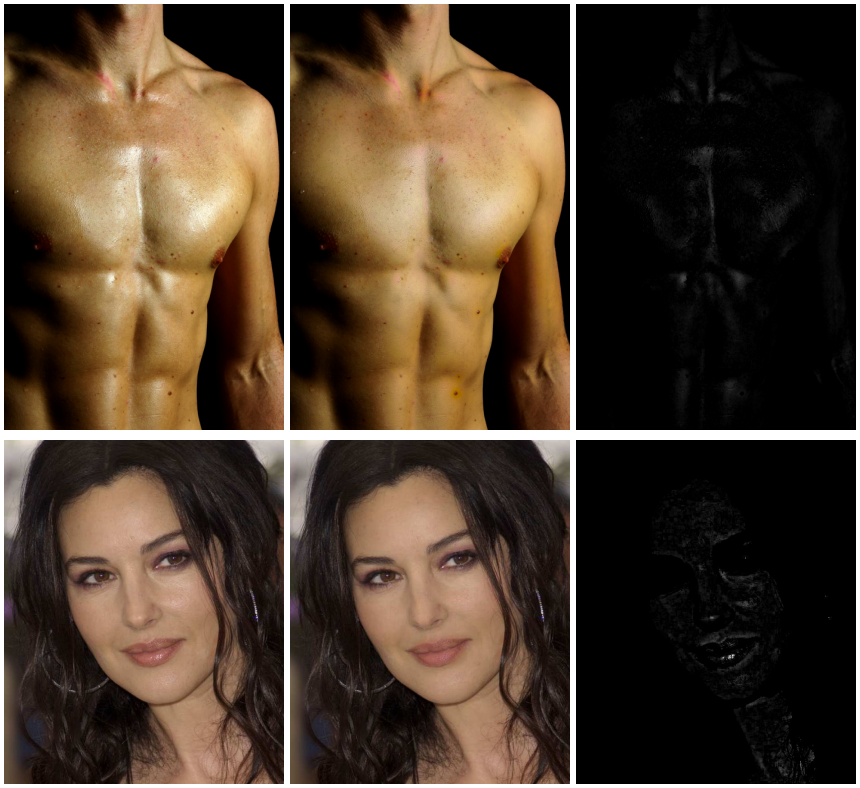


Figure 4.7 Decomposition in the presence of low dynamic range and JPEG artifacts. Eight-bit images (left) were downloaded from the Internet, the illuminant was assumed to be white, and the gamma was assumed to be 2.2. Despite these sources of uncertainity, diffuse and specular components can be recovered using anisotropic multi-scale erosion. *Images should be viewed on a monitor or in very high-quality print.*

4.3.1 Dichromatic Editing

To further demonstrate the efficacy of our approach, we use it as a means for *dichromatic editing* – the simulation of visual effects by the independent processing of reflection components. Some examples are shown in Fig. 4.9 and Fig. 4.10, where: 1) the specular and diffuse components are recovered using Eq. 4.10, 2) each component is processed individually, and 3) they are recombined. Since the diffuse and specular components often form two distinct components of visual perception, dichromatic editing can achieve a variety of visual effects, including the effects of make-up, surface roughening, and wetness.

4.4 Discussion

This chapter presents a framework for specular/diffuse separation in images and video that is based on local spatial (and spatio-temporal) interactions. Separation is framed in terms of differential morphology, which leads to a family of non-linear PDEs. By evolving these PDEs, we effectively erode the specular component at each image point. This erosion is guided by local color and shading information, so that diffuse texture is preserved without requiring an explicit segmentation of the image. Videos are naturally considered in this formulation, with the erosion equation for videos including the still-image equations as a special case.

The approach described in this chapter relies purely on local color information, and is therefore limited to dichromatic surfaces for which the diffuse and specular colors are distinct. It also requires the illuminant color to be known (at least approximately) *a priori*. In the future, we plan to overcome these limitations by exploiting additional cues, such as local shape, in addition to color.

Separation results presented in this chapter include those from low dynamic range JPEG images downloaded from the Internet. For these images, the illuminant color and gamma correction are only approximately known. The results on these images suggest that the approach is robust to various sources of noise, a property that likely stems from the continuous formulation of the problem. Indeed, by developing the problem in terms of morphological PDEs, we benefit from existing robust numerical algorithms to solve them [70]. This is an important advantage over purely discrete formulations.

This chapter is based on "Specularity Removal in Images and Videos: A PDE approach" by S. P. Mallick, T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman. I derived most of the Partial Differential Equations for specular highlight removal in images and videos. I was also responsible for implementation and experiments. In addition, I contributed toward writing the paper.

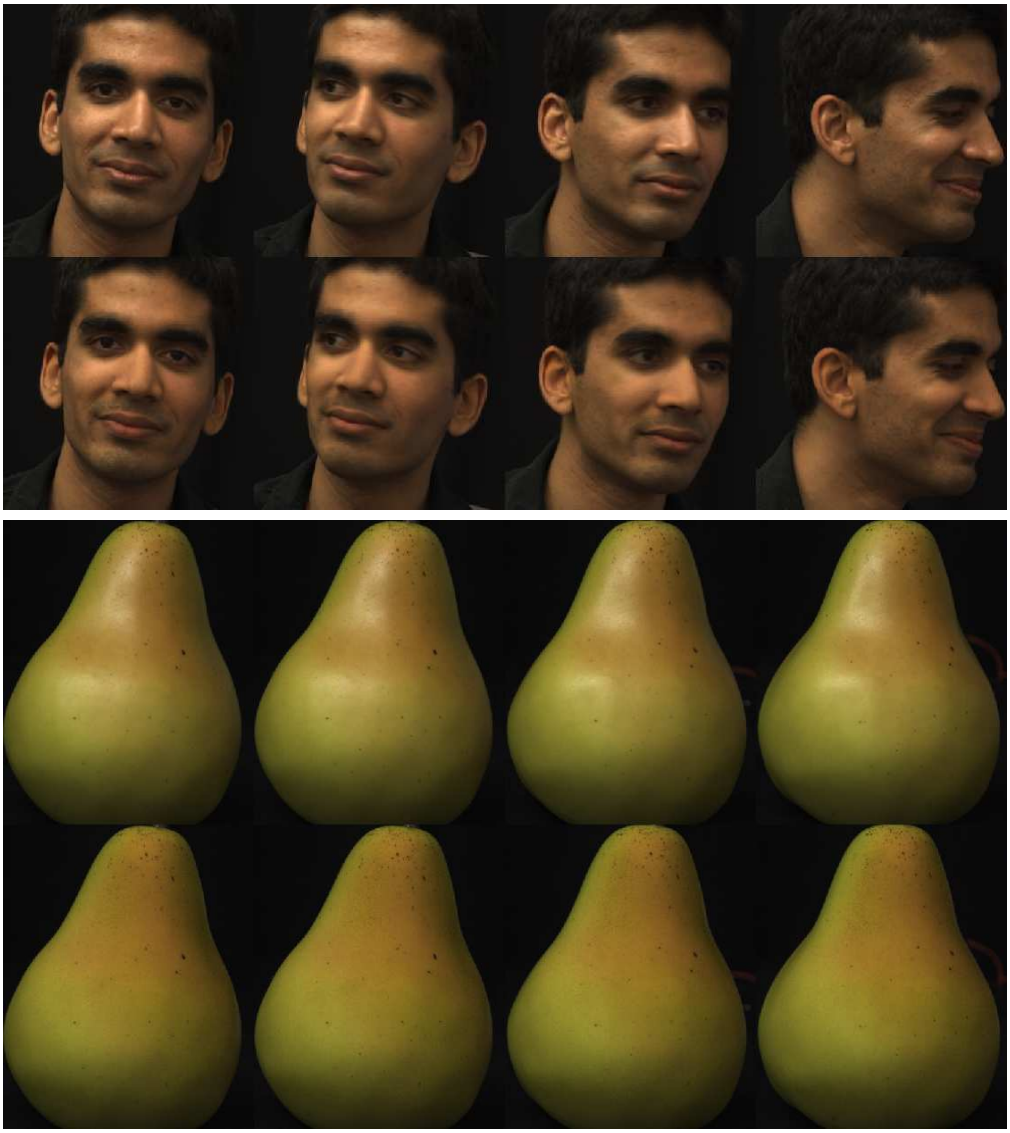
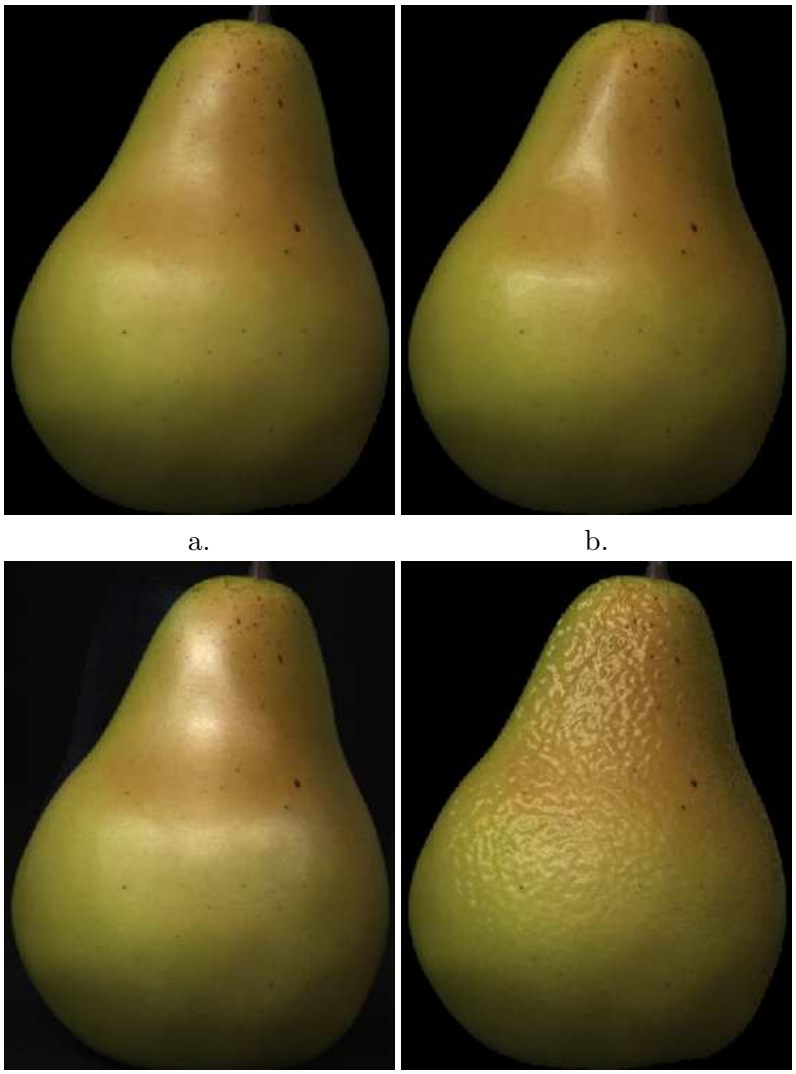


Figure 4.8 Frames representing separation in (12-bit) video. Top row includes frames from an input sequence, and bottom row shows the recovered diffuse component. The results are obtained by eroding along the iso-surfaces of $\theta(x, y, z)$ as described by Eq. 4.13. *Images should be viewed on a monitor or in very high-quality print.*



a. b.

c. d.

Figure 4.9 Dichromatic editing examples. In each case a visual effect is simulated by independent processing of the recovered specular and diffuse components. (a) Input image. (b) Sharpened specular lobe, as would occur if the surface were more smooth. This is acheived by eroding the specular component using a disk-shaped structuring element and amplifying it. (c) Effect of an additional light source obtained by exploiting the object symmetry and reflecting the specular component about the vertical axis. (d) Avocado-like appearance by modulating the specular component.

Figure 4.10 Dichromatic editing examples. In each case a visual effect is simulated by independent processing of the recovered specular and diffuse components. (a) Input image. (b) An effect of makeup is simulated by removing the gloss from the face and enhancing it on the lips. (c) Input image. (d) An effect of wetness is simulated by sharpening the specular layer of skin.

5

Refractive Optical Flow

"While we all have our own perception about what is going on that is filtered through our experience, there is at the bottom the simple truth, which is what I look for."

Michael Ramirez

The human visual system is remarkable in its ability to look at a scene through a transparent refracting object and to deduce the structural properties of that object. For example, when cleaning a wine glass, imperfections or moisture may not be visible at first, but they become apparent when one holds the glass up and moves or rotates it.

We believe that the primary cue here is the optical flow of the background image as observed through the refracting object, and our aims in this chapter are to build a theory of how motion can be used for recovering the structure of a refracting object, to introduce algorithms for estimating this structure, and to empirically validate these algorithms. By structure here we mean a representation of how the object warps and attenuates (or amplifies) the light passing through it. Recall, as a light ray enters or exits the object, its direction will change according to Snell's Law (known as Descartes' Law in France). Furthermore, the emitted radiance may differ from the incident radiance due to the difference in solid angle caused by the geometry of the interfaces between the object and the air as well as absorption along the light ray's path through the object. The geometric shape of the object itself, while of independent interest [54, 73], is not the focus of our inquiry here.

The primary contribution of this work is to generalize the optical flow equation to account for the warping and attenuation caused by a refractive object, and to present algorithms for solving for the warp and attenuation using a sequence of images obtained as a planar background moves behind the refracting object. Both the case where the

background motion is known and where it is unknown are considered. We demonstrate the performance of these algorithms on both synthetic and real data.

While there is a vast literature on motion understanding (including transparency), the warp induced by refraction appears to have been neglected hitherto [46, 49, 56, 95]. The ability to recover the refractive structure of an object has a number of applications. Recently, environment matting and compositing have been introduced as techniques for rendering images of scenes which contains foreground objects that refract and reflect light [25, 117, 132]. The proposed approach offers a new method for creating a matte of real objects without the need for extensive apparatus besides a video camera. Refractive optical flow may also be useful for visual inspection of transparent/translucent objects.

This work was inspired by the work of Zongker *et al.* [132] on environment matting and its subsequent extension by Chuang *et al.* [25]. We discuss this method in the next section. In [132] the authors utilize images of a refracting object acquired by placing it in front of a set of calibrated patterns and a combinatorial search to minimize the reconstruction error, to recover the geometrical warping, the attenuation of light through the object and specular highlights on the surface of the refracting object. More details on this method are provided in the next section. However the work that comes closest to ours in spirit is that of H. Murase [73]. Murase uses optical flow to recover the shape of a flexible transparent object, e.g., water waves that change over time. To make the problem tractable he makes a number of simplifying assumptions (a) The camera is orthographic, (b) The refracting object is in contact with the background plane which has a flat shape and a static unknown pattern on it, (c) the average shape over time of the refracting surface is known *a priori* to have zero slope. In this chapter our interest is in scenes where the refracting object is rigid and stationary. Beyond that, we make no assumptions about the number of objects, their refractive indices, or the positioning of the refracting object in the scene. We do not address effects due to specular, Fresnel or total internal reflections in this work.

There is a significant amount of work on understanding the shape of specular surfaces in a scene. While not of direct relevance to the problem of refraction, the methods used for recovering specular surface geometry serve as a rich source of techniques for analyzing refraction also. See for example [79, 93, 94]. Recently, Kutulakos and Steger [54] reduced the problem of estimating the 3D shape of refracting and reflecting surfaces to the problem of reconstructing individual 3D light paths that cross the image plane. They derive three key results – 1) An algorithm for reconstructing the interface

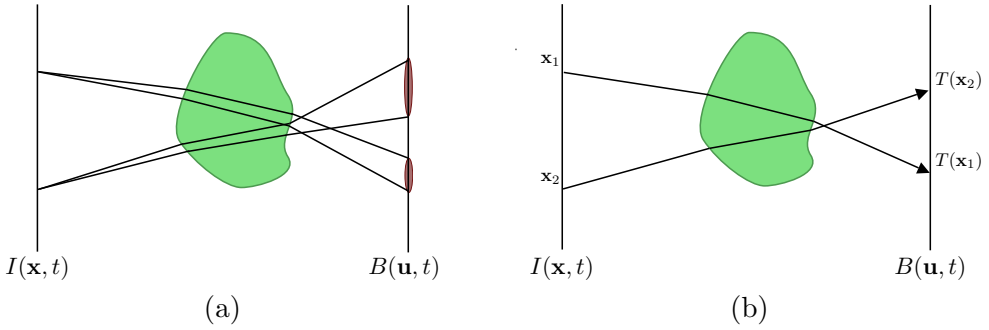


Figure 5.1 The image formation model: (a) The general image formation model, where the incident irradiance at a point $\mathbf{x}$ in the foreground plane I is a result of the emitted radiance of some number of points in the background plane B. (b) The single ray image formation model, where the incident irradiance at $\mathbf{x}$ is linear function of the emitted radiance at the point $T(\mathbf{x})$ in the background plane.

using a single view, when light rays undergo a single refraction/reflection, 2) In case when light rays undergo two refractions/reflections, three views of the surface are sufficient to reconstruct the surface, 3) It is impossible to reconstruct the light paths in case the light undergoes more than two refractions/reflections.

Detecting a refractive object in a scene is also a related problem. Recently, Kenton *et al.* [72] trained a hierarchy of classifiers to detect glass by utilizing visual cues such as systematic distortion of the background, and strong highlights.

The rest of the chapter is organized as follows. In the next section we begin by describing our image formation model and we then introduce the notion of an optical kernel and describe how the choice of a particular optical kernel leads to a new generalization of the optical flow equation. Section 5.2 describes algorithms for solving for the refractive structure using this equation. Section 5.3 demonstrates the performance of our algorithm on synthetic and real data, and its application to matting. We conclude in Section 5.4 with a discussion and directions for future work.

5.1 A Theory of Refractive Optical Flow

In this section we describe our image formation model and use it to derive the refractive optical flow equation. As illustrated in Figure 5.1(a), we assume that the scene consists of three entities.

1. A background image plane B, where $B(\mathbf{u},t)$ denotes the emitted radiance at the point $\mathbf{u} = (u,v)^\top$ and time t.

2. A foreground image plane I, where $I(\mathbf{x}, t)$ denotes the incident radiance at the point $\mathbf{x} = (x, y)^\top$ and time t.

3. A collection of one or more refracting objects between I and B.

The illumination in the scene is assumed to be ambient or coming from a directional light source, and the background plane is assumed to be an isotropic radiator. The background and the foreground planes are not constrained to be parallel. No assumptions are made about the shape of the refracting objects, or optical properties of their constituent materials. We treat the refracting objects as a "black-box" function that warps and attenuates the light rays passing through it. It is our aim to recover this function from a small set of images of the scene.

Given the assumptions about scene illumination stated earlier, the incident radiance at a point $\mathbf{x}$ in the foreground is a result of the reflected and emitted radiance from the background plane. Now let the function that indicates what fraction of the light intensity observed at a point $\mathbf{x}$ in the foreground comes from the point $\mathbf{u}$ in the background be denoted by the optical kernel $\mathcal{K}(\mathbf{u}, \mathbf{x})$. The total light intensity at $\mathbf{x}$ can now be expressed as an integral over the entire background plane,

$$I(\mathbf{x}, t) = \int \mathcal{K}(\mathbf{u}, \mathbf{x}) B(\mathbf{u}, t) d\mathbf{u} \tag{5.1}$$

The problem of recovering the refractive structure of an object can now be restated as the problem of estimating the optical kernel associated with it.

The set of all functions of the form $\mathcal{K}(\mathbf{u}, \mathbf{x})$ is huge. A large part of this set consists of functions that violate laws of physics. However the set of physically plausible optical kernels is still very big, and the reconstruction of $\mathcal{K}$ using a small number of images is an ill-posed problem. Additional constraints must be used to make the problem tractable.

One such direction of inquiry is to assume a low dimensional form for $\mathcal{K}$ by a small set of parameters. Zongker *et al.*in their work on environment matting, assume a parametric box form for $\mathcal{K}(\mathbf{u}, \mathbf{x})$,

$$\mathcal{K}(\mathbf{u}, \mathbf{x}) = \begin{cases} 1/\mu(\mathbf{x}) & \text{if } a(\mathbf{x}) \leq u \leq b(\mathbf{x}) \,\&\, c(\mathbf{x}) \leq v \leq d(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

where, $a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x}), d(\mathbf{x})$ are functions of $\mathbf{x}$. $\mu(\mathbf{x})$ is the area of the rectangle enclosed by (a, c) and (b, d). The kernel maps the average intensity of a rectangular re-

gion in the background to a point in the foreground. The values of the parameters $\{a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x}), d(\mathbf{x})\}$ for each point $\mathbf{x}$ are determined using a set of calibrated background patterns and performing a combinatorial search on them so as to minimize the reconstruction error of the foreground image. Chuang *et al.* [25] generalize $\mathcal{K}$ to be from the class of oriented two-dimensional Gaussians. The estimation procedure now is a two stage non-linear optimization procedure that minimizes the reconstruction error. In both of these cases, knowledge of the background image is assumed.

In this chapter we choose to pursue an alternate direction. We consider optical kernels of the form

$$\mathcal{K}(\mathbf{u}, \mathbf{x}) = \alpha(\mathbf{x})\delta(\mathbf{u} - T(\mathbf{x})) \tag{5.3}$$

where $\delta(\cdot)$ is Dirac's delta function, $T(\mathbf{x})$ is a piecewise differentiable function that serves as the parameter for the kernel indicating the position in the background plane where the kernel is placed when calculating the brightness at $\mathbf{x}$ in the foreground image plane. The function $\alpha(\mathbf{x})$ is a positive scalar function that accounts for the attenuation of light reaching $\mathbf{x}$. Figure 5.1 illustrates the setup.

In the following we will show how if we restrict ourselves this class of optical kernels, we can recover the refractive structure without any knowledge of the background plane. We will also demonstrate with experiments how this subclass of kernels despite having a very simple description is capable of capturing refraction through a variety of objects.

The image formation equation can now be re-written as

$$I(\mathbf{x}, t) = \int \alpha(\mathbf{x})\delta(\mathbf{u} - T(\mathbf{x}))B(\mathbf{u}, t)d\mathbf{u} \tag{5.4}$$

$$= \alpha(\mathbf{x})B(T(\mathbf{x}), t) \tag{5.5}$$

We begin by differentiating the above equation w.r.t $\mathbf{x}$, to get

$$\nabla_{\mathbf{x}} I(\mathbf{x}, t) = (\nabla_{\mathbf{x}}\alpha(\mathbf{x}))B(T(\mathbf{x})) + \alpha(\mathbf{x})\mathbf{J}^{\top}(T(\mathbf{x}))(\nabla_{T(\mathbf{x})}B(T(\mathbf{x}), t)) \tag{5.6}$$

Here, $\mathbf{J}^{\top}(T(\mathbf{x}))$ is the transpose of the Jacobian of the transformation $T(\mathbf{x})$. Using Eq. 5.5 the above equation can be written as

$$\begin{aligned}\nabla_{\mathbf{x}} I(\mathbf{x},t) &= I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})} + \mathbf{J}^{\top}(T(\mathbf{x}))\left[\alpha(\mathbf{x})\nabla_{T(\mathbf{x})}B(T(\mathbf{x}))\right] \\ \mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x},t) - I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] &= \alpha(\mathbf{x})\nabla_{T(\mathbf{x})}B(T(\mathbf{x})) \end{aligned} \tag{5.7}$$

Taking temporal derivatives of Eq. 5.5 gives

$$I_t(\mathbf{x},t) = \alpha(\mathbf{x})B_t(T(\mathbf{x}),t) \tag{5.8}$$

Now, let $\mathbf{c}(\mathbf{u},t)$ denote the velocity of the point $\mathbf{u}$ at time t in the background plane. Then from the Brightness Constancy Eq. [42] we know

$$\mathbf{c}(\mathbf{u},t)^{\top}\nabla_{\mathbf{u}}B(\mathbf{u},t) + B_t(\mathbf{u},t) = 0 \tag{5.9}$$

Now assuming that the background image undergoes in-plane translation with velocity $\mathbf{c}(\mathbf{u},t) = \mathbf{c}(T(\mathbf{x}),t)$, we take the dot product of equation 5.7 with $\mathbf{c}(T(\mathbf{x}),t)$ and add it to Eq. 5.8 to get

$$\begin{aligned}\mathbf{c}(T(\mathbf{x}),t)\mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x},t) - I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] + I_t(\mathbf{x},t) = \\ \alpha(\mathbf{x})\left[\mathbf{c}(T(\mathbf{x}),t)\nabla_{T(\mathbf{x})}B(T(\mathbf{x})) + B_t(T(\mathbf{x})\right]\end{aligned} \tag{5.10}$$

From Eq. 5.9 we know that the right hand side goes to zero everywhere, giving us

$$\mathbf{c}(T(\mathbf{x}),t)\mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla_{\mathbf{x}} I(\mathbf{x},t) - I(\mathbf{x},t)\frac{\nabla_{\mathbf{x}}\alpha(\mathbf{x})}{\alpha(\mathbf{x})}\right] + I_t(\mathbf{x},t) = 0 \tag{5.11}$$

Now for simplification's sake, let $\beta(\mathbf{x}) = \log\alpha(\mathbf{x})$. Dropping the subscript on ∇ we get

$$\boxed{\mathbf{c}(T(\mathbf{x}),t)\mathbf{J}^{-\top}(T(\mathbf{x}))\left[\nabla I(\mathbf{x},t) - I(\mathbf{x},t)\nabla\beta(\mathbf{x})\right] + I_t(\mathbf{x},t) = 0} \tag{5.12}$$

This is the refractive optical flow equation.

5.1.1 Properties of the Refractive Optical Flow Equation

Before we dive into the solution of Eq. 5.12 for recovering the refractive structure, we comment on its form and some of its properties. The first observation is that if there is

no distortion or attenuation, i.e.,

$$T(\mathbf{x}) = \mathbf{x}, \qquad \alpha(\mathbf{x}) = 1$$

the Jacobian of T reduces to the identity, and the gradient of $\beta(\mathbf{x})$ reduces to zero everywhere, giving us the familiar optical flow equation in the foreground image.

$$\mathbf{c}^\top(\mathbf{x}, t)\nabla I(\mathbf{x}, t) + I_t(\mathbf{x}, t) = 0 \tag{5.13}$$

The second observation is that the equation is independent of $B(\mathbf{u}, t)$. This means that we can solve for the refractive flow through an object just by observing a distorted version of the background image. Knowledge of the background image itself is not needed.

Finally we observe that Eq. 5.12 is in terms of the Jacobian of T, i.e. for any function $T'(\mathbf{x}) = T(\mathbf{x}) + \mathbf{u}_0$ will result in the same equation. This implies that the T can only be solved up to a translation ambiguity. Visually this is equivalent to viewing the scene through a periscope. The visual system has no way of discerning whether or not an image was taken through a periscope. A second ambiguity is introduced into the solution when we note that the velocity $\mathbf{c}(\mathbf{u}, t)$ is in the background plane and there is nothing that constrains the two coordinate systems from having different scales along each axis. Hence $T(\mathbf{x})$ can only be recovered up to a four parameter family of ambiguities corresponding to translation and scaling. The attenuation function $\beta(\mathbf{x})$ is not affected by the scaling in $\mathbf{c}(\mathbf{u}, t)$ and hence is recovered up to a translation factor which in turn means that $\alpha(\mathbf{x})$ is recovered up to a scale factor.

5.2 Solving the Equation of Refractive Flow

In this section, we further analyze Eq. 5.12 for the purposes of solving it. We begin by considering a further simplification of Eq. 5.12. We assume that the background plane translates with in plane velocity $\mathbf{c}(\mathbf{u}, t) = \mathbf{c}(t)$ that is constant over the entire background plane. We consider two cases, the calibrated case (when the motion of the background $\mathbf{c}(t)$ is known) and the uncalibrated case (when the motion of the background is unknown). In each case we describe methods for recovering $T(\mathbf{x})$ and $\alpha(\mathbf{x})$, and note the ambiguities in the resulting solution.

Let $T(\mathbf{x})$ be denoted by

$$T(x,y) = (g(x,y), h(x,y))^\top \tag{5.14}$$

The Jacobian of T and its inverse transpose are

$$\mathbf{J}(T(\mathbf{x})) = \begin{bmatrix} g_x & g_y \\ h_x & h_y \end{bmatrix} \qquad \mathbf{J}^{-\top}(T(\mathbf{x})) = \frac{1}{g_x h_y - g_y h_x} \begin{bmatrix} h_y & -h_x \\ -g_y & g_x \end{bmatrix}$$

The translation velocity in the background plane is $\mathbf{c}(\mathbf{u},t) = (\xi(t), \eta(t))^\top$. Substituting these in Eq. 5.12 we get

$$\begin{bmatrix} \xi & \eta \end{bmatrix} \frac{1}{g_x h_y - h_x g_y} \begin{bmatrix} h_y & -h_x \\ -g_y & g_x \end{bmatrix} \begin{bmatrix} I_x - \beta_x I \\ I_y - \beta_y I \end{bmatrix} + I_t = 0 \tag{5.15}$$

which rearranges to

$$\frac{\eta I_y g_x - \eta I_x g_y - \xi I_y h_x + \xi I_x h_y + \eta I(g_y \beta_x - g_x \beta_y) - \xi I(h_y \beta_x - h_x \beta_y)}{g_x h_y - g_y h_x} + I_t = 0 \tag{5.16}$$

Now let

$$p = \frac{g_x}{g_x h_y - g_y h_x} \qquad q = \frac{g_y}{g_x h_y - g_y h_x} \qquad a = \frac{g_y \beta_x - g_x \beta_y}{g_x h_y - g_y h_x} \tag{5.17}$$

$$r = \frac{h_x}{g_x h_y - g_y h_x} \qquad s = \frac{h_y}{g_x h_y - g_y h_x} \qquad b = \frac{h_y \beta_x - h_x \beta_y}{g_x h_y - g_y h_x} \tag{5.18}$$

we get

$$\eta I_y p - \eta I_x q - \xi I_y r + \xi I_x s + \eta I a - \xi I b + I_t = 0 \tag{5.19}$$

5.2.1 The Calibrated Case

In the case where the velocity of the background plane $\mathbf{c}(\mathbf{u},t)$ is known, Eq. 5.19 is linear in p, q, r, s, a, b. Given seven or more successive frames of a video or 14 or more pairs of successive frames and the associated motion of the background plane, we can solve the equation point-wise over the entire image. Given $n+1$ successive frames, we get n equations in 6 variables at each point in the foreground plane, giving us an over-constrained linear system of the form

$$\mathbf{A}\mathbf{p} = \mathbf{m} \tag{5.20}$$

here the i^{th} row of the matrix $\mathbf{A}$ is given by

$$\mathbf{A_i} = \begin{bmatrix} \eta(i)I_y(i) & -\eta(i)I_x(i) & -\xi(i)I_y(i) & \xi(i)I_x(i) & \eta(i)I(i) & -\xi(i)I(i) \end{bmatrix}$$
$$\mathbf{m} = [-I_t(1), \ldots, -I_t(n)]^\top \qquad \mathbf{p} = \begin{bmatrix} p & q & r & s & a & b \end{bmatrix}^\top \tag{5.21}$$

The above system of equations is solved simply by the method of linear least squares. The corresponding values of $g_x, g_y, h_x, h_x, \beta_x, \beta_y$ can then be obtained as follows.

$$g_x = \frac{p}{ps - qr} \qquad g_y = \frac{q}{ps - qr} \qquad \beta_x = \frac{bp - ar}{ps - qr} \tag{5.22}$$
$$h_x = \frac{r}{ps - qr} \qquad h_y = \frac{s}{ps - qr} \qquad \beta_y = \frac{bq - as}{ps - qr} \tag{5.23}$$

5.2.2 The Uncalibrated Case

If the motion of the background image is not known, Eq. 5.19 is bilinear in the variables p, q, r, s, a, b and $\xi(t), \eta(t)$. If we consider frames $i = 1, \ldots, n+1$ and pixels $j = 1, \ldots, m$ in the foreground image, then we can rewrite Eq. 5.19 in the following form

$$\mathbf{c}_i^\top \mathbf{A}_{ij} \mathbf{p}_j = 1 \qquad i = 1, \ldots, n \quad j = 1, \ldots, m \tag{5.24}$$

where

$$\mathbf{c}_i = \begin{bmatrix} \xi(i) & \eta(i) \end{bmatrix}^\top \qquad \mathbf{p}_j = \begin{bmatrix} p(j) & q(j) & r(j) & s(j) & a(j) & b(j) \end{bmatrix}^\top$$
$$\mathbf{A}_{ij} = \frac{-1}{I_t(i,j)} \begin{bmatrix} 0 & 0 & -I_y(i,j) & I_x(i,j) & 0 & -I(i,j) \\ I_y(i,j) & -I_x(i,j) & 0 & 0 & I(i,j) & 0 \end{bmatrix}$$

This gives us nm equations in $2n + 6m$ variables and we can solve them whenever $nm > 2n + 6m$.

Eq. 5.24 is a system of bilinear equations,i.e. given $\mathbf{c}_i$, the system reduces to a linear system in $\mathbf{p}_j$ and vice versa. The overall problem however is highly non-linear. Using this observation a simple alternating procedure for solving the system can be used, which starting with a random initialization, alternates between updating $\mathbf{c}_i$ and $\mathbf{p}_j$ using linear least squares. Even with the fast iterative procedure, solving for the structure in this manner is not feasible. Hence we use a small slice through image stack and use it to solve for $\mathbf{c}_j$, which is then used as input to the calibrated algorithm described in the previous section to recover the flow over the entire foreground plane.

Properties of the Solution

Observe that, $\mathbf{Q}\mathbf{A}_{ij}\mathbf{R} = \mathbf{A}_{ij} \quad \forall i, j$ where

$$\mathbf{Q} = \begin{bmatrix} q_{12} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \quad \mathbf{R} = \frac{1}{q_{11}q_{22} - q_{12}q_{21}} \begin{bmatrix} q_{11} & 0 & q_{21} & 0 & 0 & 0 \\ 0 & q_{11} & 0 & q_{21} & 0 & 0 \\ q_{12} & 0 & q_{22} & 0 & 0 & 0 \\ 0 & q_{12} & 0 & q_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{11} & q_{21} \\ 0 & 0 & 0 & 0 & q_{12} & q_{22} \end{bmatrix} \tag{5.25}$$

Hence, not knowing $\mathbf{c}(\mathbf{u}, t)$ gives rise to a solution which is ambiguous up to a 2×2 invertible transformation, and a corresponding ambiguity in the values of the various gradient estimates. Coupled with this is the translation ambiguity in g and h. This gives rise to a six parameter family of ambiguities in the overall solution.

5.2.3 Integration

The methods described in the previous two sections give us estimates of the partial derivatives of $g(x, y)$, $h(x, y)$ and $\beta(x, y)$. The final step is integrating the respective gradient fields to recover the actual values of the functions. Reconstruction of a function from its first partial derivatives is a widely studied problem. We use an iterative least squares optimization that minimizes the reconstruction error [40].

5.3 Experiments

All experiments were done with video sequences of 200 frames each. The synthetic data were generated using a combination of MATLAB and POVRay, a public domain ray tracer. The real data was captured by placing a refracting object in front of an LCD screen, and imaging the setup using a firewire camera. Figure 5.2 illustrates the data acquisition setup. Calculation of image derivatives is very sensitive to noise. Smoothing was performed on the images using anisotropic diffusion [86], which has superior behavior along sharp edges as compared to Gaussian filtering. This is important for objects that cause light rays to be inverted, which in turn causes the optical flow across the boundary to be opposite sign; a naive Gaussian based smoothing procedure will result in significant loss of signal.

The least squares estimation step in the calibrated estimation algorithm was

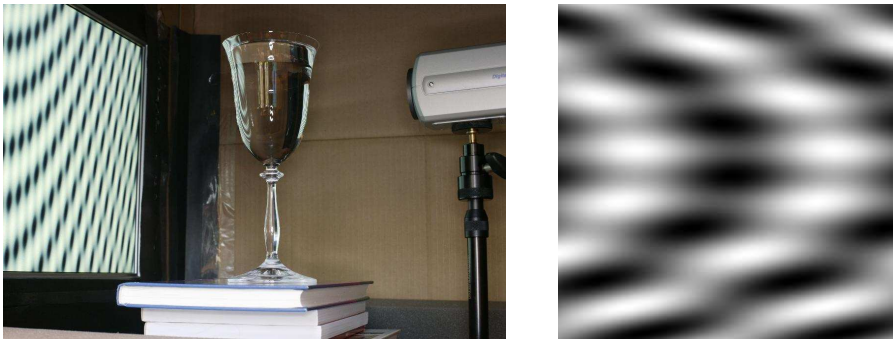


Figure 5.2 (a) Shown above is a photograph of the data acquisition system used in our experiments. It consists of Samsung 19" LCD display screen on the left, a Sony DFW-VL500 firewire camera on the right and a refracting object between the screen and the camera. (b) shows a frame from the background image sequence for all our experiments.

made robust by only considering equations for which the temporal gradient term I_t was within 85% of the maximum temporal gradient at that pixel over time. This choice results in only those constraints being active where some optical flow can be observed.

The boundary of refracting objects typically have little or no optical flow visible. This results in the refractive optical flow constraint breaking down along the boundary as well as certain medium interfaces. We mask these pixels out by considering the average absolute temporal gradient at each point in the foreground image plane and ignoring those pixels that fall below a chosen threshold. This results in a decomposition of the image into a number of connected components. All subsequent calculations are carried out separately for each connected component.

5.3.1 Results

We begin by considering a synthetic example to illustrate the performance of our algorithm in the calibrated and the uncalibrated case. The warping function used was $T(x, y) = (xe^{-(x^2+y^2)}, ye^{-(x^2+y^2)})^\top$ and the attenuation factor was $\alpha(x, y) = e^{-(x^2+y^2)}$. Figure 5.3.1 shows a comparison between the estimated and the true motion in the uncalibrated case. Figure 5.4 illustrates the results of the experiment. The estimated warp and attenuation functions are virtually identical to the true warp and attenuation functions. In the uncalibrated case, the estimated motion was disambiguated by doing a least squares fit to the ground truth for the purposes of illustration.

Figures 5.4 and 5.4 illustrate the result of applying the refractive structure of a glass with and without water inside it to the task of environment matting. Note that

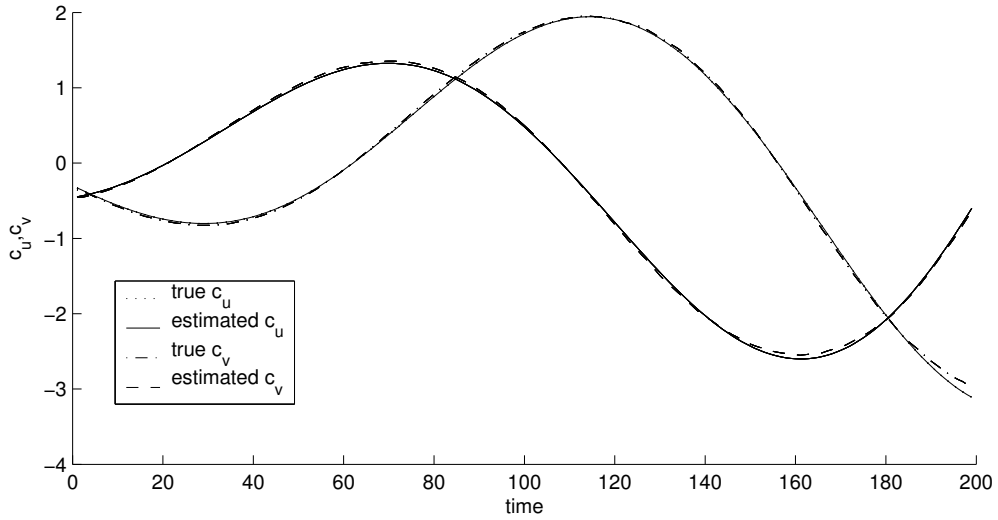


Figure 5.3 Estimation of background motion in the uncalibrated case. This figure plots the true and estimated motion in the background plane. The two curves show motion along the $\xi(t)$ and $\eta(t)$ along the u and v axis respectively. As can be seen there is virtually no difference between the true and estimated values for $\xi(t)$ and $\eta(t)$.

the region along the rim of the glass is missing. This is more so in case of the glass filled with water. These are the regions where the refractive optical flow equation breaks down. The black band in the case of the filled glass is due to a combination of the breakdown of the refractive optical flow equation along the air/water/glass interface and the finite vertical extent of the image.

5.4 Discussion

We have introduced a generalization of the optical flow constraint, described methods for solving for the refractive structure of objects in the scene, and shown that this can be readily computed from images. We now comment on the limitations of our work and directions of future work.

First, our method does not address Fresnel and total internal reflection. This places a limit on our analysis to the case where all the illumination comes from behind the object being observed. Methods for recovering surface geometry from specular reflections are an active area of research and are better suited for this task [79, 93, 94].

Second, the presented approach, like [25,132], formulates the objective as determining a plane-to-plane mapping, i.e., it only informs us about how the object distorts a plane in 3-space. A more satisfactory solution will be to recover how the object dis-

torts arbitrary light rays. We believe this is solvable using two or more views of the background plane and is the subject of our ongoing work.

The choice of the optical kernel that resulted in the single ray model was made for reasons of tractability. This however leaves the possibility of other optical kernels that account for multiple ray models. Chuang et al [25] extend the box kernel from [132] to include oriented variable width Gaussian optical kernels. Their work however assumes knowledge of the background patterns.

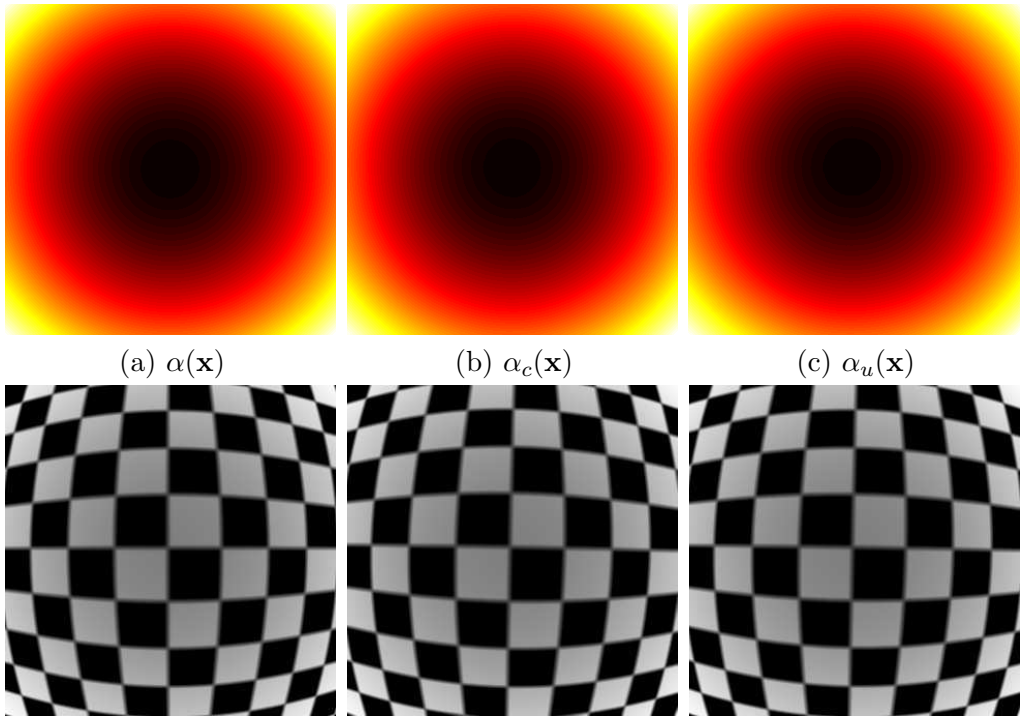
Our experiments were carried out using a 19 inch LCD screen as the background plane. For most refracting objects, as their surface curvature increases, the area of the background plane that they project onto a small region in the foreground image can increase rapidly. If one continues to use a flat background one would ideally require an infinite background plane to be able to capture all the optical flow. Obviously that is not practical. An alternate approach is to use a curved surface as the background, perhaps a mirror which reflects an image projected onto it.

The algorithms described in this chapter estimates the function gradients g_x, g_y, h_x, h_y and integrates them in two separate stages. This still leaves open the possibility of a direct approach that discretizes Eq. 5.12 and solves it directly in one step based on methods borrowed from the study of partial differential equations.

The current work only considers gray scale images; extensions to color or multi-spectral images are straightforward. There are two cases here: if the distortion T is assumed to be the same across spectral bands, the generalization can be obtained by modeling $\alpha(\mathbf{x})$ as a vector valued function that accounts for attenuation in each spectral band. In case T is dependent on the wavelength of light, each spectral band results is an independent version of Eq. 5.12 and can be solved using the methods described.

Finally, the refractive optical flow equation is a very general equation describing optical flow through a distortion function. This allows us to address distortion not due to transmission through transparent objects, but also due to reflection from non-planar mirrored and specular surfaces. We believe that the problem of specular surface geometry can be addressed using this formalism.

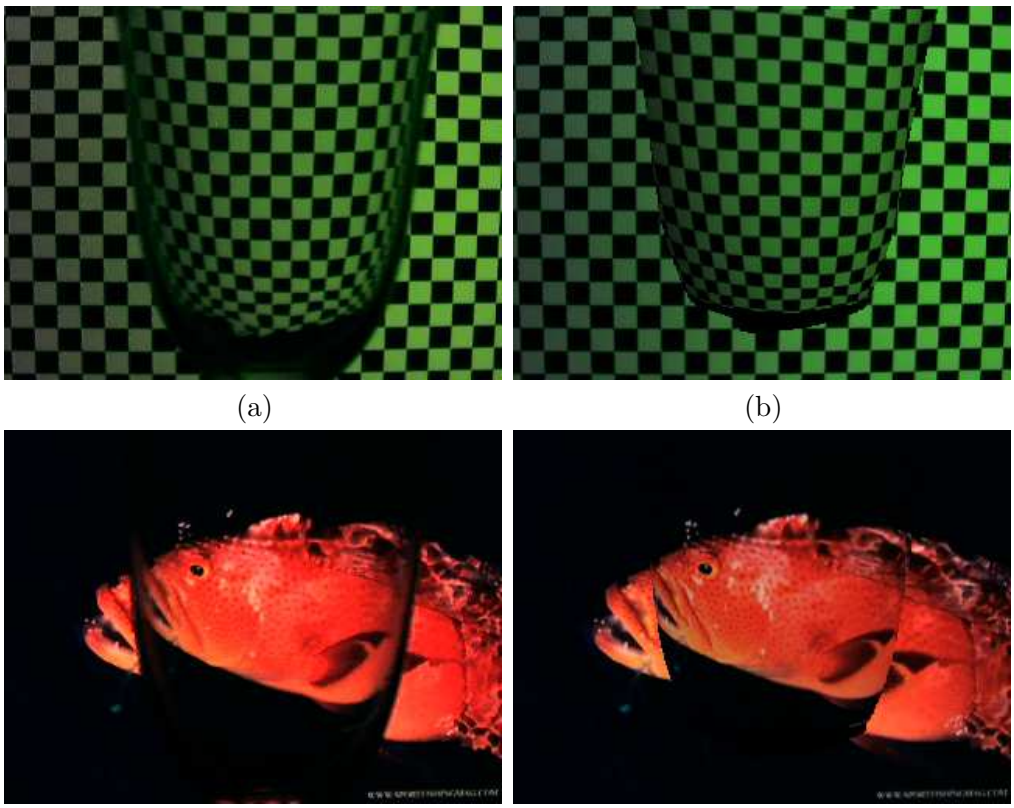
This chapter is based on "On Refractive Optical Flow" by S. Agarwal, S. P. Mallick, D. J. Kriegman, and S. J. Belongie. I played a small part in theoretical development, and was responsible for implementation and experiments.



(a) $\alpha(\mathbf{x})$ (b) $\alpha_c(\mathbf{x})$ (c) $\alpha_u(\mathbf{x})$

(d) $T(\mathbf{x})$ (e) $T_c(\mathbf{x})$ (f) $T_u(\mathbf{x})$

Figure 5.4 This figure shows compares the performance of refractive structure estimation in the calibrated and the uncalibrated case. (a) and (d) show the true warp $T(\mathbf{x})$ applied to a checkerboard pattern and the attenuation factor $\alpha(\mathbf{x})$. (b) and (e) show the estimated warp and the alpha for the calibrated case, and (c) and (f) show the estimated warp and alpha for the uncalibrated case.

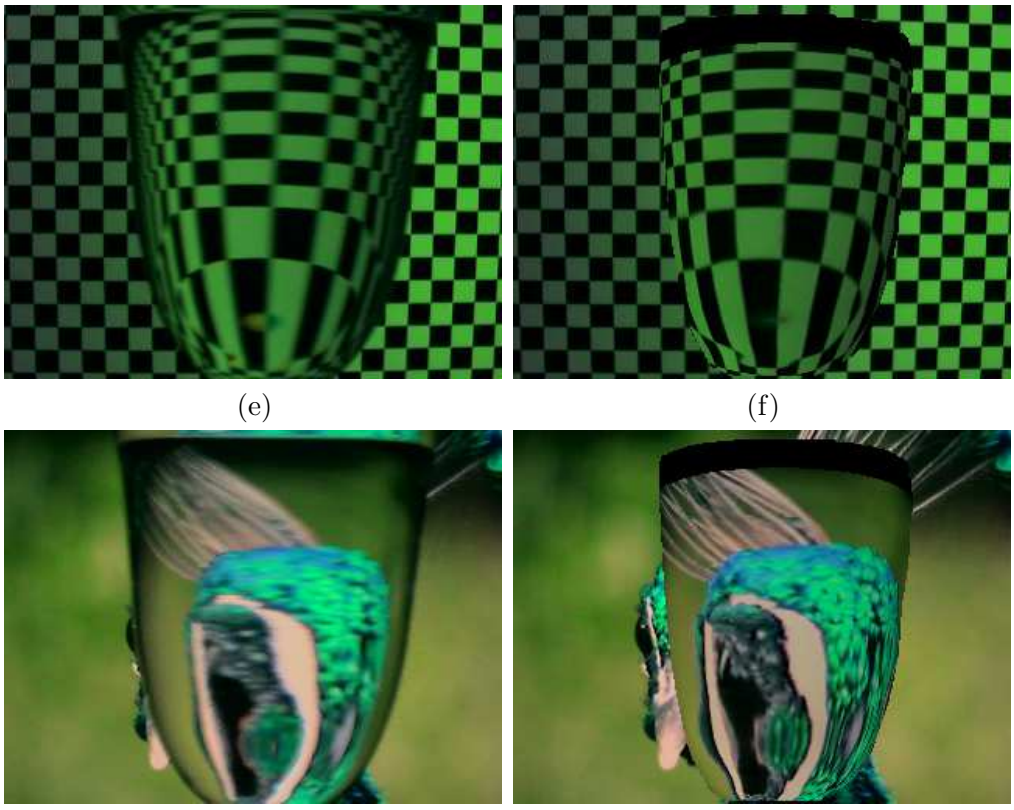


(a) (b)

(c) (d)

Without Water

Figure 5.5 Results of using the refractive structure for environment matting. (a), (c) show the true warping of a background image when an empty glass is placed in front of it, (b) and (d) show the estimated refractive structure applied to the same images.



(e) (f)

(g) (h)

With Water

Figure 5.6 Results of using the refractive structure for environment matting. (a) and (c) show the true warping of a background image when a glass filled with water is placed in front of it, (b) and (d) show the estimated refractive structure applied to the same images.

6

Conclusion and Future Work

"The black belt represents the beginning – the start of a never ending journey of discipline, work, and the pursuit of an ever-higher standard"

A Parable of the Black Belt
Built to Last: Successful Habits of Visionary Companies
James C. Collins and Jerry I. Porras

As the field of computer vision matures and vision algorithms make their way into real world applications, there is need to question long held assumptions that can potentially cripple the utility of these algorithms. This thesis provides simple, yet elegant algorithms to relax the Lambertian assumption in computer vision. More significantly, this thesis does not demand new algorithms be developed to handle non-Lambertian surfaces. Instead, it proposes algorithms that provide access to the Lambertian components of color images which can be used as inputs to Lambertian based algorithms.

At a philosophical level, our approach is in contrast to traditional approaches. While traditionally specialized algorithms were developed to handle non-Lambertian surfaces in each problem domain, this thesis argues for a unified solution to problems arising because of the Lambertian assumption. The development of unified solutions can benefit vision algorithms across the board and lead to a novel methods of relaxing the Lambertian assumption.

This thesis also clears a general misunderstanding that a complete specular-diffuse separation (an ill-posed problem) is necessary to extend algorithms for Lambertian surfaces to dichromatic surfaces. We show that we only need to solve a well-posed problem of recovering purely diffuse channel(s) of a color image. These diffuse channels can be used as inputs to Lambertian based algorithms.

In the real world, illumination can be complex due to the presence of multiple

colored light sources in a scene. While several aspects of complex illumination (for example, the proximity of light source to the object, the size of the light source(s), the number of light sources, *etc.*) have been addressed to some extent in the existing literature, the issue of multiple color illuminants, which arguably cause the most dramatic change in image appearance, is rarely dealt with. In this thesis, we propose algorithms to handle complex illumination for color as well as hyper spectral images. A class of photometric invariants are defined that are invariant to specular highlights and/or shading under complex illumination.

The theory developed in the thesis is supplemented with empirical evidence demonstrating the usefulness of the proposed algorithms. The thesis also identifies the limitations of the proposed solutions. In particular, conditions on light source color and object color are identified under which the proposed algorithms become ineffective due to decrease in signal to noise ratio in the photometric invariant images. The identification of these limitations can also serve as guidelines for carefully choosing components of computer vision systems. For example, in a photometric stereo set up for recovering the 3D shape of human face, the knowledge of the distribution of human skin colors can be utilized to carefully choose the light source color to maximize the signal to noise ratio in specular-invariant image.

A possible direction of future work is the estimation of light source color from the image(s). The estimation of the light source color using a single image is dealt with in color constancy literature. However, there are possible opportunities to get better estimates of light source color when multiple images of a scene are acquired either by moving the camera (as in the case of stereo) or by moving the light source (as in the case of photometric stereo). In the case of complex illumination the estimation of light source color provides even greater challenges.

As mentioned earlier the signal to noise ratio of invariant images is much lower than the original image. This primarily due to the fact that a portion of the diffuse component is lost while calculating the invariants. For example, the S-channel in the SUV color space contains a portion of the diffuse color, and as the light source color approaches the object color, the signal in the U and V channels approach zero. It needs to be investigated whether it is possible make geometric assumptions (for example, an upper bound on surface curvature) about the surfaces in the scene to isolate part of the diffuse component from the S-channel.

The thesis also proposes a unified PDE based approach for separating the diffuse and specular components of textureless and textured images and videos. To the best

of our knowledge, this is the only algorithm in its class that generalizes to videos and takes into account the temporal coherence of neighboring frames by treating videos as a spatio-temporal volume of pixels. A novel application of specular-diffuse separation called *Dichromatic Editing* is presented that independently processes and then combines the two dichromatic components to produce a variety of visual effects. In particular we show that this reflectance aware processing allows a digital artist to manipulate the appearance of objects in a digital image in physically meaningful ways (for example, broadening the specular lobe of the BRDF by dilating the separated specular layer).

While, the theory for specular-diffuse separation is cast as an anisotropic erosion equation, there are several other approaches using which the same equation may be derived. Due to a close link between Differential Morphology, Level Set Methods, and Curve Evolution, it would be a good theoretical exercises to derive the same equations using the other two frameworks as part of future work. In addition to providing greater theoretical insight, and showing mathematical harmony, such an exercise might help come up with methods for faster convergence and greater stability of the PDE.

Saturation of pixels is currently not handled explicitly in the PDE; although in the implementation, saturated regions are inpainted (*i.e.* information from other parts of the image is borrowed to fill in saturated regions). Inpainting is the best available option when all three channels of a pixel are saturated. However, when only one or two channels of a pixel are saturated, new methods need to be developed so as to incorporate unsaturated channels of the pixel in the estimation procedure.

Finally, the thesis ventures into a difficult and neglected territory of problems involving refraction. We show that it is possible to recover the path of light transport as it moves from a background plane to the image plane of a camera through a refracting object by observing the differential motion of the warped background. This work generalizes the optical flow equation to the case when light rays bend (due to reflection or refraction) as they travel from the object to the camera. Results of the algorithm are shown using Environment Matting, where a refractive objected is matted into a scene.

There are several avenues of future work, as not many vision algorithms deal with refraction. In our current framework, a point in the background image maps to a point the image plane. However, if the scene in which the refractive object needs to be matted is not planar (for example, when a 3D model of the scene is available), we need mappings of rays from background 3D structure to point in the image plane. This may be accomplished by finding the point to point mappings corresponding to two background planes separated by a distance. Yet another interesting direction of future

work is to recover the 3D shape of a refractive material using motion cues, or by using multiple views.

Bibliography

[1] Middlebury Stereo Page. `http://www.middlebury.edu/stereo`.

[2] A. H. Ahmed and A. A. Farag. A new formulation for shape from shading for non-lambertian surfaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1817–1824, Washington, DC, USA, 2006. IEEE Computer Society.

[3] L. Alvarez, F. Guichard, P. Lions, and J. Morel. Axiomatisation et nouveaux operateurs de la morphologie mathematique. *C. R. Acad. Sci., Paris*, pages 265–268, 1992.

[4] R. Andrews and B. Lovell. Color optical flow. In *Proc. of the APRS Workshop on Digital Image Computing*, Brisbane, 2003.

[5] M. Ashikhmin, S. Premoze, and P. Shirley. A microfacet-based BRDF generator. In *Proc. SIGGRAPH*, pages 65–74, 2000.

[6] B. K. P. Horn. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, MIT, 1970.

[7] R. Bajcsy, S. Lee, and A. Leonardis. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *IJCV*, 17(3):241–272, March 1996.

[8] S. Bakshi and Y.-H. Yang. Shape from shading for non-lambertian surfaces. In *Proc. IEEE Int. Conf. Image Processing*, volume 2, pages 130–134, 1994.

[9] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms–part I: Methodology and experiments with synthesized data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(9), 2002.

[10] S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, 1982.

[11] J. L. Barron and R. Klette. Quantitative color optical flow. In *Proc. International Conference on Pattern Recognition*, volume 4, pages 251–255, Washington, DC, USA, August 2002. IEEE Computer Society.

[12] S. Barsky and M. Petrou. Colour photometric stereo: Simultaneous reconstruction of local gradient and colour of rough textured surfaces. In *Proc. Int. Conf. on Computer Vision*, pages 600–605, 2001.

[13] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, October 2003.

[14] J. Besag. On the statisical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48:259–302, 1986.

[15] D. Bhat and S. Nayar. Stereo and specular reflection. *Int. Journal of Computer Vision*, 26(2):91–106, February 1998.

[16] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. In *Proc. Int. Conf. on Computer Vision*, pages 1073–1080, 1998.

[17] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. on Computer Vision*, pages 231–236, 1993.

[18] A. Blake and G. Brelstaff. Geometry from specularities. In *ICCV*, pages 394–403, 1988.

[19] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9), september 2003.

[20] R. Bolles, H. Baker, and D. Marimont. Epipolar-plane image anlysis: An approach to determining structure from motion. *IJCV*, 1:7–55, 1987.

[21] R. Boomgaard. *Mathematical morphology: Extensions towards computer vision.* PhD thesis, University of Amsterdam, 1992.

[22] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 648–655, 1998.

[23] G. Brelstaff and A. Blake. Detecting specular reflection using lambertian constraints. In *Proc. Int. Conf. on Computer Vision*, pages 297–302, 1988.

[24] R. Brockett and P. Maragos. Evolution equations for continuous scale morphology. *IEEE trans. on Sig. Proc.*, 42:3377–3386, 1994.

[25] Y. Y. Chuang, D. E. Zongker, J. Hindorff, B. Curless, D. H. Salesin, and R. Szeliski. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proc. of ACM SIGGRAPH 2000*, pages 121–130. ACM Press, July 2000.

[26] E. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Vision, Graphics and Image Processing*, 18(4):309–328, April 1982.

[27] R. Cook and K. Torrance. A reflectance model for computer graphics. In *Proc. SIGGRAPH*, pages 307–316, 1981.

[28] M. G. Crandall, H. Ishii, and P. L. Lions. *User's Guide to Viscosity Solutions of Second Order Partial Differential Equations*, volume 26. American Mathematical Society, July 1992.

[29] J. E. Davis, R. Yang, and L. Wang. Brdf invariant stereo using light transport constancy. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 436–443, Washington, DC, USA, 2005. IEEE Computer Society.

[30] G. Finlayson and G. Schaefer. Constrained dichromatic colour constancy. In *Proc. European Conf. on Computer Vision*, volume 1, pages 342–358, 2001.

[31] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.

[32] A. Georghiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 816–823, 2003.

[33] M. D. Grossberg and S. K. Nayar. High dynamic range from multiple images: Which exposures to combine? In *Proc. IEEE Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, October 2003.

[34] P. W. Hawkes. The electron microscope as a structure projector. *Electron tomography*, pages 17–38, 1992.

[35] P. W. Hawkes and E. Kasper. *Principles of electron optics: Wave Optics*, volume 3. Academic Press, London, 1994.

[36] X. D. He, P. Heynen, R. Phillips, K. Torrance, D. Salesin, and D. Greenberg. A fast and accurate light reflection model. In *Proc. SIGGRAPH*, pages 253–254, 1992.

[37] G. Healey and T. O. Binford. Local shape from specularity. *CVGIP*, 42(1):62–86, 1988.

[38] A. Hertzmann and S. Seitz. Shape and material by example: a photometric stereo approach. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003.

[39] S. D. Hordley, G. D. Finlayson, and M. S. Drew. Removing shadows from images. In *Proc. European Conf. on Computer Vision*, pages 823–836, 2002.

[40] B. Horn. *Robot Vision*. MIT Press, Cambridge, Mass., 1986.

[41] B. Horn. Height and gradient from shading. *Int. Journal of Computer Vision*, 5(1):37–75, 1990.

[42] B. K. P. Horn and B. G. Schunck. Determinig optical flow. *Artificial Intelligence*, 17(1-3), August 1981.

[43] B. K. P. Horn and B. G. Schunck. Determining optical flow: a retrospective. In *Artificial intelligence in perspective*, pages 81–87, Cambridge, MA, USA, 1994. MIT Press.

[44] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3(6):661–669, 1981.

[45] K. Ikeuchi and B. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.

[46] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motion. *International Journal of Computer Vision*, 12(1):5–16, 1994.

[47] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004. To Appear.

[48] H. Jin, S. Soatto, and A. J. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *Int. J. Comput. Vision*, 63(3):175–189, 2005.

[49] S. X. Ju, M. J. Black, and A. D. Jepson. Skin and bones: Multilayer, locally affine, optical flow and regularization with transparency. In *CVPR 1996*, pages 307–314, June 1996.

[50] J. Kim, V. Kolmogorov, and R. Zabih. "visual correspondence using energy minimization and mutual information", 2003.

[51] G. Klinker, S. Shafer, and T. Kanade. The measurement of highlights in color images. *Int. Journal of Computer Vision*, 2(1):7–32, 1988.

[52] J. Koenderink and A. van Doorn. Bidirectional reflection distribution function expressed in terms of surface scattering modes. In *Proc. European Conf. on Computer Vision*, pages II:28–39, 1996.

[53] M. Koudelka, S. Magda, P. Belhumeur, and D. Kriegman. Image-based modeling and rendering of surfaces with arbitrary BRDFs. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 568–575, 2001.

[54] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *Proc. Int. Conf. on Computer Vision*, pages 1448–1455, Washington, DC, USA, 2005. IEEE Computer Society.

[55] H.-S. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *J. Optical Society of America A*, 3(10):1694–1699, October 1986.

[56] A. Levin, A. Zomet, and Y. Weiss. Learning to perceive transparency from the statistics of natural scenes. In *Neural Information Processing Systems, 2002*, 2002.

[57] Y. Li, S. Lin, H. Lu, S. B. Kang, and H.-Y. Shum. Multibaseline stereo in the presence of specular reflections. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 3*, page 30573, Washington, DC, USA, 2002. IEEE Computer Society.

[58] J. Lim, J. Ho, M.-H. Yang, and D. Kriegman. Passive photometric stereo from motion. In *Proc. Int. Conf. on Computer Vision*, 2005.

[59] S. Lin, Y. Li, S. Kang, X. Tong, and H. Shum. Diffuse-specular separation and depth recovery from image sequences. In *ECCV*, page III: 210 ff., 2002.

[60] S. Lin, Y. Li, S. B. Kang, X. Tong, and H.-Y. Shum. Diffuse-specular separation and depth recovery from image sequences. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part III*, pages 210–224, London, UK, 2002. Springer-Verlag.

[61] J. Lu and J. Little. Reflectance and shape from images using a collinear light source. *Int. Journal of Computer Vision*, 32(3):1–28, 1999.

[62] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision (ijcai). In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pages 674–679, April 1981.

[63] S. Ludtke, P. Baldwin, and W. Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, 122:82–97, 1999.

[64] S. Ludtke, J. Jakana, J. Song, D. Chuang, and W. Chiu. An 11.5 A Single Particle Reconstruction of GroEL Using EMAN. *Journal of Molecular Biology*, 314:241–250, 2001.

[65] S. Magda, D. Kriegman, T. Zickler, and P. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *Proc. Int. Conf. on Computer Vision*, pages II: 391–398, 2001.

[66] S. P. Mallick, S. Agarwal, D. J. Kriegman, S. J. Belongie, B. Carragher, and C. S. Potter. Structure and View Estimation for Tomographic Reconstruction: A Bayesian Approach. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume II, pages 2253–2260, New York, June 2006.

[67] S. P. Mallick, B. Carragher, C. S. Potter, and D. J. Kriegman. ACE: Automated CTF Estimation. *Ultramicroscopy*, 104:8–29, August 2005.

[68] S. P. Mallick, Y. Zhu, and D. J. Kriegman. Detecting particles in cryo-EM micrographs using learned features. *Journal of Structural Biology*, 145:52–62, Jan/Feb 2004.

[69] S. P. Mallick, T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman. Specularity Removal In Images and Videos: A PDE Approach. In *Proc. European Conf. on Computer Vision*, 2006.

[70] P. Maragos and M. A. Butt. Curve evolution and differential morphology. *Fundamenta Informaticae*, 41:91–129, 2000.

[71] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietikäinen. Physics-based database for color research. *Journal of Electronic Imaging*, 9(1):32–38, January 2000.

[72] K. McHenry, J. Ponce, and D. Forsyth. Finding glass. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, 2005.

[73] H. Murase. Surface shape reconstruction of a nonrigid transparent object using refraction and motion. *IEEE. Trans. on Pattern Analysis and Machine Intelligence*, 14(10):1045–1052, October 1992.

[74] S. G. Narasimhan, V. Ramesh, and S. K. Nayar. A class of photometric invariants: Separating material from shape and illumination. In *Proc. Int. Conf. on Computer Vision*, volume 2, pages 1387–1394, 2003.

[75] S. Nayar, X. Fang, and T. Boult. Separation of reflection components using color and polarization. *Int. Journal of Computer Vision*, 21(3):163–186, 1997.

[76] S. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE J. of Robotics and Automation*, 6(4):418–431, 1990.

[77] S. K. Nayar and M. Bolle. Reflectance based object recognition. *Int. Journal of Computer Vision*, 17(3):219–240, 1996.

[78] M. Oren and S. Nayar. Generalization of the Lambertian model and implications for machine vision. *Int. Journal of Computer Vision*, 14:227–251, 1996.

[79] M. Oren and S. Nayar. A theory of specular surface geometry. *IJCV*, 24(2):105–124, September 1997.

[80] M. Osadchy, D. Jacobs, and R. Ramamoorthi. Using specularities for recognition. In *ICCV*, pages 1512–1519, 2003.

[81] J. Park and J. Tou. Highlight separation and surface orientation for 3-D specular objects. In *ICPR90*, pages Vol–I 331–335, 1990.

[82] J. B. Park. Efficient color representation for image segmentation under nonwhite illumination. In *SPIE, Volume 5267*, pages 163–174, 2003.

[83] A. Pentland. Local shading analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(2):170–187, March 1984.

[84] A. Pentland. Shape information from shading: A theory about human perception. *Proc. Int. Conf. on Computer Vision*, 88:404–413, 1988.

[85] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *PAMI*, 12(7):629–639, 1990.

[86] P. Perona, J. Malik, and T. Shiota. *Geometry-Driven Diffusion in Computer Vision*, chapter Anisotropic Diffusion, pages 73–92. Kluwer, Amsterdam, 1995.

[87] B. Phong. Illumination for computer-generated pictures. *Comm. of the ACM*, 18(6):311–317, 1975.

[88] T. Ping, S. Lin, L. Quan, and H.-Y. Shum. Highlight removal by illumination-constrained inpainting. In *ICCV*, pages 164–169, Nice, France, 2003.

[89] E. Prados and O. Faugeras. Shape from shading: A well-posed problem? In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages II: 870–877, 2005.

[90] H. Ragheb and E. Hancock. Separating Lambertian and specular reflectance components using iterated conditional modes. In *BMVC*, 2001.

[91] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proc. SIGGRAPH*, pages 117–128, 2001.

[92] Y. Sato and K. Ikeutchi. Temporal-color space analysis of reflection. *J. Optical Society of America A*, 11(11):2990–3002, 1994.

[93] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces. In *CVPR 2001*, pages 738–745, 2001.

[94] S. Savarese and P. Perona. Local analysis for 3d reconstruction of specular surfaces: Part ii. In *ECCV 2002*, 2002.

[95] Y. Schechner, N. Kiryati, and J. Shamir. Blind recovery of transparent and semireflected scenes. In *CVPR 2000*, volume 1, pages 38–43. IEEE Computer Society, 2000.

[96] K. Schlüns and O. Wittig. Photometric stereo for non-Lambertian surfaces using color information. In *Proc. Int. Conf. on Image Analysis and Processing*, pages 505–512, September 1993.

[97] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 519–528, Washington, DC, USA, 2006. IEEE Computer Society.

[98] S. Shafer. Using color to separate reflection components. *COLOR research and applications*, 10(4):210–218, 1985.

[99] W. Silver. Determining shape and reflectance using multiple images. Master's thesis, MIT, 1980.

[100] H. Tagare and R. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(2):133–152, February 1991.

[101] R. Tan and K. Ikeuchi. Reflection components decomposition of textured surfaces using linear basis functions. In *CVPR*, pages I: 125–131, 2005.

[102] R. T. Tan and K. Ikeuchi. Separating reflection components of textured surface using a single image. In *Proc. Int. Conf. on Computer Vision*, pages 870–877, 2003.

[103] R. T. Tan, K. Nishino, and K. Ikeutchi. Color constancy through inverse-intensity chromaticity space. *J. Optical Society of America A*, 21(3):321–334, 2004.

[104] T. Tan, R, K. Nishino, and K. Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *PAMI*, 26(10):1373–1381, October 2004.

[105] Y. Tian and H. Tsui. Shape recovery from a color image for non-lambertian surfaces. *J. Optical Society of America A*, 14(2):397–404, February 1997.

[106] S. Tominga and B. Wandell. Standard surface-reflectance model and illuminant estimation. *J. Optical Society of America A*, 6(4):576–584, April 1989.

[107] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt Soc. Am.*, 57:1105–1114, 1967.

[108] P. Tsai and M. Shah. Shape from shading using linear-approximation. *Image and Vision Computing*, 12(8):487–498, October 1994.

[109] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. In *Proc. SIGGRAPH*, pages 770–779, 2003.

[110] S. Umeyama and G. Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *PAMI*, 26(5):639–647, 2004.

[111] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–87, 1989.

[112] O. E. Vaga and Y. H. Yang. Shading logic: A heuristic approach to recover shape from shading. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):592–597, 1993.

[113] J. van de Weijer and T. Gevers. Robust optical flow from photometric invariants. In *Proc. IEEE Int. Conf. Image Processing*, pages 1835–1838, 2004.

[114] P. Viola and M. Jones. Robust real-time face detection. *Int. Journal of Computer Vision*, 57(2):137–154, May 2004.

[115] H. Wann Jensen, S. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proc. SIGGRAPH*, pages 511–518, 2001.

[116] J. Weijer and T. Gevers. Robust optical flow from photometric invariants, 2004.

[117] Y. Wexler, A. W. Fitzgibbon, and A. Zisserman. Image-based environment matting. In *Proc. of the 13th Eurographics workshop on Rendering*, 2002.

[118] V. Willert, J. Eggert, S. Clever, and E. Körner. Probabilistic color optical flow. In *DAGM-Symposium*, pages 9–16, 2005.

[119] L. Wolff and E. Angelopoulou. Three-dimensional stereo by photometric ratios. *J. Optical Society of America A*, 11:3069–3078, 1994.

[120] L. B. Wolff and T. E. Boult. Constraining object features using a polarization reflectance model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):635–657, 1991.

[121] R. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intesity. In *Proc. SPIE*, volume 155, pages 136–143, 1978.

[122] R. Yang, M. Pollefeys, and G. Welch. Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 576, Washington, DC, USA, 2003. IEEE Computer Society.

[123] K. Yoon and I. Kweon. Correspondence search in the presence of specular highlights using specular-free two-band images. In *Proc. Asian Conf. on Computer Vision*, pages II:761–770, 2006.

[124] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[125] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proc. Int. Conf. on Computer Vision*, pages 618–625, October 2003.

[126] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah. Shape from Shading: A Survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8):690–706, August 1999.

[127] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(7):680–702, 1991.

[128] Y. Zhu, B. Carragher, R. M. Glaeser, D. Fellmann, C. Bajaj, M. Bern, F. Mouche, F. de Haas, R. J. Hall, D. J. Kriegman, S. J. Ludtke, S. P. Mallick, P. A. Penczek, A. M. Roseman, F. J. Sigworth, N. Volkmann, and C. S. Potter. Automatic particle selection: results of a comparative study. *Journal of Structal Biology*, 145:3–14, Jan-Feb 2004.

[129] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *Int. Journal of Computer Vision*, 49(2/3):215–227, September 2002.

[130] T. Zickler, P. Belhumeur, and D. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *Proc. European Conf. on Computer Vision*, pages III: 869–884, May 2002.

[131] T. E. Zickler, J. Ho, D. J. Kriegman, J. Ponce, and P. N. Belhumeur. Binocular helmholtz stereopsis. In *Proc. Int. Conf. on Computer Vision*, page 1411, Washington, DC, USA, 2003. IEEE Computer Society.

[132] D. E. Zongker, D. M. Werner, B. B Curless, and D. H. Salesin. Environment matting and compositing. In *Proc. of ACM SIGGRAPH 99*, pages 205–214. ACM Press, July 1999.