UNIVERSITY OF CALIFORNIA, SAN DIEGO

Models and Methods for Recovering Shape, Reflectance, and Illumination From Images

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

in

Computer Science and Engineering

by

Neil Gordon Alldrin

Committee in charge:

Professor David Kriegman, Chair
Professor Serge Belongie
Professor Mohan Trivedi
Professor Nuno Vasconcelos
Professor Matthias Zwicker

2008

The dissertation of Neil Gordon Alldrin is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

DEDICATION

This document is dedicated to my wife Kristen, whose love, support, and patience helped make this possible.

EPIGRAPH

Men are born soft and supple; dead, they are stiff and hard. Plants are born tender and pliant; dead, they are brittle and dry. Thus whoever is stiff and inflexible is a disciple of death. Whoever is soft and yielding is a disciple of life. The hard and stiff will be broken. The soft and supple will prevail. —Lao-tzu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The work in Chapter 4 first appeared in the proceedings of the 2007 Conference on Computer Vision and Pattern Recognition (CVPR) as (Alldrin et al., 2007). Chapter 5 first appeared in the proceedings of the 2007 International Conference on Computer Vision (ICCV) as (Alldrin and Kriegman, 2007). Chapter 6 first appeared in CVPR 2008 as (Alldrin et al., 2008). Chapter 8 first appeared in CVPR 2006 as (Alldrin and Kriegman, 2006).

VITA

2002	Bachelor of Science, University of California, Berkeley
2005	Master of Science, University of California, San Diego
2008	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

N. Alldrin, T. Zickler, & D. Kriegman. "Photometric Stereo With Non-Parametric and Spatially-Varying Reflectance." In Proc. of 2008 Conf. on Comp. Vision and Pattern Recognition (CVPR), June, 2008.

N. Alldrin & D. Kriegman. "Toward Reconstructing Surfaces With Arbitrary Isotropic Reflectance : A Stratified Photometric Stereo Approach." In Proc. of 2007 International Conf. on Computer Vision (ICCV), October, 2007.

N. Alldrin, S. Mallick, & D. Kriegman. "Resolving the Generalized Bas-Relief Ambiguity by Entropy Minimization." In Proc. of 2007 Conf. on Comp. Vision and Pattern Recognition (CVPR), June, 2007.

N. Alldrin & D. Kriegman. "A Planar Light Probe." In Proc. of 2006 Conf. on Comp. Vision and Pattern Recognition (CVPR), June, 2006.

D. Reznik, J. Canny, & N. Alldrin. "Leaving on a Plane Jet." In Proc. of 2001 Int. Conf. on Intell. Robots and Systems (IROS), October 2001.

FIELDS OF STUDY

Major Field: Computer Science and Engineering
Studies in Computer Vision
Professors David Kriegman and Serge Belongie

ABSTRACT OF THE DISSERTATION

Models and Methods for Recovering Shape, Reflectance, and Illumination From Images

by

Neil Gordon Alldrin

Doctor of Philosophy in Computer Science and Engineering

University of California, San Diego, 2008

Professor David Kriegman, Chair

Recovery of scene shape, reflectance, and illumination are of fundamental importance to computer vision. However, the image formation process involves complex interactions between all three components, making inference difficult or impossible in the absence of simplifying models or prior scene knowledge. Unfortunately, real-world scenes often violate these approximations, leading to biased or incorrect reconstructions. Thus, there is a constant struggle between model complexity and tractability. In this dissertation, new models and methods are presented for recovering shape, reflectance, and illumination which are valid for broader classes of scenes than competing techniques. Underlying all of the research presented is the ability to handle objects with complex reflectance.

First, a novel approach is presented for resolving the generalized bas-relief ambiguity which arises in uncalibrated photometric stereo. Previous work showed that the ambiguity can be resolved for textureless objects; however, as shown in this disserta-

tion, it is also possible to resolve the ambiguity for textured objects, provided there is statistical regularity in the distribution of albedo values across the surface.

Next, a photometric stereo algorithm is presented that is capable of handling nearly arbitrary reflectance. The main contribution is the utilization of bilateral symmetry in the reflectance function, a property shared by most real-world materials. By explicitly utilizing symmetry, surface shape can be constrained without relying on parametric models; a significant advance over most photometric stereo algorithms which depend on simple parametric models of surface reflectance, such as the Lambertian model.

Another photometric stereo algorithm is also presented that is capable of fully recovering the surface shape as well as the reflectance function across the surface. While a few additional constraints are necessary, this is one of only a handful of photometric stereo methods capable of simultaneously recovering shape and complex reflectance; of these, the reflectance model is by far the least restrictive.

It is also shown that bilateral symmetry of the reflectance function can be exploited for multi-view shape reconstruction. The method presented handles both textured and textureless surfaces and is capable of recovering surface concavities. Finally, a novel technique for measuring illumination is presented which relies on spatially varying reflectance.

Chapter 1

Introduction

Computer vision can be defined as the inference of information about a scene from images of the scene. Potentially useful scene information includes 3D geometry (shape reconstruction), the presence and location of specific objects (object detection and recognition), illumination incident on the scene (lighting estimation), surface reflectance properties (BRDF estimation), motion understanding of video, and camera position and intrinsic parameters (camera calibration). A standard approach to solving such inference problems involves breaking the problem into three components: (1) a model for the properties to be inferred, (2) features taken from the image data, and (3) an algorithm that finds parameters of the model that best fit the features. Each component is critical to the performance and usefulness of a given method.

In this dissertation, we are interested in recovering the shape, reflectance, and illumination of a scene or objects in a scene. Recovering these properties is fundamental to computer vision and ripe with potential applications. For example, in computer graphics physically accurate models of shape, material, and lighting can significantly influence the photo-realism of rendered images. Computer vision algorithms are increasingly being used to acquire these models, such as in Matusik et al. (2003) and Debevec (1998) where measurements of real-world material reflectance and real-world illumination environments were respectively utilized; two benefits of such techniques are im-

proved realism of rendered images and reduction of tedious hand-modeling by artists. Another important application is to "capture" culturally significant artifacts and locations around the world, both for historical posterity and for remote interaction. Recent projects involving 3D digitization of historical artifacts include the Digital Michelangelo Project (Levoy et al., 2000), reconstruction of the Parthenon (Debevec, 2004a,b; Stumpfel et al., 2003), and the the Great Buddha Project (Ikeuchi et al., 2007).

One of the primary contributions of our work is that we utilize features that are as general as possible. To illustrate what we mean by "general", consider collecting a set of images of an object, in which the scene varies only in the position of a point light source illuminating the object. *Photometric stereo* is a class of algorithms that seeks to recover the shape of an object given such images as input. If we assume that the object's surface reflects light equally in all directions (known as *Lambertian* reflectance), then there is a very simple algorithm capable of recovering the object's shape. However, if the object is, say, a glazed ceramic teapot which does not satisfy the Lambertian assumption, then the shape estimated by this method will be biased. Ideally, we would like to use an algorithm that is unbiased for a wide variety of objects, from highly specular glazed ceramic teapots to diffuse clay bunny statues (see Figure 1.1). In Chapters 5 and 6 we present photometric stereo algorithms that are unbiased for any object with *isotropic* reflectance – a property that holds for the majority of objects encountered in everyday life.

The example just presented highlights a conundrum in computer vision: utilizing prior knowledge can greatly simplify inference problems, but also can decrease the applicability of a given method since only scenes satisfying the assumptions lead to correct results. For this reason, it is important to exploit *domain knowledge* whenever possible, since domain knowledge captures properties that are generally applicable. At a physical level, the laws of physics dictate the transport, emission, and absorption of photons which are then measured as images in a camera. Moreover, most scenes encountered in everyday life are similar in many respects : they usually consist of a set of objects separated by a transparent and minimally interactive medium; the objects are

Figure 1.1: (Left) The original Utah teapot, made of glazed ceramic. (Right) The original Stanford bunny, made of terra cota clay. Images from (Astor, 2008) and (Levoy, 2008) respectively.

usually opaque (*i.e.*, they absorb and reflect, but do not transmit light); and illumination often results from a small set of *emitting* light sources.

In this dissertation, we are concerned with three complimentary goals: (a) reconstruction of the shape of an object or scene, (b) estimation of the reflectance properties across a surface, and (c) estimation of the light sources and illumination in a scene. These three components, briefly described in the following sections, provide a nearly complete description of any scene. The overarching contribution of this work is to provide new models and methods for handling and utilizing complex reflectance.

1.1 Shape Reconstruction

Shape reconstruction, as the name implies, involves recovering the shape of an object or scene from images. Three common computer vision approaches to recovering shape are stereo, structure from motion (SFM) and photometric stereo. Stereo, or more generally, multi-view reconstruction algorithms utilize images taken from multiple viewpoints. If a point on the surface of the object is visible in two or more images and the location of the point is known in each of those images, then 3D position can be

recovered by triangulation. Thus, recovering the shape of an object reduces to finding corresponding points in each of the images. Structure from motion is closely related to multi-view reconstruction, except that motion of the scene or camera over time is utilized instead of multiple viewpoints. Moreover, SFM is typically applied to video sequences where motion between frames is usually small. Photometric stereo algorithms utilize images taken from *the same* viewpoint, but with varying illumination. For each point on the surface of an object, the amount of light reflected toward the camera is a function of the light hitting the object at that point, as well as the *surface normal* at the point. Thus, with sufficient knowledge of the illumination and surface reflectance, the surface normal at each point on the object can be recovered. Chapters 4 to 6 present novel photometric stereo algorithms that recover the shape of objects. Chapter 7 presents a multi-view shape reconstruction algorithm that utilizes photometric constraints.

1.2 Reflectance Estimation

Reflectance estimation involves measuring or recovering a model for how an object's surface reflects light. Typically, the reflectance at a point on the surface of an object is modeled with a *bi-direction reflectance distribution function* (BRDF), which specifies how light incident on the surface is reflected or scattered in each exitant direction (Nicodemus et al., 1922). Traditionally, BRDFs were measured with *gonioreflectometers* – special devices that measure the response of incoming and outgoing light applied to a planar patch of material. These devices, such as the Cornell gonioreflectometer (Foo, 1996), are highly accurate, but require significant time to sample the BRDF of a material since only a single point in the BRDF domain is captured at each measurement. For this reason, image-based measurement techniques such as Matusik et al. (2003) have become much more common in recent years. The methods presented in Chapters 4 and 6 recover shape and reflectance simultaneously.

1.3 Illumination Estimation

Illumination or lighting estimation involves recovering or measuring the lighting environment in a scene. For example, in an outdoor setting the primary illumination source is the sun, but the sky and reflections of light off of other surfaces in the scene also contribute to the lighting environment. The *plenoptic* function, which specifies the intensity of light in every direction at each point in a scene, is perhaps the most general illumination model used in computer vision (Levoy and Hanrahan, 1996; Gortler et al., 1996). However, it is often reasonable to assume that the illumination does not vary spatially; for example, this is approximately true when all light sources are far away from objects in the scene. In this case, illumination can be modeled by an *environment map* that records the intensity of light in each direction (Blinn and Newell, 1976). The simplest way to measure an environment map is to capture an image of a mirrored sphere placed in the scene. However, this method is not always possible or practical. In Chapter 4 we present a photometric stereo algorithm that recovers both shape and illumination simultaneously and in Chapter 8 we present a novel type of light probe that overcomes a shortcoming of the standard mirrored sphere approach.

Chapter 2

Background

Much of the theory presented in this dissertation requires basic knowledge of the physical processes underlying image formation, as well as an understanding of standard assumptions and mathematical models widely used within the computer vision research community. Specifically, the reader should have a basic understanding of radiometry, the interaction of light and matter, methods for measuring light, and models for representing surfaces. In the following sections we provide this necessary background, as well as introduce notation used throughout the dissertation.

2.1 Radiometry

The field of *radiometry* studies the transport and measurement of electromagnetic radiation, including visible light. The quantities used in radiometry are heavily employed in computer vision and graphics as they provide the foundation for how images are formed. The field of *photometry* – which actually predates radiometry – is closely related, but studies how *humans* respond to light. This section borrows heavily from Glassner (1994), which the reader may wish to reference for a more comprehensive review of radiometry.

Electromagnetic radiation can be thought of as the passage of energy through a

Table 2.1: Radiometric quantities.

Symbol	Name	Units
Q	Radiant energy	J (Joule)
u	Radiant flux density	J/m^3
Φ	Radiant power (flux)	$W = J/s$ (Watt)
W	Radiant power density	W/m^2
E	Irradiance	W/m^2
M	Radiant exitance	W/m^2
I	Radiant intensity	W/sr
L	Radiance	$W/(sr \cdot m^2)$

medium. Because of the discrete nature of energy, we define a *photon* to be a particle carrying a single quanta of energy. Due to the wave / particle duality of matter and energy, photons exhibit both wave and particle-like properties. Thus, associated with each photon is a wavelength λ. Photons always travel at the speed of light, c, and can only interact with matter by transferring an amount of energy $\frac{hc}{\lambda}$, where h is Planck's constant. Additionally, photons carry *momentum* and *polarization* states.

The most fundamental unit of radiometry is *radiant energy* Q, measured in Joules (J). Each photon carries some amount of radiant energy. The *radiant energy density* w is defined as the amount of radiant energy per unit volume,

$$w = Q/V. \tag{2.1}$$

Since we are concerned with the movement of energy, we define *radiant power* or *radiant flux* Φ as the amount of energy flowing through a surface per unit time,

$$\Phi = dQ/dt. \tag{2.2}$$

The incident or departing flux per unit of surface area is called *radiant flux area density* u,

$$u = d\Phi/dA. \tag{2.3}$$

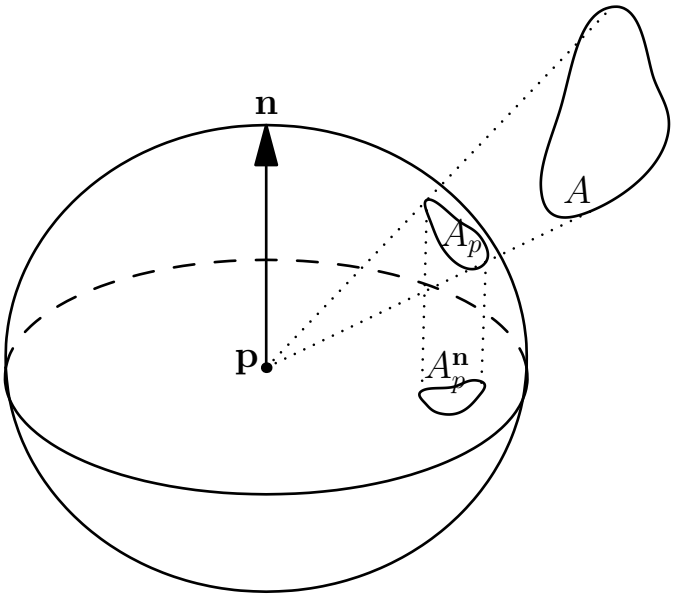


Figure 2.1: The solid angle A_p and projected solid angle $A_p^{\mathbf{n}}$ of a surface patch A.

We further distinguish two types of radiant flux density. If energy is arriving at a surface we call it *irradiance* E; if energy is departing a surface we call it *radiant exitance* M.

Before proceeding in the discussion, we need to make a brief detour to define the concepts of *solid angle* and *projected solid angle*, as well as discuss finite and differential solid angles and area patches. A *solid angle* is the 3-dimensional equivalent of a standard angle. Consider a point $\mathbf{p}$ surrounded by a unit sphere (see Figure 2.1). Further, consider a surface patch A and it's projection onto the sphere through point $\mathbf{p}$. Then the solid angle subtended by the surface patch with respect to $\mathbf{p}$ is simply the area A_p occupied on the sphere by the projection of A onto $\mathbf{p}$. We can augment our figure so that $\mathbf{p}$ is *oriented* in direction $\mathbf{n}$. A *projected solid angle* is the area $A_p^{\mathbf{n}}$ occupied by a solid angle when projected onto a plane with normal $\mathbf{n}$ (typically, one envisions the plane passing through point $\mathbf{p}$, but any parallel plane will yield identical projected solid angle). Projected solid angles will prove useful, for example, when considering light incident and exitant on points on a surface with surface normal $\mathbf{n}$.

Because we are often concerned with integrating quantities incident and exitant from a point, it is useful to consider *differential* patches of solid angle. It is convenient

to utilize spherical coordinates to represent points on the unit sphere: we denote the polar angle with θ and the azimuthal angle with ϕ. If we consider an infinitely small rectangular patch of the sphere in direction (θ, ϕ) with size $(d\theta, d\phi)$, it can be shown that the area of this patch is equal to $\sin\theta d\theta d\phi$. These patches are referred to as *differential solid angles*, which we denote $d\omega$, so that,

$$d\omega = \sin\theta d\theta d\phi. \tag{2.4}$$

Analogous to finite projected solid angle, we define *projected differential solid angle* $d\omega^{\mathbf{n}}$ as the area occupied by the projection of a differential solid angle onto a plane with normal $\mathbf{n}$. If $\mathbf{n}$ is coincident with the pole of our spherical coordinate system, then it is straightforward to show that the projected differential solid angle of a patch on the sphere in direction (θ, ϕ) with size $(d\theta, d\phi)$ is,

$$d\omega^{\mathbf{n}} = \sin\theta\cos\theta d\theta d\phi. \tag{2.5}$$

Returning to our discussion of energy, we define *intensity* I as the radiant energy leaving a point in the direction Φ, per unit solid angle,

$$I = d\Phi/d\omega. \tag{2.6}$$

Finally, we define perhaps the most important radiometric quantity for computer vision, radiance. *Radiance*, L, is defined as the power arriving or leaving from a surface per unit solid angle per unit projected area. Equivalently, one can also define radiance as the power arriving or leaving from a surface per unit projected solid angle per unit area,

$$L = \frac{d^2\Phi}{dA^{\Phi}d\omega} = \frac{d^2\Phi}{dAd\omega^{\Phi}}. \tag{2.7}$$

Radiance is important because its value does not change as a function of distance. For example, consider a point light source illuminating a surface patch. If the patch is moved away from the light, the power received by the patch will decrease proportional to the distance squared, but the radiance incident on the patch remains the same because the definition of radiance exactly compensates for this falloff. This invariance to distance makes radiance the preferred quantity when describing the transport of light and significantly simplifies analysis since one can model light as directed rays of constant radiance.

2.2 Interaction of Light and Matter

In the previous section, we defined various radiometric quantities to describe the transport of light through a transparent medium. However, what happens when light interacts with matter, or moves between different transparent mediums (*e.g.*, air to water)? To answer this question, it is useful to view the world as a partitioned set of mediums, where light travels in a uniform manner in each given medium. Viewing the world in this way, we only need to define the behavior of light at the boundary of two different mediums. Examples of different mediums include empty space, air, water, rocks, tree bark, milk, etc.; basically, any material (or lack of material) that one can think of.

At an atomic or microscopic scale, photons can interact in very complicated ways at the boundary of two mediums. Photons can exhibit *direct reflection*, *transmission*, *absorption* and possible *re-emission*, *scattering*, or a combination of any of these effects (Glassner, 1994). Moreover, if the new medium is not transparent, some light may transmit to the interior of the medium, get reflected and exit the second medium at a different location – this is known as *subsurface scattering*. The direction that photons travel after interacting with a boundary are also highly dependent on the *geometry* of the boundary. Light entering a boundary at an angle interacts differently than light entering in a frontal direction.

It is standard practice to differentiate between *reflectance*, where light arrives and exits on the same side of a boundary, and *transmission*, where light arrives and exits on different sides of a boundary. As we are primarily concerned with reflectance, we limit our subsequent discussion to that case. However, transmission can be described in an analogous way to reflectance.

To quantify reflectance, consider two points $\mathbf{x}_i$ and $\mathbf{x}_o$ on a surface. Now suppose light with radiance $L(\mathbf{x}_i, \omega_i)$ arrives at point $\mathbf{x}_i$ from differential solid angle $d\omega_i$ and we would like to know the radiance of light exiting point $\mathbf{x}_o$ in direction ω_o. Assuming reflectance is a linear process, we expect the exitant radiance to be proportional to the

incident *flux* $d\mathbf{\Phi} = L(\mathbf{x}_i, \omega_i) d\omega_i^{\mathbf{n}} dA$, where $d\omega_i^{\mathbf{n}}$ is the projected differential solid angle of the incident light and dA is the differential surface area of a patch at point $\mathbf{x}_i$. Thus,

$$\begin{aligned} L(\mathbf{x}_o, \omega_o) &= d\mathbf{\Phi} \rho(\mathbf{x}_i, \omega_i, \mathbf{x}_o, \omega_o) \\ &= L(\mathbf{x}_i, \omega_i) \rho(\mathbf{x}_i, \omega_i, \mathbf{x}_o, \omega_o) d\omega_i^{\mathbf{n}} dA. \end{aligned} \tag{2.8}$$

The proportionality constant $\rho(\mathbf{x}_i, \omega_i, \mathbf{x}_o, \omega_o)$ is called the bi-directional scattering surface reflectance distribution function (BSSRDF) and specifies the fraction of incident to exitant energy and generally depends on both the position and direction of incident and exitant light. To account for light arriving at different positions and directions, one can simply integrate over all incident rays,

$$L(\mathbf{x}_o, \omega_o) = \int_{\mathbf{x}_i \in A} \int_{\omega_i \in \Omega} L(\mathbf{x}_i, \omega_i) \rho(\mathbf{x}_i, \omega_i, \mathbf{x}_o, \omega_o) d\omega_i^{\mathbf{n}} dA. \tag{2.9}$$

Because of it's high dimensionality, BSSRDFs are difficult to work with directly. It is therefore common to assume that all light enters and leaves from the same surface point, which eliminates dependence on $\mathbf{x}_o$. This function, denoted $\rho(\mathbf{x}, \omega_i, \omega_o)$ is called the bi-directional reflectance distribution function (BRDF). If the surface does not vary spatially, then $\mathbf{x}$ can also be dropped and the BRDF reduces to a function of incident and exitant directions, $\rho(\omega_i, \omega_o)$. Applying these simplifications to Equation 2.9 yields the commonly utilized expression,

$$L(\omega_o) = \int_{\omega_i \in \Omega} L(\omega_i) \rho(\omega_i, \omega_o) d\omega_i^{\mathbf{n}} \tag{2.10}$$

which specifies exitant radiance at a surface point in terms of the BRDF and the incident radiance over the upper hemisphere of directions.

BRDFs are typically defined relative to the *local* coordinate system of a given point on a surface. Consider Figure 2.2a. Using spherical coordinates (relative to the coordinate system attached to a differential patch on a surface), we see that any direction can be defined by two variables. Thus, we can specify incident and exitant directions as $\omega_i = (\theta_i, \phi_i)$ and $\omega_o = (\theta_o, \phi_o)$. Thus, ignoring spatial (and spectral) variation, a BRDF

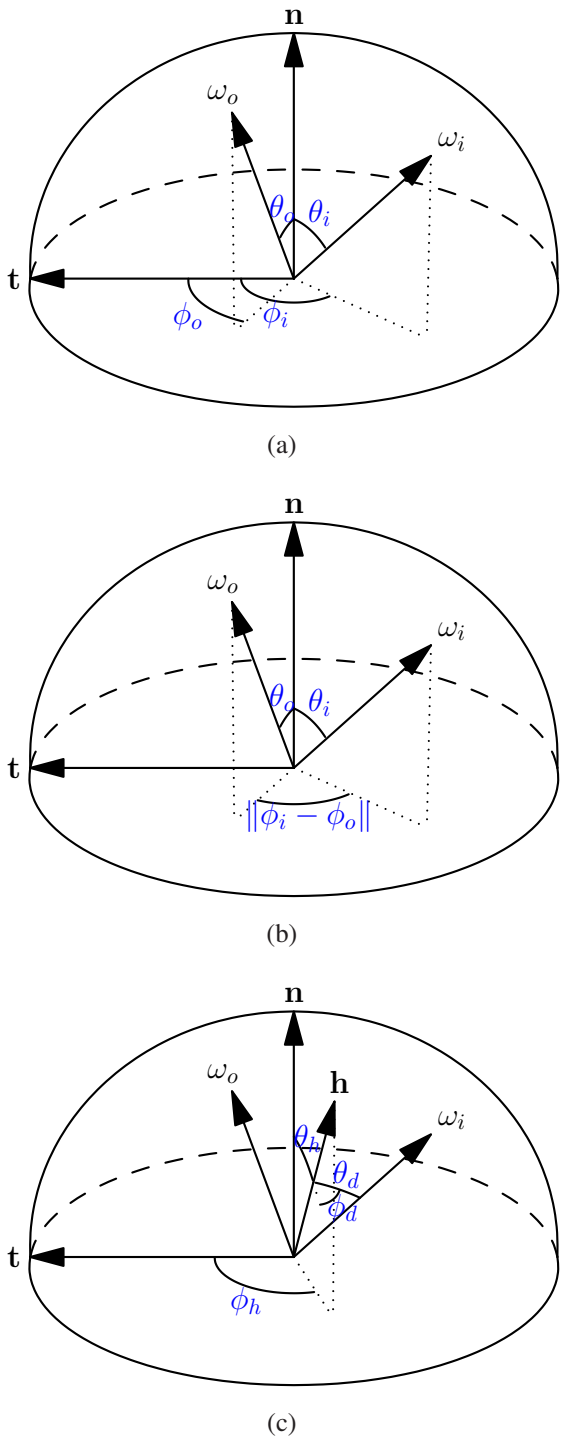


Figure 2.2: Common parametrizations of the BRDF domain. (a) The local coordinate system attached to a surface point. (b) Isotropic BRDF parameterization. (c) Half-angle parameterization.

is a function of four scalar variables: $\rho(\theta_i, \phi_i, \theta_o, \phi_o)$. Re-expressing Equation 2.10 in spherical coordinates results in the following expression,

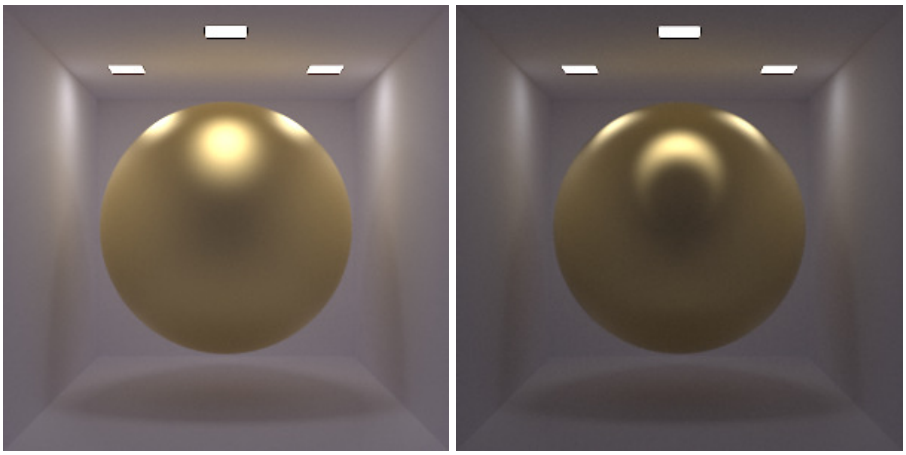
$$L(\theta_o, \phi_o) = \int_{\phi_i \in 0}^{2\pi} \int_{\theta_i \in 0}^{\frac{\pi}{2}} L(\theta_i, \phi_i) \rho(\theta_i, \phi_i, \theta_o, \phi_o) \cos\theta_i \sin\theta_i d\theta_i d\phi_i. \tag{2.11}$$

The $\cos\theta_i$ term accounts for foreshortening and $\sin\theta_i d\theta_i d\phi_i$ is the solid angle formed by the differential portion of the sphere in direction (θ_i, ϕ_i).

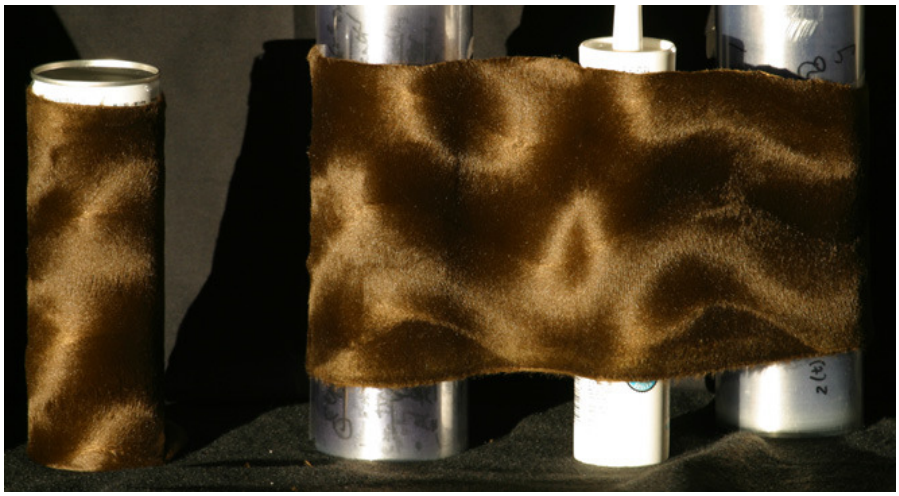
An alternative parameterization of the BRDF domain is the *half-angle* parameterization (see Figure 2.2). Empirically, it has been shown that the BRDFs of many real-world materials vary more smoothly in the half-angle domain than in the spherical coordinate parameterization (Rusinkiewicz, 1998; Stark et al., 2005). This property is useful, for example, in data-fitting and compression.

Fully general 4D BRDFs, while simpler than BSSRDFs, can still be difficult to analyze, particularly in computer vision applications attempting to infer scene properties from images. A common simplification is to assume the BRDF is invariant to rotations and reflections about the surface tangent plane. Such BRDFs are called *isotropic*, while general 4D BRDFs are referred to as *anisotropic*. Isotropic BRDFs are functions of three variables (see Figure 2.2b): $\rho(\theta_i, \theta_o, \|\phi_i - \phi_o\|)$. To make the notion of isotropy more concrete, imagine looking at a planar surface, say a piece of paper on a table. If the material is isotropic then as the sheet is rotated on the table the intensity reaching one's eye will remain constant. If the material is anisotropic (say, if the sheet is made of velvet or brushed steel), then the intensity reaching one's eye will vary as the sheet is rotated. Figure 2.3 illustrates the difference between isotropic and anisotropic materials.

To simplify BRDFs even further, it is very common to use *parametrized* models. Many parametrized models have been proposed which try to capture the appearance of materials commonly encountered in the real world. These models can be broadly grouped into *empirical models* and *physically based models*. Physically based models are derived from properties that result from physics (or approximations thereof), typically for specific classes of real-world materials. For example, the Cook-Torrance reflectance model is based on the idea of *micro-facets*, or small scale geometric variations



(a) (b)



(c)

Figure 2.3: Comparison of (a) isotropic and (b) anisotropic materials. Note the non-circular specular lobe of the anisotropic sphere. (c) A real image of anisotropic cloth (Hertzmann and Seitz, 2005).

on an object's surface. The parameters of the model dictate statistical properties of the underlying micro-facet distribution, from which a Cook-Torrance BRDF is derived[1],

$$\begin{aligned}
\rho &= d\frac{1}{\pi} + s\frac{1}{\pi}\frac{FDG}{\cos\theta_i \cos\theta_o} \\
F &= \frac{1}{2}\frac{(g-c)^2}{(g+c)^2}\left(1 + \frac{[c(g+c)-1]^2}{[c(g+c)+1]^2}\right), c = \cos\theta_d, g^2 = \eta^2 + c^2 - 1 \\
D &= ke^{-(\alpha/m)^2} \\
G &= \min\left\{1, \frac{2\cos\theta_h \cos\theta_i}{\cos\theta_d}, \frac{2\cos\theta_h \cos\theta_o}{\cos\theta_d}\right\}.
\end{aligned} \tag{2.12}$$

Empirical models, on the other hand, were developed in an ad-hoc manner and thus can yield materials that are improbable or even physically impossible (for example, some models violate energy conservation and/or Helmholtz reciprocity which are held by almost all real-world materials). However, empirical models have an advantage in that they are often designed to be simple and easy to analyze, in addition to being visually plausible to human observers. An example of an empirical model is the Phong BRDF,

$$\rho(\theta_i, \phi_i, \theta_o, \phi_o; \rho_d, k, \alpha) = \rho_d + k\cos^\alpha \theta_i. \tag{2.13}$$

In computer vision, the *Lambertian* reflectance model is by far the most utilized. The Lambertian model consists of a single parameter, the *albedo* α,

$$\rho(\theta_i, \phi_i, \theta_o, \phi_o; \alpha) = \alpha. \tag{2.14}$$

In essence, an ideal Lambertian material scatters incoming light equally in all directions, so that a viewer in any direction will observe the same exitant radiance. The reason the Lambertian model is so commonly used is that it leads to a simple linear model of image formation and is also a reasonable approximation for many materials, such as chalk, unglazed ceramics, and matte paper (see Figure 2.4).

[1] Equation adapted from (Glassner, 1994, chap. 15).

Figure 2.4: Examples of materials closely approximated by the Lambertian reflectance model.

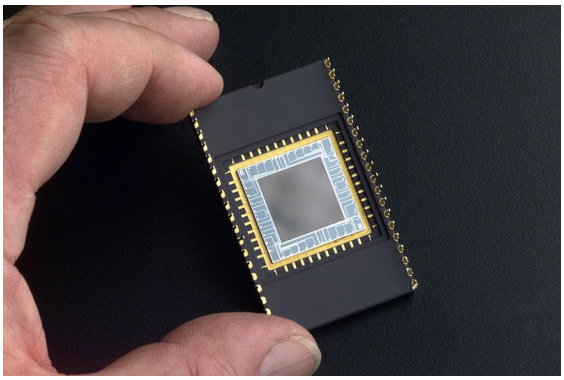


Figure 2.5: A typical CCD sensor.

2.3 Measuring Light

At this point, we have discussed models for how light travels and interacts with matter. Equally important, however, is an understanding of how light is *measured*. At the core, we want to know the radiance of light at specific locations and directions. While there are many potential measurement devices, depending on which rays one wants to measure, we are primarily concerned with *camera* systems, which measure the projection of light rays on a 2D imaging surface. In the following subsections, we cover various camera models, concluding with a discussion on practical issues that arise in real-world camera systems. A more in-depth treatment of these topics can be found in standard computer vision textbooks such as Forsyth and Ponce (2003), Trucco and Verri (1998), or Ma et al. (2005).

2.3.1 Imaging Sensors

At the heart of any camera is a sensor that records the flux / irradiance of incident light. In digital cameras, *charge coupled devices* (CCDs) or *CMOS* based sensors are typically used (see Figure 2.5), while film cameras rely on chemical reactions. CCD sensors are based on the photovoltaic effect, whereby photons striking the sensor are converted into electrical charges that can then be measured as voltages. The irradiance

of light arriving at a sensor can be derived from the voltage accumulated by the CCD over a fixed period of time (the *exposure* time of the camera). Usually, a large set of CCD (or CMOS) sensors is arranged in a regular grid pattern across the *image plane* of the camera, providing a spatially varying measurement of incident irradiance. The set of values thus obtained is commonly referred to as an *image*.

2.3.2 The Ideal Pinhole Camera Model

An ideal *pinhole camera* measures the set (or a subset) of light rays projected onto the *image plane* through a single *focal point* (see Figure 2.6). If we consider a canonical coordinate system centered at the focal point with x and y spanning the image plane and z orthogonal to the image plane, we arrive at the following projection equation,

$$\begin{aligned} x &= -fX/Z \\ y &= -fY/Z \end{aligned} \tag{2.15}$$

where $\mathbf{X} = (X, Y, Z)^\top$ is a 3D point, $\mathbf{x} = (x, y)^\top$ is its projection onto the image plane, and f is the focal length of the camera.

Such projection models are studied in the branch of mathematics known as *projective geometry*, where it is often convenient to utilize *homogeneous* coordinates. Homogeneous coordinates are related to standard Cartesian coordinates in the following way. Consider an $n+1$ dimensional vector $\mathbf{v}_{\text{homogeneous}} = (v_1, v_2, ..., v_n, w)^\top$. One can map $\mathbf{v}_{\text{homogeneous}}$ to an n dimensional vector as $\mathbf{v} = (v_1/w, v_2/w, ..., v_n/w)^\top$. Thus, any n dimensional vector can be represented in homogeneous coordinates as an $n+1$ dimensional vector. A property of homogeneous coordinates is that they are only defined up to scale; hence, $\lambda\mathbf{v} \equiv \mathbf{v}\ \forall\lambda \neq 0$. Moreover, there are special points, called *points at infinity*, that occur when $w = 0$. In homogeneous coordinates, the pinhole projection

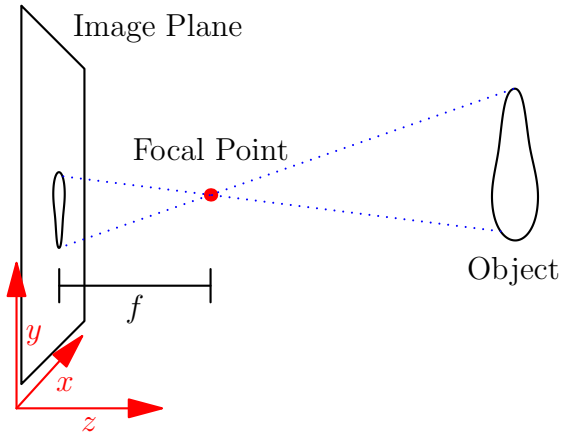


Figure 2.6: The pinhole camera projection model. Rays passing through the focal point are projected onto the image plane. The focal distance f is the distance from the focal point to the image plane.

model can be expressed as,

$$\lambda \mathbf{x} = \lambda \begin{pmatrix} -fX \\ -fY \\ Z \end{pmatrix} = K\mathbf{X} = \begin{pmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \tag{2.16}$$

Equation 2.16 requires 3D points to be expressed in the canonical camera centered coordinate frame. However, in practice it is often convenient to express these points relative to a different world coordinate frame. The camera and world coordinate frames are related to each other by a rigid transformation consisting of a 3×3 rotation matrix $\mathbf{R}$ and a 3×1 translation vector $\mathbf{t}$, which can be incorporated into Equation 2.16 as,

$$\lambda \mathbf{x} = KP\mathbf{X} = \begin{pmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{2.17}$$

The rigid transformation mapping world to camera coordinates is referred to as the *extrinsic* camera parameters.

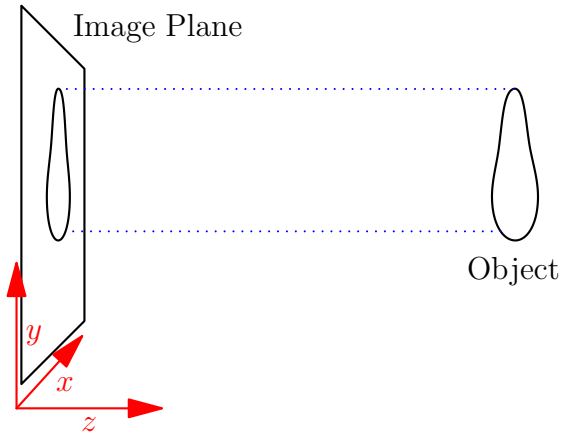


Figure 2.7: The orthographic camera projection model. Rays are projected in parallel onto the image plane in the direction orthogonal to the image plane.

2.3.3 Alternative Projection Models

The *affine* camera model results when rays are projected in parallel onto the image plane. Pinhole (and real) cameras are well approximated by affine cameras when the object being imaged is distant from the camera relative to the object's diameter. Affine projection maps points onto the image plane as,

$$\mathbf{x} = P\mathbf{X} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.18}$$

where $\mathbf{A}$ is a 2×3 rank 2 matrix and $\mathbf{b}$ is a 2×1 vector.

The *orthographic* camera model is a special case of affine projection resulting when rays are projected in the direction orthogonal to the image plane (see Figure 2.7). Orthographic projection maps points onto the image plane as,

$$\mathbf{x} = \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = P\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{2.19}$$

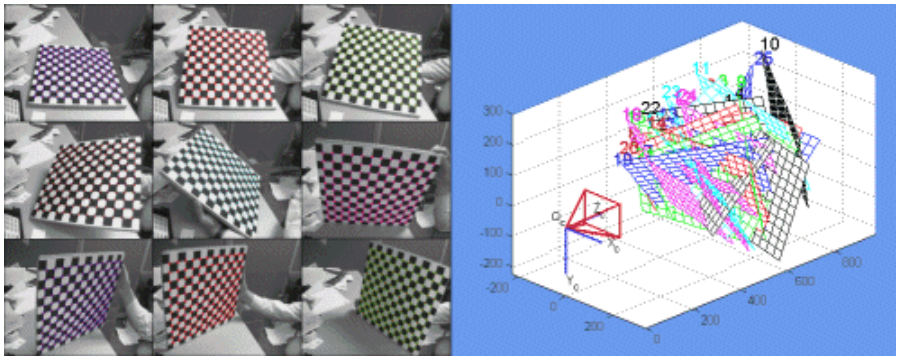


Figure 2.8: (Left) A set of imaged checkerboard patterns. (Right) Recovered extrinsic camera parameters. Image from (Bouguet, 2008).

Orthographic projection is commonly assumed in computer vision applications such as photometric stereo and structure from motion.

2.3.4 Real Camera Systems

Real cameras deviate from ideal camera models in a number of respects, mostly due to the use of lenses and non-ideal image sensors. Fortunately, most of these effects can be mitigated by appropriate calibration processes. Broadly speaking, there are two types of camera calibration: *geometric* calibration and *radiometric* calibration.

Geometric calibration involves finding a mapping between the rays of light measured by the camera and the rays that would have been recorded by an ideal pinhole camera. Geometric calibration is often subdivided into two parts: estimation of *extrinsic* camera parameters (*i.e.*, the location and orientation of the camera relative to the scene) and estimation of the *intrinsic* camera parameters that model effects such as lens distortion, focal length, translation of the optical center, skew, and aspect ratio. A thorough treatment of geometric calibration techniques is beyond the scope of this document, but a popular approach is to image a checkerboard pattern (or other pattern with known geometry) in multiple poses. This allows one to fit parameters of a camera model by comparing the measured locations of the checkerboard pattern with those predicted by the model (see Figure 2.8).

Radiometric calibration involves finding a mapping between measured pixel values and the irradiance (ideally) arriving at the image sensor. Ideally, the measured pixel intensities $I(x, y)$ should be proportional to the irradiance arriving at the sensor $E(x, y)$,

$$I(x, y) \propto E(x, y). \tag{2.20}$$

However, some sensors have a non-linear response curve f, so that

$$I(x, y) \propto f[E(x, y)]. \tag{2.21}$$

One method for recovering the response curve is to image a scene multiple times with varying shutter speed. Another problem arising in real camera systems is *vignetting*, which results in spatially varying falloff of pixel intensity across the image plane. Vignetting is primarily caused by the *aperture* and lens of the camera, which cause more light to enter at the center of the image than at the edges. Vignetting is typically modelled as a multiplicative falloff $V(x, y)$ across the image plane,

$$I(x, y) \propto V(x, y)E(x, y). \tag{2.22}$$

A simple technique to recover V is to image a scene with uniformly emitted radiance. Because vignetting falloff is often close to radially symmetric, it is also common to fit a function of the form $V(r = \sqrt{x^2 + y^2})$, which can be further simplified to a low-order polynomial $V(r) = a + br^2 + cr^4 + dr^6$.

2.4 Surface Models

An issue of critical importance to any shape reconstruction algorithm is how to mathematically represent shape. There are many possible representations, but here we present three of the most commonly utilized representations – height maps, meshes, and voxel grids.

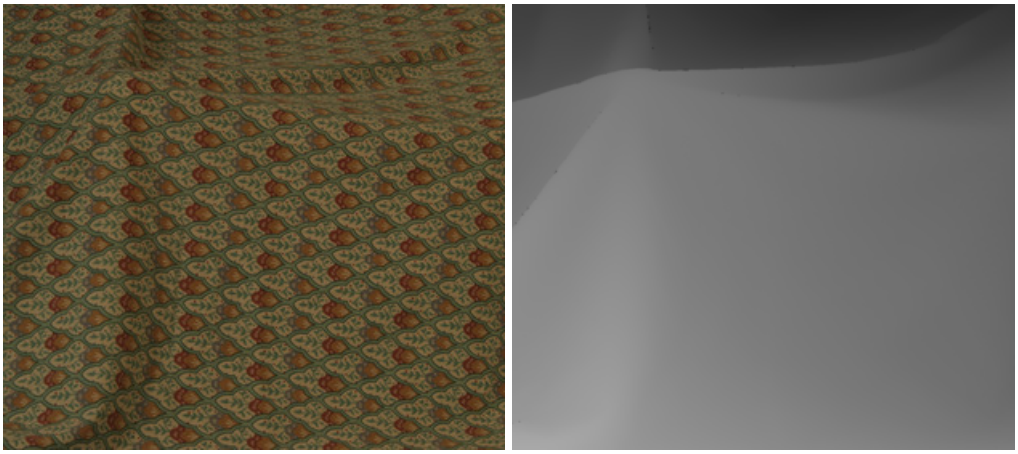


Figure 2.9: (Left) Image of a scene. (Right) Associated height map, intensity proportional to distance from camera. Images from (Scharstein and Pal, 2007; Hirschmüller and Scharstein, 2007).

2.4.1 Height Maps

The shape representation used most frequently throughout this dissertation are height maps. A height map stores 3D shape as a set of height or depth values defined over a plane (See Figure 2.9). Mathematically, a height map can be specified as the set of points $(x, y, z = f(x, y))$ for some function f. Because only one depth value can be specified per (x, y) coordinate, height maps are inherently limited in their representational power, leading some to refer to them as "2.5D" shapes. However, height maps are very natural for computer vision applications, since all points in the world are projected on the the image plane – meaning there is at most a single visible surface point per pixel in an image.

If the height map is continuous and differentiable – that is, if $z_x = \frac{\partial z}{\partial x}$ and $z_y = \frac{\partial z}{\partial x}$ exist for all (x, y) – then we define the *gradient* as (p, q), where $p = z_x$ and $q = z_y$. Moreover, the *surface normal map* can be defined in terms of the gradient,

$$\mathbf{n} = \frac{(-p, -q, 1)^\top}{\sqrt{1 + p^2 + q^2}}. \tag{2.23}$$

In some computer vision algorithms such as photometric stereo, the output is a normal or gradient map. If the surface is differentiable (or *integrable*), one can convert a

normal/gradient map into a height map by "integrating" the normal field. A standard approach is to minimize the following objective function over surface height z,

$$F(z) = \int\int \left[(z_x - p)^2 + (z_y - q)^2\right] dxdy \tag{2.24}$$

where p and q are known and z, z_x, and z_y are unknown. This partial differentiable equation can be solved by applying the Euler-Lagrange equation from variational calculus. The resulting *Poisson equation* yields constraints of the form,

$$z_{xx}^2 + z_{yy}^2 = p_x + q_y. \tag{2.25}$$

For an overview of this approach, see (Horn, 1986, chap. 11). In the presence of noise, outliers, or non-integrable surfaces, other solution techniques may give better results than the Poisson solver (*e.g.*, Agrawal et al. (2006, 2005)).

2.4.2 Meshes

Another popular model for representing shape is a *surface mesh*. A mesh consists of a set of 2D polygons embedded in 3d space. The faces of these polygons define the surface of the object. Typically, triangles are preferred over polygons of higher degree, for the following reasons: (1) they are the simplest possible polygon; (2) arbitrary polygons can be formed from adjacent triangles; and (3) they are always convex, which simplifies analysis of certain computations such as testing whether a ray intersects a given triangle. Because surfaces are typically composed of large sets of adjacent triangles, vertexes are usually shared by multiple triangles. In this scenario, it is most efficient to store vertex positions in a list and specify triangle faces as triplets of indices into the vertex list.

Meshes are often augmented in order to specify additional information about a surface. Examples include face orientations to disambiguate the interior and exterior of a surface; "photometric" normal maps to specify more accurate surface normals than the implicit normals defined by each triangle face; spatially varying texture and / or

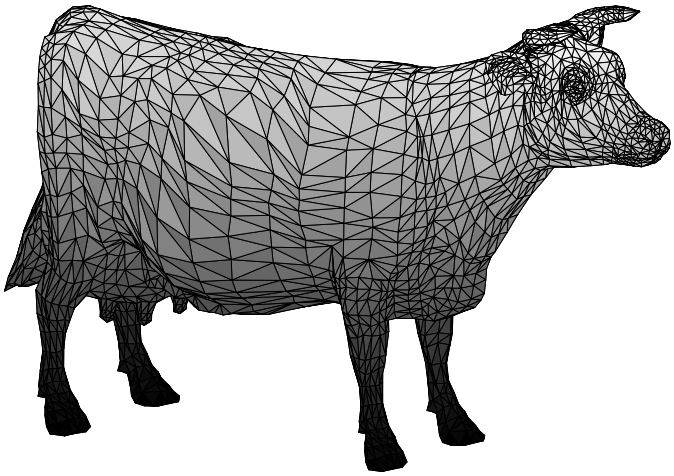


Figure 2.10: Example of a triangle mesh.

reflectance; etc. Naturally, these additional properties can be stored either per-vertex or per triangle face, depending on the needs of a given application.

When working with triangles, it is often useful to utilize barycentric coordinates. In barycentric coordinates, points are defined as linear combinations of the triangle vertex positions. For example, a point with barycentric coordinates (α, β, γ) has Cartesian coordinates $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3$, where $\mathbf{v}_{\{1,2,3\}}$ are the Cartesian coordinates of the three vertexes of some triangle. Barycentric coordinates are restricted so that the coordinate variables sum to unity, (*i.e.*, $\alpha+\beta+\gamma = 1$). Thus, $\gamma = 1-\alpha-\beta$. Moreover, barycentric coordinates have an interesting geometric interpretation: if we label the triangle vertexes A, B, and C and a point lying on the plane of the triangle as P, then the barycentric coordinates of P are proportional to the areas defined by triangles $\overline{BCP}$, $\overline{ACP}$, and $\overline{ABP}$ (see Figure 2.11).

The most common use of barycentric coordinates is to interpolate data along the face of a triangle. In fact, the coordinate variables exactly correspond to weights that linearly interpolate between the vertexes of a triangle. To illustrate, suppose we have texture values t_1, t_2, and t_3 stored at vertexes $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ and we want to linearly interpolate to find the texture value at some point $\mathbf{p}$ along the plane of the triangle. To do this, we first compute the barycentric coordinates of $\mathbf{p}$: (α, β, γ). Then the interpolated

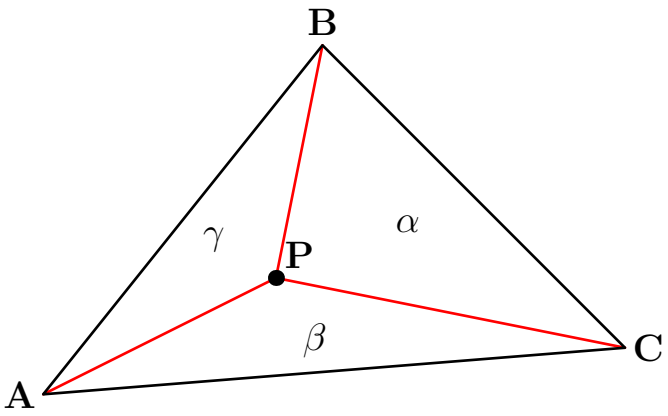


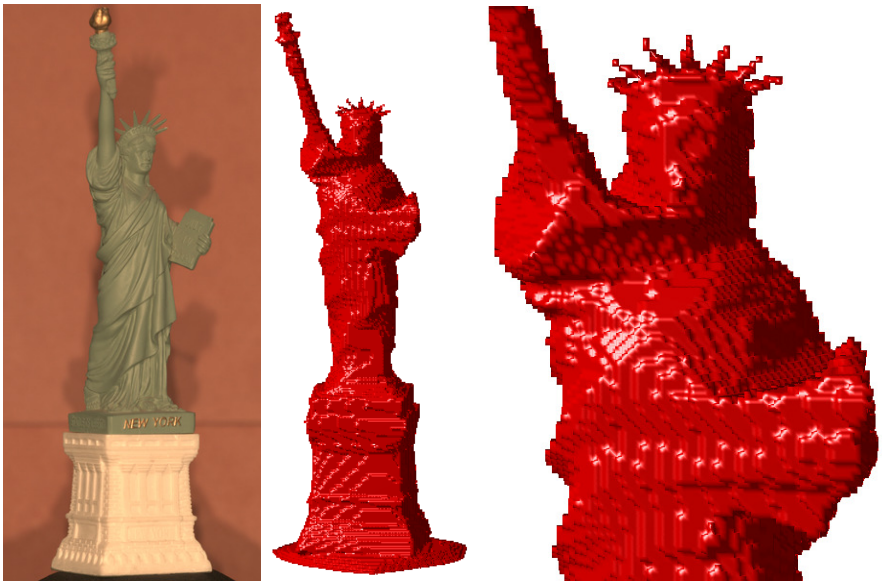
Figure 2.11: A geometric interpretation of barycentric coordinates. Point P has Cartesian coordinates $P = \alpha A + \beta B + \gamma C$, where α, β, and γ are proportional to the areas of triangles $\overline{BCP}$, $\overline{ACP}$, and $\overline{ABP}$ respectively.

texture value is easily computed as $\alpha t_1 + \beta t_2 + \gamma t_3$.

2.4.3 Voxel Grids

Voxel grids are perhaps the simplest representation of shape one could imagine. The basic idea is to discretize 3D space into a set of cells, or *voxels*, with each voxel labeled to indicate whether it is in the interior or exterior of a given shape. Voxel grids are fundamentally different from both height maps and meshes in that they model 3D *volumes* rather than 2D surfaces. For this reason, voxel-grids are particularly well-suited for algorithms that operate on volume data. Examples include *shape-from-silhouette* algorithms such as Laurentini (1994) and *space carving* (Kutulakos and Seitz, 2000). A simple way to utilize silhouette or boundary information is to first distinguish between foreground and background pixels in each image (*i.e.*, pixels that project onto an object or not). Any voxel which projects to a background region in *any* image must lie outside the object. By combining information from multiple viewpoints, one can construct the *visual hull* of the object (Laurentini, 1994), as seen in Figure 2.12.

Voxel grids, like meshes, can be augmented with auxiliary data depending on the application. For example, for fluid simulations in computer graphics one can assign



(a) (b) (c)

Figure 2.12: A voxelgrid obtained by shape-from-silhouettes. (a) 1 of 64 input images. (b) Resulting voxel grid. (c) Detail of voxel grid.

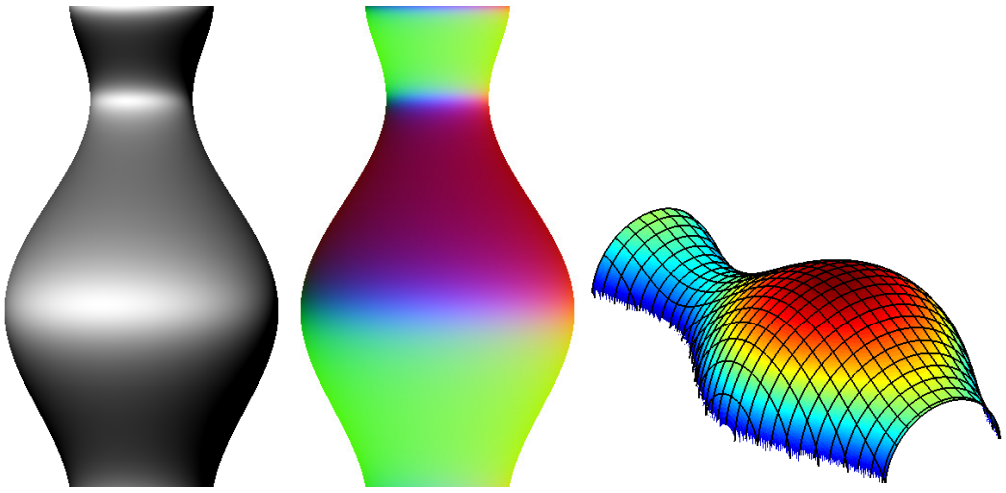
a time-varying fluid velocity at each voxel position; in medical imaging, one can assign values specifying tissue density; etc.

The primary downside to using voxel-grids (or any volumetric representation) is that they are relatively inefficient when compared to surface representations such as meshes. This is because surface representations grow quadratically in size while volume representations grow at a cubic rate. Thus, to achieve equivalent levels of detail in a voxel-grid as in a mesh requires significantly more storage and processing time.

Chapter 3

Photometric Stereo

Photometric stereo refers to a class of algorithms based on observing the appearance of an object under varying illumination. Shape reconstruction is the primary goal, but it is also common to recover surface reflectance and/or illumination as well. Typically, images are acquired from a fixed viewpoint relative to the object and the goal is to recover a surface normal map, which can then be converted into a height map as described in Chapter 2. Closely related to photometric stereo is *shape from shading* (SFS). The primary difference between SFS and photometric stereo is that in SFS only a single input image is given. Thus, SFS is inherently more difficult than photometric stereo because SFS operates on a subset of the images available to photometric stereo algorithms. Given their similarity, ideas from SFS are immediately applicable to photometric stereo (but not necessarily the other way around). In the following sections, we provide an overview and literature review of shape from shading and photometric stereo algorithms. The intent is to provide a backdrop on which to evaluate the research contributions of the subsequent chapters in this dissertation.



(a) (b) (c)

Figure 3.1: Overview of SFS. From an input image, the goal is to obtain the surface orientation at each pixel which can then be "integrated" to form a height map. (a) Input image of a vase with Lambertian reflectance and illumination in direction $\mathbf{s} \propto (-0.3, -0.2, 1)^\top$. (b) Normal map of vase, color encoded as $r = (n_x + 1)/2$, $g = (n_y + 1)/2$, $b = n_z$. (c) Height map of the surface.

3.1 Shape From Shading

The idea behind shape from shading is to utilize *shading* cues to reconstruct the shape of an object from a single image. In this context, shading refers to changes in intensity caused by varying surface orientation across an object. Shape from shading was first introduced by Horn in the early 1970s (Horn, 1970). Since then, SFS has received significant research attention, which is briefly summarized in this section. For a more thorough overview of SFS, see (Zhang et al., 1999) or (Horn, 1986).

Because shading is dependent on surface orientation and not on surface height, it is common for SFS algorithms to output a representation of surface orientations, such as a normal map or gradient map. Just to review, height maps, gradients, and surface

normals are related to each other by,

$$\begin{aligned} \text{Height map} &: z = f(x, y) \\ \text{Surface gradient} &: (p, q) = (z_x, z_y) = (\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}) \\ \text{Surface normal} &: \mathbf{n} = \frac{(-p, -q, 1)^\top}{\sqrt{p^2 + q^2 + 1}}. \end{aligned} \tag{3.1}$$

Assuming the surface is integrable, a normal map or gradient map can be transformed into a height map, up to a global additive height ambiguity (see Figure 3.1).

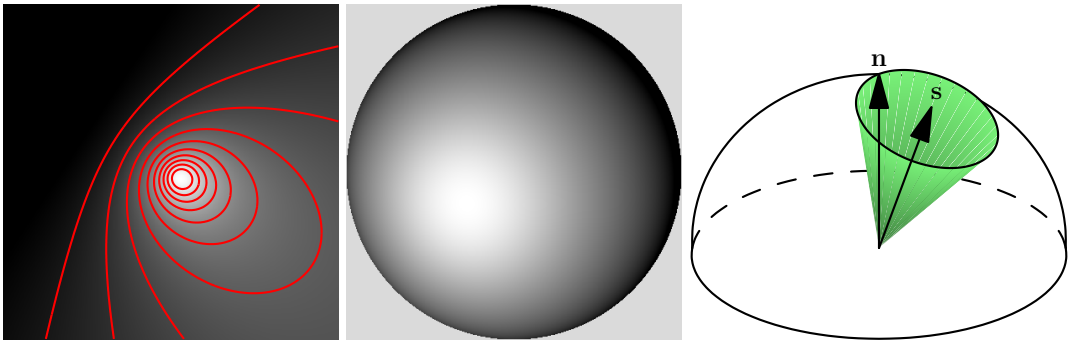
For arbitrary surfaces and / or camera and illumination configurations SFS is an under-constrained problem. It is therefore common to make the following simplifying assumptions:

- Uniform, known surface reflectance
- Illumination from a single directional light source
- Orthographic camera projection
- Negligible global illumination effects (*e.g.*, interreflections, cast shadows, sub-surface scattering).

Under these assumptions, the image formation model is quite simple:

$$e = \rho(\mathbf{n}, \mathbf{s})\mathbf{n}^\top\mathbf{s}, \tag{3.2}$$

where e is the image intensity of a single surface point, $\mathbf{n}$ is the surface normal, $\mathbf{s}$ is the light source direction, and ρ is the BRDF parametrized in terms of $\mathbf{n}$ and $\mathbf{s}$ with implicit viewing direction $\mathbf{v} = (0, 0, 1)^\top$. If $\mathbf{s}$ and ρ are known, then e reduces to a function of the surface orientation; the set of possible image intensities as a function of surface orientation is called a *reflectance map* (see Figure 3.2a). Minimizing the difference between measured surface brightness and the brightness predicted by a reflectance map is the most fundamental goal of SFS algorithms. Usually this is expressed by minimizing



(a) (b) (c)

Figure 3.2: (a) A reflectance map parametrized in terms of the surface gradient (p, q). Each isocontour curve corresponds to a single image intensity value. (b) An image of a sphere is a practical way to measure the reflectance map of otherwise unknown materials. (c) Under Lambertian reflectance, each cone of surface normals centered about the light source direction $\mathbf{s}$ yields identical image intensity.

the following objective,

$$\int\int (E - R)^2 \, dxdy \tag{3.3}$$

where E is the measured brightness value and R represents the value of the reflectance map at each point. Another alternative is to minimize deviations in the gradient of the image intensity,

$$\int\int \left[(E_x - R_x)^2 + (E_x - R_x)^2\right] dxdy. \tag{3.4}$$

An obvious concern in practice is that the reflectance map of a given surface is rarely known a-priori. One way to address this problem is to assume Lambertian reflectance. Under Lambertian reflectance, Equation 3.2 simplifies to,

$$e = \alpha \mathbf{n}^\top \mathbf{s}. \tag{3.5}$$

The surface albedo α can be easily obtained in a number of ways (for example, knowing the orientation of a single point on the surface reveals the albedo). Another approach

is to directly measure the reflectance map using a calibration object. For example, one can image a sphere composed of the same material as the test object under identical illumination (Woodham, 1980; Hertzmann and Seitz, 2005, 2003). Because a sphere contains surface points at every possible orientation, such an image provides a direct measurement of the reflectance map for that material / illumination (see Figure 3.2b).

A serious problem with minimizing the objective in Equation 3.3 is that the solution is not unique. In fact, given a single image, the surface orientation at each point is only constrained to lie on an *isocontour* curve of the reflectance map, a 1D family of solutions in general (see Figure 3.2a). That said, there are a few special conditions where it is possible to uniquely recover the surface: (a) At singular points in the reflectance map, and (b) along the occluding contour of a curved surface. For the special case of Lambertian reflectance, it follows from Equation 3.5 that a single measurement constrains the surface normal to lie in a cone centered about the light source direction (see Figure 3.2c).

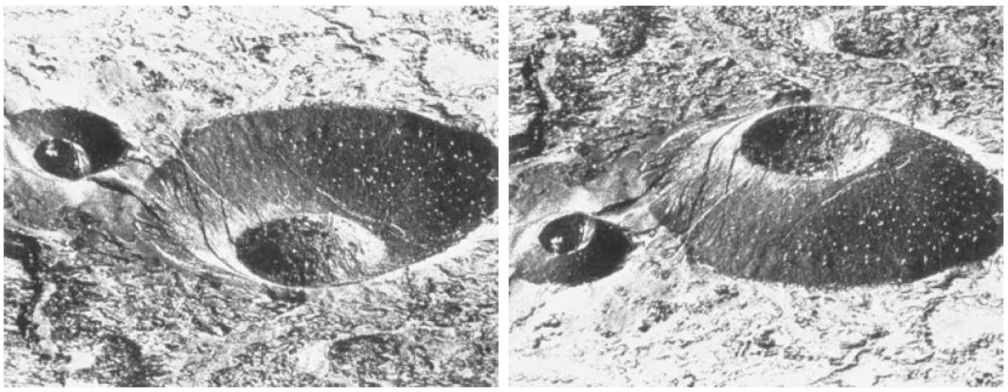
Clearly, knowledge of the reflectance map on its own is insufficient to reconstruct the full surface shape, so it is natural to consider what additional constraints can be utilized to reduce or remove ambiguities. One of the most common approaches is to impose either surface *smoothness* or surface *integrability*. Surface smoothness is an assumption or heuristic that is often mathematically measured in terms of the second order derivatives of the surface height,

$$\int\int \left(p_x^2 + p_y^2 + q_x^2 + q_y^2\right) dxdy. \tag{3.6}$$

Intuitively, this says that smooth surfaces are flat. Surface integrability is a constraint (which holds for smooth surfaces) based on the fact that following the surface gradient along any closed loop of a differentiable surface should not result in a change of depth. Surface integrability is often expressed as,

$$\int\int \left(z_{xy} - z_{yx}\right)^2 dxdy \tag{3.7}$$

where z_{xy} and z_{yx} are the mixed second order partial derivatives of the surface height.



(a) (b)

Figure 3.3: Illustration of the convex / concave ambiguity. The images are identical except for a reflection about the x-axis. Human observers interpret (a) to be a crater and (b) to be a cone due to an implicit assumption in the human visual system that illumination comes from above. Images from (Pentland, 1984).

Even after imposing surface smoothness or integrability, there are still ambiguities that can arise. For example, if the surface is allowed to have arbitrary spatially varying albedo, it becomes impossible to distinguish shading patterns from texture without resorting to ad-hoc techniques. The most impactful ambiguity to classical SFS (where known light source direction and uniform reflectance are assumed) is the *convex / concave* ambiguity under which a pair of surfaces, one convex and one concave, give rise to the same image (see Figure 3.3). In practice, the convex / concave ambiguity can be resolved by prior knowledge (*i.e.*, an operator specifies convexity or concavity to the algorithm). Another ambiguity, which can arise if the light source direction is unknown, is the *generalized bas-relief* (GBR) ambiguity (see Figure 3.4). The GBR ambiguity is most commonly encountered in *uncalibrated photometric stereo* and will be covered in more detail in subsequent sections.

Given a brightness constraint (Equations 3.4 or 3.3) and a surface constraint such as smoothness or integrability (Equations 3.6 or 3.7), the next step is to find a shape that satisfies the constraints. The most common approach is to apply minimization

Figure 3.4: Bas-relief sculptures illustrate a special case of the generalized bas-relief ambiguity. When viewed from a frontal direction, such statues look normal. Viewed from the side, their flattened shape becomes apparent. Images from (Belhumeur et al., 1999).

techniques to solve the problem. For example, combining the brightness constraint with the integrability constraint yields the following cost function,

$$F = \left[\int\int (E - R)^2 \, dxdy\right] + \lambda \left[\int\int (z_{xy} - z_{yx})^2 \, dxdy\right], \tag{3.8}$$

which we want to minimize with respect to the surface gradient $\mathbf{p}$ and $\mathbf{q}$. This system can be solved using variational calculus, provided appropriate boundary conditions. *Propagation* techniques are also commonly used to find SFS solutions. The basic idea is to propagate surface shape starting at points with known geometry. Figure 3.5 shows results from a few two SFS algorithms, one a minimization method and another a propagation method.

While shape from shading is a noble goal, it is beset by problems in practice. First, the requirements imposed by even state-of-the-art methods are quite restrictive. For example, many real-world objects are textured and violate the requirement of uniform surface reflectance. It is also unreasonable to assume known reflectance or access to a calibration object in many circumstances. Unfortunately, it is difficult to relax these requirements for technical reasons. A second problem with SFS is that it is highly sensitive to noise and outliers. This is because only a single measurement is acquired per

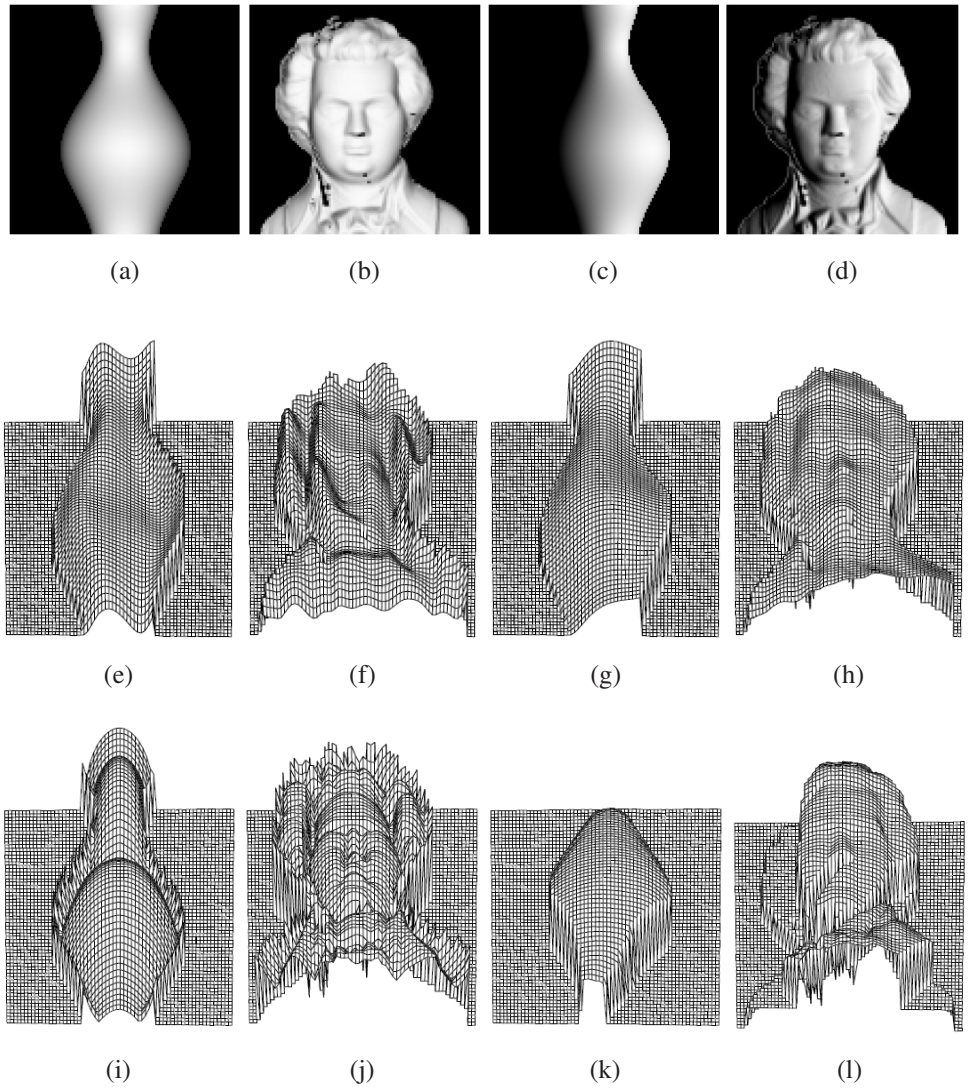
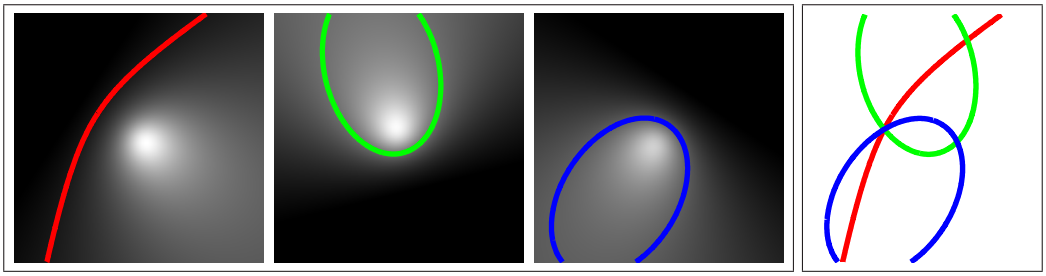


Figure 3.5: (a-d) Synthetic input images. (e-h) Results from (Lee and Kuo, 1993). (i-l) Results from (Bichsel and Pentland, 1992). Images from (Zhang et al., 1999).



(a) (b)

Figure 3.6: (a) Reflectance maps and isocontour curves corresponding to a single surface point under three different illumination conditions. (b) Intersection of all three isocontour curves. Generally, three or more measurements provides a unique solution.

surface point – thus, it is difficult to distinguish noise from the underlying signal.

3.2 Classical Photometric Stereo

Photometric stereo, first introduced by Woodham (1980), is in many ways simply an extension of shape from shading in which multiple images acquired under different illumination conditions are utilized. Early work, such as Woodham (1980) and Silver (1980) directly employ the idea of reflectance maps. It was noted that different illumination conditions resulted in reflectance maps with different sets of isocontours; thus, the surface orientation at a point can be obtained by simply intersecting the isocontour curves obtained from each reflectance map (see Figure 3.6). In general, two measurements constrain the surface orientation of a point to one of two possibilities. Three measurements fully constrains the surface orientation.

It was also observed that under Lambertian reflectance, surface orientations and *spatially varying* albedo could be simultaneously obtained from three or more images. The classical Lambertian photometric stereo algorithm assumes a Lambertian surface illuminated by distant point light sources and viewed by an orthographic camera. Consider acquiring M images, each taken with a different lighting configuration, with N

pixels in each image. Ignoring shadows and interreflections, the image intensity at the ith pixel in the jth image is given by,

$$e_{ij} = \alpha_i \mathbf{n}_i^\top \mathbf{s}_j \tag{3.9}$$

where α_i and $\mathbf{n}_i$ are the albedo and surface normal at the ith pixel position and $\mathbf{s}_j$ encodes the magnitude and direction of the jth directional light source. To simplify analysis, it is convenient to express Equation 3.9 in matrix form. By *vectorizing* the pixels in each image (in, say, column-major ordering), we can compactly express Equation 3.9 as,

$$\mathbf{E} = \mathbf{BS} \tag{3.10}$$

where,

- $\mathbf{E} \in \mathbb{R}^{N \times M}$ stores the pixels from all input images. Each column contains the set of pixels from a single image, and each row corresponds to a different lighting condition for a single pixel.
- $\mathbf{B} \in \mathbb{R}^{N \times 3}$ encodes the surface normal and albedo at each pixel. The ith row corresponds to the product of the albedo with the surface normal vector at the ith pixel; *i.e.*, $\mathbf{B}_{i,:} = \alpha_i \mathbf{n}_i^\top$.
- $\mathbf{S} \in \mathbb{R}^{3 \times M}$ encodes the light source direction and intensity for each image. The jth column contains the jth light source vector $\mathbf{s}_j$.

If $\mathbf{S}$ is known, $\mathbf{B}$ can be obtained by solving the over-constrained linear system in Equation 3.10; *e.g.*,

$$\mathbf{B} = \mathbf{E}\mathbf{S}^\dagger \tag{3.11}$$

where $\mathbf{S}^\dagger$ is the pseudo-inverse of matrix $\mathbf{S}$. Once $\mathbf{B}$ is obtained, the albedo and surface normal at each pixel are given respectively by the magnitude and direction of each row of $\mathbf{B}$. Figure 3.7 shows typical inputs and outputs of the Lambertian photometric stereo algorithm.

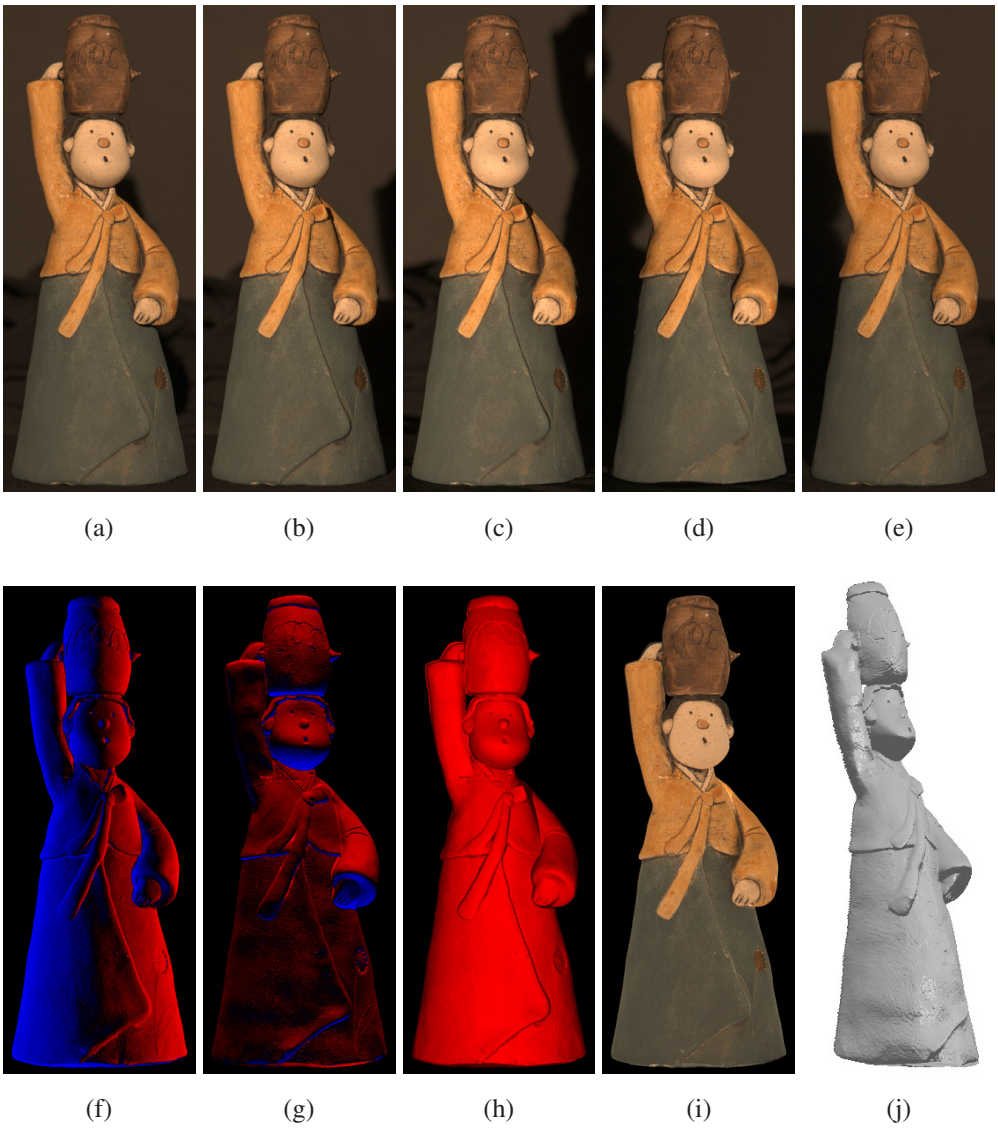


Figure 3.7: Demonstration of the Lambertian photometric stereo algorithm. (a-e) Input images, each obtained with a different light source direction. (f-h) x, y, and z components of the recovered normal map. (i) Recovered albedo map. (j) Height map obtained by integrating the recovered normal map.

There are a few complications that arise in practice when using the simple linear model of Equation 3.10. First, the model does not faithfully represent attached shadows[1]; this can be easily fixed by incorporating a $\max$ function, but this introduces non-linearity making the least-squares solution harder to obtain. A more difficult problem to overcome are *global illumination* effects, which include cast shadows, interreflections, and subsurface scattering. These effects are challenging because any surface point can potentially influence the brightness of any other surface point. Thus, it no longer suffices to treat each surface point independently, they have to all be considered simultaneously. While there has been some effort to incorporate cues from cast shadows and interreflections, *e.g.* Chandraker et al. (2007), the typical approach is to either ignore global illumination effects or to treat them as outliers.

3.2.1 Uncalibrated Photometric Stereo

When the light source direction and strengths, $\mathbf{S}$, are unknown, a case known as *uncalibrated photometric stereo*, the solution is not so simple. Without additional constraints, the best one can do is recover $\mathbf{B}$ and $\mathbf{S}$ up to an unknown general linear transformation. From the form of Equation 3.10 it is clear that the matrix $\mathbf{E}$ is constrained to be at most rank 3, since it is the product of an $N \times 3$ and a $3 \times M$ matrix; this enables $\mathbf{E}$ to be *factorized*. Because of measurement noise, $\mathbf{E}$ may not be exactly rank 3, but the closest rank 3 approximation can be obtained using singular value decomposition (SVD) or similar algorithms. For example, the rank 3 SVD decomposition of $\mathbf{E}$ yields,

$$\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}^\top \tag{3.12}$$

$$\mathbf{E} \simeq \mathbf{U}_3\mathbf{D}_3\mathbf{V}_3^\top \tag{3.13}$$

$$\mathbf{E} \simeq \hat{\mathbf{B}}\hat{\mathbf{S}} \tag{3.14}$$

[1] Attached shadows occur when the surface normal points away from the light source, *i.e.*, $\mathbf{n}^\top\mathbf{s} < 0$.

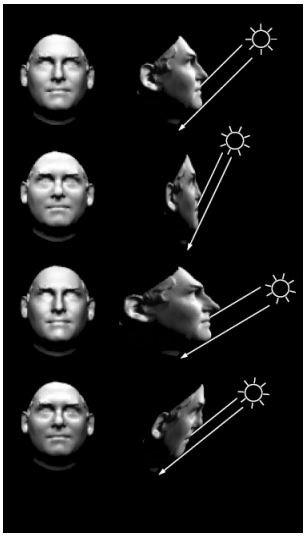


Figure 3.8: Illustration of the generalized bas-relief ambiguity. There is a family of integrable, Lambertian surfaces (right column) that result in identical images (right column). A GBR transforms the shape, albedo, and illumination of a surface. Image from (Belhumeur et al., 1999).

where $\hat{\mathbf{B}} \in \mathbb{R}^{N\times 3}$ and $\hat{\mathbf{S}} \in \mathbb{R}^{3\times M}$. $\hat{\mathbf{B}}$ and $\hat{\mathbf{S}}$ are not uniquely defined, however, since $\hat{\mathbf{B}}\mathbf{A}$ and $\mathbf{A}^{-1}\hat{\mathbf{S}}$ is an equally valid solution for any invertible 3×3 matrix $\mathbf{A} \in GL(3)$.

It has further been shown that only a 3-parameter subset of these transformations, known as the Generalized Bas-Relief (GBR) ambiguity, preserve surface integrability (Belhumeur et al., 1999; Kriegman and Belhumeur, 1998, 2001). Thus, given three or more images of a Lambertian scene acquired under light sources of unknown direction and strength, the surface can be reconstructed up to a GBR transformation by enforcing surface integrability. An analytic method for doing this was introduced by Yuille and Snow (1997).

To understand the GBR ambiguity, consider a Lambertian surface defined by a height map $z = f(x, y)$ and albedo map $\rho(x, y)$ with surface normal $\mathbf{n}$. Then a GBR transformation has the following effect on the albedo and surface normal,

$$\hat{\rho} = \rho\|\mathbf{n}^\top\mathbf{G}^{-1}\| \qquad\qquad \hat{\mathbf{n}}^\top = \frac{\mathbf{n}^\top\mathbf{G}^{-1}}{\|\mathbf{n}^\top\mathbf{G}^{-1}\|} \tag{3.15}$$

where $\hat{\rho}$, $\hat{\mathbf{n}}$, and $\mathbf{G}$ are the transformed albedo, transformed surface normal, and GBR

transformation matrix respectively. Qualitatively, a GBR transforms the depth of each surface point by multiplicatively scaling the depth combined with an additive plane. The albedo and light source vector are simultaneously transformed in such a way that the resulting image from a given viewpoint remains the same (see Figure 3.8). A GBR transformation depends on three parameters μ, ν, and λ and has the following form (Belhumeur et al., 1999),

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix} \qquad \mathbf{G}^{-1} = \frac{1}{\lambda} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{pmatrix}. \tag{3.16}$$

In terms of Equation 3.10, a GBR yields transformed $\mathbf{B}$ and $\mathbf{S}$ matrices,

$$\mathbf{E} = \mathbf{B}\mathbf{G}^{-1}\mathbf{G}\mathbf{S} = \hat{\mathbf{B}}\hat{\mathbf{S}}. \tag{3.17}$$

To resolve the GBR ambiguity, additional constraints must be imposed. Prior works on resolving the GBR ambiguity involve assumptions about (a) the light sources (Yuille and Snow, 1997), (b) the distribution of albedos on the surface (Belhumeur et al., 1999; Hayakawa, 1994), (c) the surface reflectance (Belhumeur et al., 1999; Hayakawa, 1994; Drbohlav and Sara, 2002), and (d) the geometry of the surface (Chandraker et al., 2005; Georghiades et al., 2001). Yuille and Snow (1997) assume knowledge of light source intensities, while several algorithms (Belhumeur et al., 1999; Hayakawa, 1994) assume constant albedo over the entire surface to resolve the parameters of the GBR. Drbohlav and Sara (2002) assume non-Lambertian reflectance, and recover the parameters of the GBR by assuming that the surface normal corresponding to a specular highlight bisects the viewing direction and the light source direction. Georghiades (2003) shows that the GBR can be resolved by assuming that the reflectance of the surface is well described by the Torrance-Sparrow reflectance model. Georghiades et al. (2001) use priors on surface geometry, while Chandraker *et al*. (Chandraker et al., 2005) assume that the surface geometry leads to interreflections which can then be utilized to resolve the GBR. In Chapter 4 a new method for resolving the GBR ambiguity is presented, based on the

fact that most surfaces contain a small number of distinct albedo values, a property not preserved under a GBR transformation.

3.3 Non-Lambertian Photometric Stereo

Much of the emphasis in photometric stereo research has been to relax the assumption of Lambertian reflectance, thus enabling photometric stereo to work on broader classes of objects. For example, it has been observed that the reflectance of many materials is well approximated by the sum of a specular and a diffuse lobe, which has motivated an entire line of research (Coleman and Jain, 1982; Barsky and Petrou, 2003; Ikeuchi, 1981; Nayar et al., 1990). Many of these approaches assume a Lambertian diffuse lobe, while not imposing a parametric form on the specular lobe. Examples include Coleman and Jain (1982) and Barsky and Petrou (2003) who treat specular pixels as outliers as well as Schlüns and Wittig (1993), Sato and Ikeuchi (1994), and Mallick et al. (2005) who assume the color of the specular lobe differs from the color of the diffuse lobe, allowing separation of the specular and diffuse components.

A different approach is to place reference objects in the scene that have similar reflectance to the test object. This method was used in early photometric stereo research (Silver, 1980) and was later reexamined by Hertzmann and Seitz (2003, 2005). The basic idea is that the reference objects provide a direct measurement of the BRDFs in the scene, which is then matched to points on the test object. This works for arbitrary BRDFs, but requires reference objects of the same material as the test object. Spatially varying BRDFs can also be handled by assuming that the BRDF at each point on the test object is a linear combination of the "basis" BRDFs defined by the set of reference objects. This approach to spatially varying BRDFs is similar in spirit to work by Lensch et al. (2001), although their method uses parametric (Lafortune) BRDFs and assumes known surface shape.

Building upon the idea of considering the reflectance at each surface point to be a linear combination of a small set of BRDFs, Goldman et al. (2005) removed the need for

reference objects by iteratively estimating the basis BRDFs and surface normals. Their method assumes an isotropic Ward model for each basis material, whose parameters are estimated at each iteration. While it requires the solution of a difficult optimization problem, their approach is still one of a very few capable of recovering surface normals and relatively flexible parametric BRDFs in tandem.

While parametric models are very good at reducing the complexity of BRDFs, they are usually only valid for a limited class of materials (Ngan et al., 2005; Stark et al., 2005). An alternative is to exploit physical properties common to large classes of BRDFs. For example, it is well known that all real-world BRDFs satisfy energy conservation, non-negativity, and Helmholtz reciprocity. Utilizing these properties, while not as simple as utilizing parametric models, is nonetheless possible. Helmholtz stereopsis, introduced by Zickler et al. (2002a,b), is one such technique, exploiting reciprocity to obtain (multi-view) surface reconstruction with no dependence the BRDF. Isotropy is another physical property which holds for materials without "grain". While isotropy is implicitly assumed in almost all parametric models used in computer vision, only recently has it been explicitly utilized for photometric stereo. Tan et al. (2007) use both symmetry and reciprocity present in isotropic BRDFs to resolve the generalized bas-relief ambiguity. In Chapter 5, it is shown that isotropy, with no further assumptions on surface shape or BRDF, can be utilized to recover the surface normal at each surface point up to a plane. In particular, no parametric model is used, and the BRDF is allowed to vary arbitrarily across the surface.

Another recent development in non-parametric BRDF acquisition is the concept of factorizing sampled BRDF values into the product of a material weight matrix and a BRDF matrix. The most prominent of these approaches is work by Lawrence et al. (2006) who solve the factorization problem using alternating constrained least squares. Their algorithm is again based on the assumption that spatially varying reflectance can be represented as a weighted sum of a small set of materials. Although their technique is primarily focused on BRDF acquisition, they also show limited examples of surface normal estimation. In Chapter 6 a method for simultaneously estimating material weights,

non-parametric basis BRDFs, and surface normals is presented.

Chapter 4

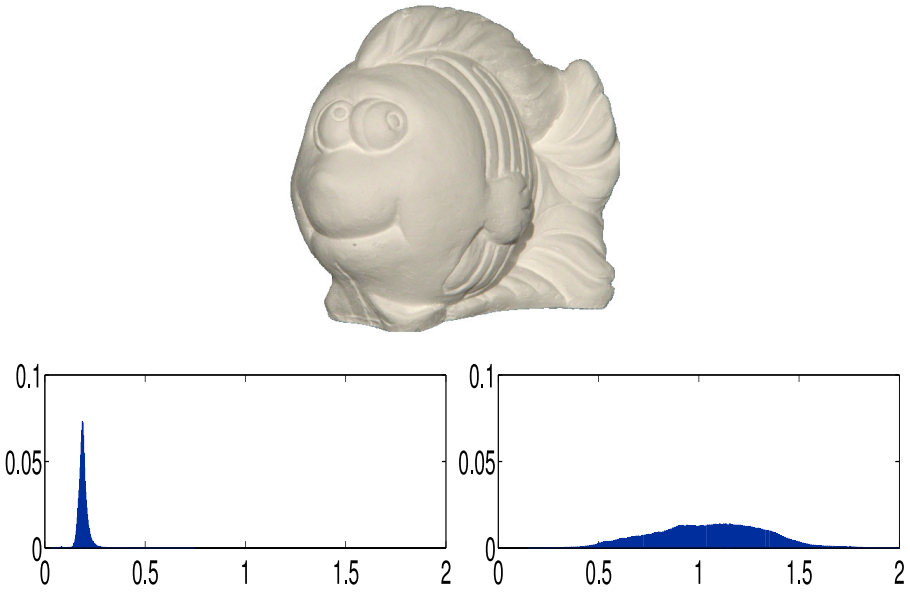
Resolving the Generalized Bas-Relief Ambiguity by Entropy Minimization

In uncalibrated photometric stereo, one is given a set of images taken from the same viewpoint under varying, *unknown* illumination with the goal of recovering the surface shape and albedo, as well as the light source directions and intensities. In this scenario, it has been shown that a family of surfaces give rise to the same image, even if surface integrability is enforced. This ambiguity is called the generalized bas-relief (GBR) ambiguity (see Chapter 3 for more details).

In this chapter, we propose a novel technique for resolving the GBR ambiguity based on minimization of the entropy of the recovered albedos. The hypothesis is based on the observation that in general a GBR transformation smears the distribution of albedos. For example, if the albedo is constant over the entire surface, the distribution of albedo values is a delta function with zero entropy. After a GBR transformation, the observed albedo distribution is a function of the surface geometry and the GBR parameters, and is therefore no longer constant across the surface. This smearing of the albedo values in turn increases the entropy of the albedo distribution (see Figure 4.1).

This chapter makes the following contributions,

- Presents an intuitive conjecture that the GBR can be resolved by minimizing the



(a) Albedo distribution (b) Albedo distribution (GBR)

Figure 4.1: Illustration of the effect of a GBR transformation on the distribution of surface albedos. The true albedo distribution is very peaked while the GBR transformed ($\mu = 3, \nu = 1.5, \lambda = 0.5$) albedo distribution is significantly spread out.

entropy of the GBR transformed albedos.

- Proves the above conjecture under certain assumptions about the distribution of normals and albedos.
- Empirically shows that the above conjecture has excellent performance on synthetic as well as real images.
- Identifies degenerate configurations of the surface normals and the albedos for which the conjecture fails to resolve the GBR.

4.1 Image Formation Model and the GBR Ambiguity

Consider a Lambertian surface defined by a height map $z = f(x, y)$ and albedo map $\rho(x, y)$ with surface normal $\mathbf{n} = \frac{\left(-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1\right)}{\sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1}}$. The intensity of the ith pixel under the jth light source $\mathbf{s}_j$ is given by,

$$e_{ij} = \rho_i \mathbf{n}_i^\top \mathbf{s}_j \tag{4.1}$$

and the set of images formed can be compactly expressed as,

$$\mathbf{E} = \mathbf{BS} \tag{4.2}$$

where $\mathbf{E}$ encodes pixel intensities, $\mathbf{B}$ encodes surface normals and albedo values, and $\mathbf{S}$ encodes light source vectors as described in Chapter 3.

A GBR transformation has the following effect on the albedo and surface normal,

$$\hat{\rho} = \rho \|\mathbf{n}^\top \mathbf{G}^{-1}\| \qquad\qquad \hat{\mathbf{n}}^\top = \frac{\mathbf{n}^\top \mathbf{G}^{-1}}{\|\mathbf{n}^\top \mathbf{G}^{-1}\|} \tag{4.3}$$

where $\hat{\rho}$, $\hat{\mathbf{n}}$, and $\mathbf{G}$ are the transformed albedo, transformed surface normal, and GBR transformation matrix respectively. A GBR transformation depends on three parameters μ, ν, and λ and has the following form (Belhumeur et al., 1999),

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix} \qquad\qquad \mathbf{G}^{-1} = \frac{1}{\lambda} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{pmatrix}. \tag{4.4}$$

In terms of Equation 4.2, a GBR yields transformed $\mathbf{B}$ and $\mathbf{S}$ matrices,

$$\mathbf{E} = \mathbf{B}\mathbf{G}^{-1}\mathbf{G}\mathbf{S} = \hat{\mathbf{B}}\hat{\mathbf{S}}. \tag{4.5}$$

The important thing to note is that under a GBR transformation, the transformed albedo $\hat{\rho}$ is the product of the true albedo and and the norm of vector $\mathbf{n}^\top\mathbf{G}^{-1}$ which varies spatially according to the surface normal at each point unless $\mathbf{G}$ is the identity matrix (*i.e.*, when no GBR is present).

4.2 Entropy Minimization

Differential entropy is a natural measure of the "peakiness" of a probability density function f and is defined as

$$H(f) = -\int_S f(x)\log\left(f(x)\right)dx \tag{4.6}$$

where S is the support of f (*i.e.*, $f(x \notin S) = 0, \forall x$). Entropy minimization has been previously used in vision algorithms to estimate parameters that result in a peaky distribution of some observed quantity. Finlayson et al. (2004) use an intuitive argument to justify the use of entropy minimization to estimate the direction of projection in log-chromaticity space for obtaining an intrinsic image. Similarly, entropy has been used to define a prior in unsupervised clustering when the number of clusters are not known (Palubinskas et al., 1998).

In the following subsection, in addition to providing intuitive arguments in support of minimizing the entropy for resolution of the GBR parameters, we prove that under certain assumptions about the distribution of the albedos and the surface normals, the minimum entropy solution is locally optimal.

4.2.1 An Entropy Based Cost Function

A large percentage of man-made (for example, toys) and natural objects (for example, fruits and vegetables) are composed primarily of a small set of dominant albedo

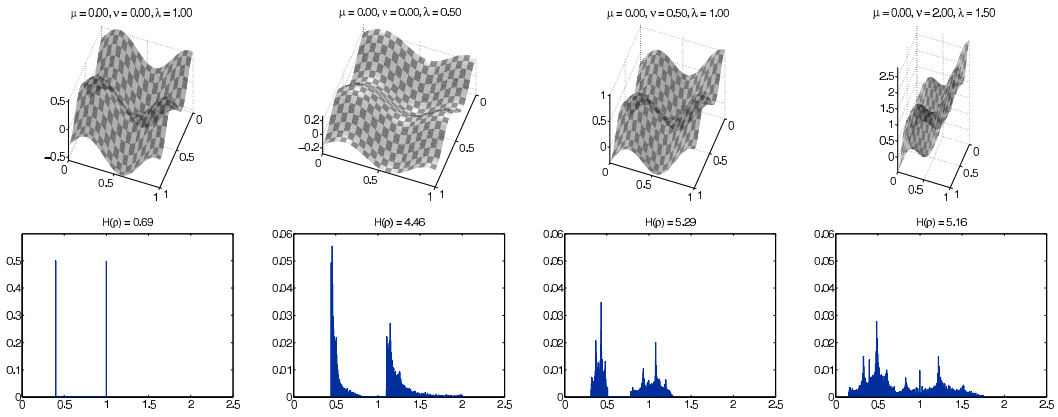


Figure 4.2: The effect of GBR transformations on a surface $f(x, y)$ with two albedos. (Top) A GBR transformed surface, $\tilde{f}(x, y) = (x, y, f(x, y))\mathbf{G}^{\top}$ with albedo $\tilde{\rho} = \|\tilde{\mathbf{n}}^{\top}\mathbf{G}^{-1}\|$. (Bottom) The histogram of GBR transformed albedo values.

values[1]. The probability density function (pdf) of the albedos of such objects will be very close to the sum of a set of delta functions. Even the objects that don't consist of a small set of dominant albedos (for example, human skin) will typically have pdf's that are peaked (*i.e.*, it is very uncommon for an object to have a truly uniform distribution of albedos).

Equation 4.3 suggests that the distribution of GBR transformed albedos depends on the the distribution of the true albedos as well as the distribution of $\gamma = \|\mathbf{n}^{\top}\mathbf{G}^{-1}\|$. In the absence of a GBR transformation (*i.e.* $G = I_{3\times 3}$), the pdf of γ is a delta function centered at one (*i.e.* $\delta(\gamma - 1)$). The GBR has the effect of smearing this delta function, and thereby increasing the entropy of the distribution of the GBR transformed albedos. To motivate this, Figure 4.2 shows the effects of a few GBR transformations on both the surface geometry and the albedo distribution of a synthetic surface. For this surface, GBR transformations clearly smooth out the distribution of albedos.

[1] In fact, in a recent work Ke et al. (2006), a lower hue count in images was used as a measure of good quality.

4.2.2 The k-Albedo Configuration

We now consider a special case when the minimum entropy solution is a local optimum. Consider an object consisting of k different albedo values ρ_i where $i = 1...k$, and $\rho_{i+1} > \rho_i \ \forall i$. Let α_i denote the fraction of all pixels that have albedo ρ_i, and n_i denote the set of normals at these pixels. Note that $\sum_{i=1}^{k} \alpha_i = 1$. We assume that the corresponding sets of surface normals $\mathbf{n}_i$ are identically distributed on a Gauss sphere for all $i = 1...k$. This implies that after GBR transformation, the density of the term $\gamma_i = \|\mathbf{n}_i^\top \mathbf{G}^{-1}\|$ will be identical to the density of $\gamma_j = \|\mathbf{n}_j^\top \mathbf{G}^{-1}\|$ for all $i, j = 1...k$. The density of the true albedos can be written as

$$f_\rho(x) = \sum_{i=1}^{k} \alpha_i \delta(x - \rho_i) \tag{4.7}$$

and the entropy can be written as

$$H(f_\rho) = -\sum_{i=1}^{k} \alpha_i \log \alpha_i \tag{4.8}$$

In the presence of a GBR transformation, the density of $\gamma_i = \|\mathbf{n}_i^\top \mathbf{G}^{-1}\|$ is no longer a delta function. For simplicity of derivation, we assume that f_γ represents the distribution of any γ_i $i = 1...k$ after the GBR transformation (recall that each γ_i is identically distributed). We also assume that f_γ is a small perturbation of the delta function, and has finite support. In other words,

$$f_\gamma(x) = 0, \ \ x \notin [1 - \Delta, 1 + \Delta] \tag{4.9}$$

and,

$$\int_{1-\Delta}^{1+\Delta} f_\gamma(x)\, dx = 1. \tag{4.10}$$

The GBR transformed albedo $\hat{\rho}$ can be treated as a product of two independent random variables ρ and γ. Therefore the probability distribution of $\hat{\rho}$ can be written as (Rohatgi,

1976),

$$\begin{aligned} f_{\hat{\rho}}(y) &= \int_0^\infty f_\rho(x) f_\gamma\left(\frac{y}{x}\right)\frac{1}{x}\,dx \\ &= \int_0^\infty \sum_{i=1}^{k} \alpha_i \delta(x-\rho_i) f_\gamma\left(\frac{y}{x}\right)\frac{1}{x}\,dx \\ &= \sum_{i=1}^{k} \frac{\alpha_i}{\rho_i} f_\gamma\left(\frac{y}{\rho_i}\right) \end{aligned} \tag{4.11}$$

where, f_γ is the same for all values of i owing to identical distribution of n_i and n_j for all $i, j = 1...k$ over the Gauss sphere.

The entropy of the distribution of the GBR transformed albedos can be written as

$$H(f_{\hat{\rho}}) = -\int_0^\infty \sum_{i=1}^{k} \frac{\alpha_i}{\rho_i} f_\gamma\left(\frac{y}{\rho_i}\right) \log\left(\sum_{j=1}^{k} \frac{\alpha_j}{\rho_j} f_\gamma\left(\frac{y}{\rho_j}\right)\right) dy. \tag{4.12}$$

If $\Delta < \frac{\rho_{i+1}-\rho_i}{2(\rho_{i+1}+\rho_i)}$ holds for $i = 1...k-1$, then the distributions of transformed albedos corresponding to two different albedos will not overlap[2]. Under this condition, Equation 4.12 can be re-written as,

$$\begin{aligned} H(f_{\hat{\rho}}) &= -\sum_{i=1}^{k} \frac{\alpha_i}{\rho_i} \int_{(1-\Delta)\rho_i}^{(1+\Delta)\rho_i} f_\gamma\left(\frac{y}{\rho_i}\right) \log \frac{\alpha_i}{\rho_i} f_\gamma\left(\frac{y}{\rho_i}\right) dy \\ &= -\sum_{i=1}^{k} \alpha_i \int_{(1-\Delta)\rho_i}^{(1+\Delta)\rho_i} f_\gamma\left(\frac{y}{\rho_i}\right) \log f_\gamma\left(\frac{y}{\rho_i}\right) d\frac{y}{\rho_i} \\ &\quad - \sum_{i=1}^{k} \alpha_i \log \alpha_i + \sum_{i=1}^{k} \alpha_i \log \rho_i \\ &= H(f_\rho) + H(f_\gamma) + \sum_{i=1}^{k} \alpha_i \log \rho_i. \end{aligned} \tag{4.13}$$

Therefore the entropy of the distribution of the transformed albedos is greater than the entropy of the distribution of the correct albedos when $H(f_\gamma) + \sum_{i=1}^{k} \alpha_i \log \rho_i > 0$.

[2]This follows since $(1-\Delta)\rho_{i+1} > (1+\Delta)\rho_i$.

4.2.3 Degenerate Configurations

In the previous section we showed that minimum entropy solution corresponds to the correct solution under certain assumptions. Empirical evidence suggests that the minimum entropy hypothesis works for a wide variety of configurations. However, we have theoretically identified degenerate cases in which the distribution of surface normals and the surface albedo conspire such that the entropy of the distribution of the GBR transformed albedos is lower than or equal to the entropy of true albedo distribution.

Consider the image of a polyhedron with k visible faces, and let $\mathbf{n}_i$ be the normal associated with the i^{th} face. The transformed albedo of the i^{th} face is

$$\hat{\rho}_i = \rho_i \|\mathbf{n}_i^\top \mathbf{G}^{-1}\| \tag{4.14}$$

where, ρ_i is the albedo associated with the i^{th} face, and $\mathbf{G}$ is the GBR transform. Given a GBR $\mathbf{G}$ and a collection of normals n_i $i = 1...k$, the i^{th} face of the polyhedron can be painted with albedo $\rho_i = 1/\|\mathbf{n}_i^\top \mathbf{G}^{-1}\|$ so that all GBR transformed albedos have value equal to unity, and their distribution is a delta function and the entropy is minimum. This is clearly a degenerate configuration.

Consider a second case in which the surface is a plane, and it contains k distinct albedos. It is easy to see that under this condition, the distribution of $\hat{\rho}$ is a sum of k delta functions, regardless of what GBR transform is chosen. Therefore, this configuration is also degenerate, and entropy does not provide information about the GBR.

4.3 Experimental Validation

To empirically validate that the minimum entropy solution indeed resolves the GBR, we formulate an optimization algorithm that solves for the parameters of the GBR (μ, ν, and λ) given an uncalibrated set of surface normals and albedos. Our experiments consist of the following high-level steps,

- Photograph or render multiple images of a Lambertian object under different lighting conditions.

- Recover the surface and light sources $\mathbf{I} = \hat{\mathbf{B}}\hat{\mathbf{S}}$ up to a GBR transformation using the algorithm of Yuille and Snow Yuille and Snow (1997).

- Find the GBR parameters that minimize the entropy of the albedo distribution and apply the GBR to recover the true surface $\rho_i \mathbf{n}_i^\top = \mathbf{B}_i = \hat{\mathbf{B}}_i \mathbf{G}$.

- Validate the solution by comparing to calibrated photometric stereo results.

The following subsections describe our methodology in more detail.

4.3.1 Approximating the Differential Entropy

In Section 4.2 we assumed knowledge of the underlying probability density function of $\hat{\rho}$. In practice, we have measurements of this quantity at each pixel location and seek to approximate the underlying distribution using these samples. While there are many powerful methods for approximating continuous distributions from samples (e.g., Parzen windows, the mean-shift algorithm, kernel based estimators, etc.), for the purpose of computing entropy we find it sufficient to use a histogram approximation.

Histograms are beneficial because (1) they are simple to generate from sampled data and (2) the entropy of a histogram can be computed as a discrete summation instead of an integral. These properties are particularly important in our case since we need to evaluate the entropy of the albedo distribution under many different GBR transformations. Consider an m-bin histogram $\{a_i\}$, $i = 1...m$ generated from n i.i.d. samples with underlying distribution f. Then the simplest estimator of the entropy $H(f)$ is,

$$\hat{H}_{MLE}(f) = -\sum_{i=1}^{m} \frac{a_i}{n} \log \frac{a_i}{n} \tag{4.15}$$

which is the maximum likelihood estimator of $H(f)$ given its histogram. While $\hat{H}_{MLE}$ can exhibit significant bias[3], its variance is typically low and thus should be well suited in a minimization framework.

[3] See Paninski (2003) for a bias corrected estimator.

A downside to using histograms is that the number of bins and the support region must be chosen appropriately. For the purpose of comparing albedo distributions we fix the number of bins to some constant (256 bins were used for most of the experimental results shown in Section 4.3.4). Moreover, since we don't want to overly favor distributions with narrow support, we set the support to the range of the albedo samples.

4.3.2 Optimization

We now turn to the problem of finding the GBR parameters μ, ν, and λ that minimize the entropy of the albedo distribution. Let $\{\hat{\mathbf{b}}_i = \hat{\rho}_i \hat{\mathbf{n}}_i\}$, $i = 1...N$ be a set of GBR transformed surface normals scaled by the GBR transformed albedo. Then, under a *proposed* set of GBR parameters encoded in matrix $\tilde{\mathbf{G}}$, the obtained set of albedos is $\{\tilde{\rho}_i = \|\hat{\mathbf{b}}_i \tilde{\mathbf{G}}\|\}$ from which the entropy $H(f_{\tilde{\rho}}) \simeq \hat{H}_{MLE}(hist(\tilde{\rho}))$ can be estimated as described in Section 4.3.1. The first thing to note is that the entropy estimate is neither convex, nor differentiable with respect to the GBR parameters. Given these facts, it seems finding the minima will be very difficult. Luckily, the number of parameters is low and the error surface is empirically smooth. Moreover, by making weak assumptions on the surface geometry or lighting configuration we can bound the parameters. Empirically, we find that the GBR parameters induced by the uncalibrated photometric stereo algorithm very rarely exceed an absolute value of 5 – even with varied light source intensity. Also, because of the concave/convex ambiguity, we can restrict λ to be positive. Based on these observations, we restrict our search space throughout our experiments to,

$$\begin{aligned} -5 \leq \mu \leq 5 \\ -5 \leq \nu \leq 5 \\ 0 \leq \lambda \leq 5. \end{aligned} \tag{4.16}$$

Given these bounds on the parameters, along with our observation of smoothness in the error surface, we solve for the GBR parameters using discrete search with

coarse-to-fine refinement. The algorithm works by iteratively sampling the error function at uniform positions within the parameter space. At each iteration, the bounds are tightened and centered around the best solution found in the previous iteration. This process continues until the sampling interval falls below some threshold. While other strategies such as simulated annealing could also have been used, we found that our method worked well enough to find an optimum in most cases[4].

4.3.3 Image Acquisition and Pre-Processing

Prior to using any input images, we perform a number of pre-processing steps. First of all, whenever possible we remove ambient lighting by subtracting each image from an ambient image. Next, we generate a mask image to isolate the object from the background. This is either done manually or automatically using thresholding. Once this is done, we randomly scale each image by a scalar between 0.5 and 1.5. This has the effect of randomizing the light source magnitude, which is important because many datasets are collected using uniform magnitude which can resolve the GBR. By randomizing the intensities we can be certain that our algorithm is not taking advantage of uniform light source magnitude.

After this step, we compute a visibility matrix, $\mathbf{V}$, which is used to reduce the effect of outlier pixels in the input images. $\mathbf{V}$ is the same size as the image matrix $\mathbf{I}$ and takes value 1 if the corresponding pixel in $\mathbf{I}$ is an inlier or value 0 if the corresponding pixel is an outlier. We compute the visibility matrix by excluding pixels that don't lie close to the linear subspace predicted by the Lambertian model (this can be done using SVD).

Next we perform uncalibrated photometric stereo on the set of input images, discounting pixels masked by V[5], which yields a set of GBR transformed surface normals and albedos. Prior to searching for the GBR parameters as described in Section 4.3.2,

[4]Depending on the "coarseness" of the search grid, as well as the number of samples, the optimization takes anywhere from a few minutes to a few hours.

[5]To handle missing values, we utilize the SVD algorithm derived by Brand (2002).

we aggressively filter any surface points that could be outliers. This includes edge points (in any image) as well as points that do not lie close to the predicted Lambertian subspace. Finally, we search over the GBR parameter space to find GBR that minimizes the entropy of the albedo distribution.

4.3.4 Results

We show results on three datasets. Figure 4.4 shows the reconstruction of a synthetically generated Stanford bunny. There are a total of 6 input images in this dataset, each rendered in POV-Ray (2008) with a different light source position. As is clearly seen in the figure, the output of the uncalibrated photometric stereo algorithm results in an extremely flat distribution of albedos. More importantly, our optimization routine clearly finds the correct GBR parameters.

Our second dataset consists of 15 images from the YaleB face database (Georghiades et al., 2001). Looking at figure 4.5, two things are noteworthy : first, our minimum entropy solution appears qualitatively better than the reconstruction obtained from calibrated photometric stereo. For example, the calibrated photometric stereo result is clearly skewed in the x direction, while the minimum entropy solution is not. We conjecture that slight errors in the the light source vectors are biasing the reconstructed surface in this case. There are also many shadow and non-Lambertian effects in these images that may be affecting the calibrated case. These results empirically show that the albedo entropy can resolve the GBR for images consisting of k albedos as well as more general albedo distributions such as human faces.

Finally, Figure 4.6 shows the output of our algorithm on a real surface with two albedos. Our solution is very close to that obtained by calibrated photometric stereo where the light source matrix has been provided.

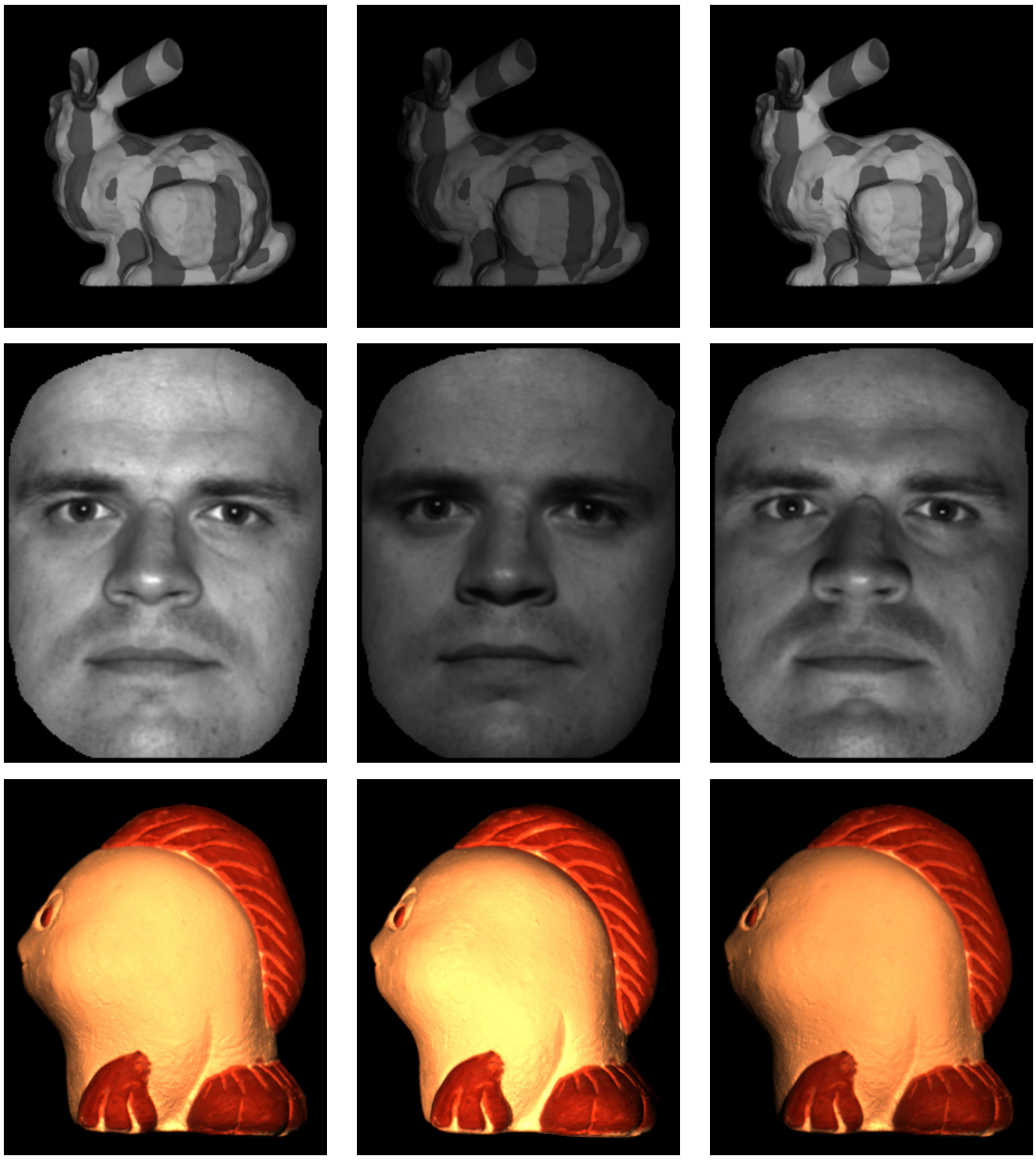


Figure 4.3: Input images. (Top row) Three of the 6 synthetic input images used to generate the results in Figure 4.4. (Middle row) Three of the 15 images (obtained from the Yale face database) used to generate the results in Figure 4.5. (Bottom row) Three of the 5 input images used to generate the results in Figure 4.6.

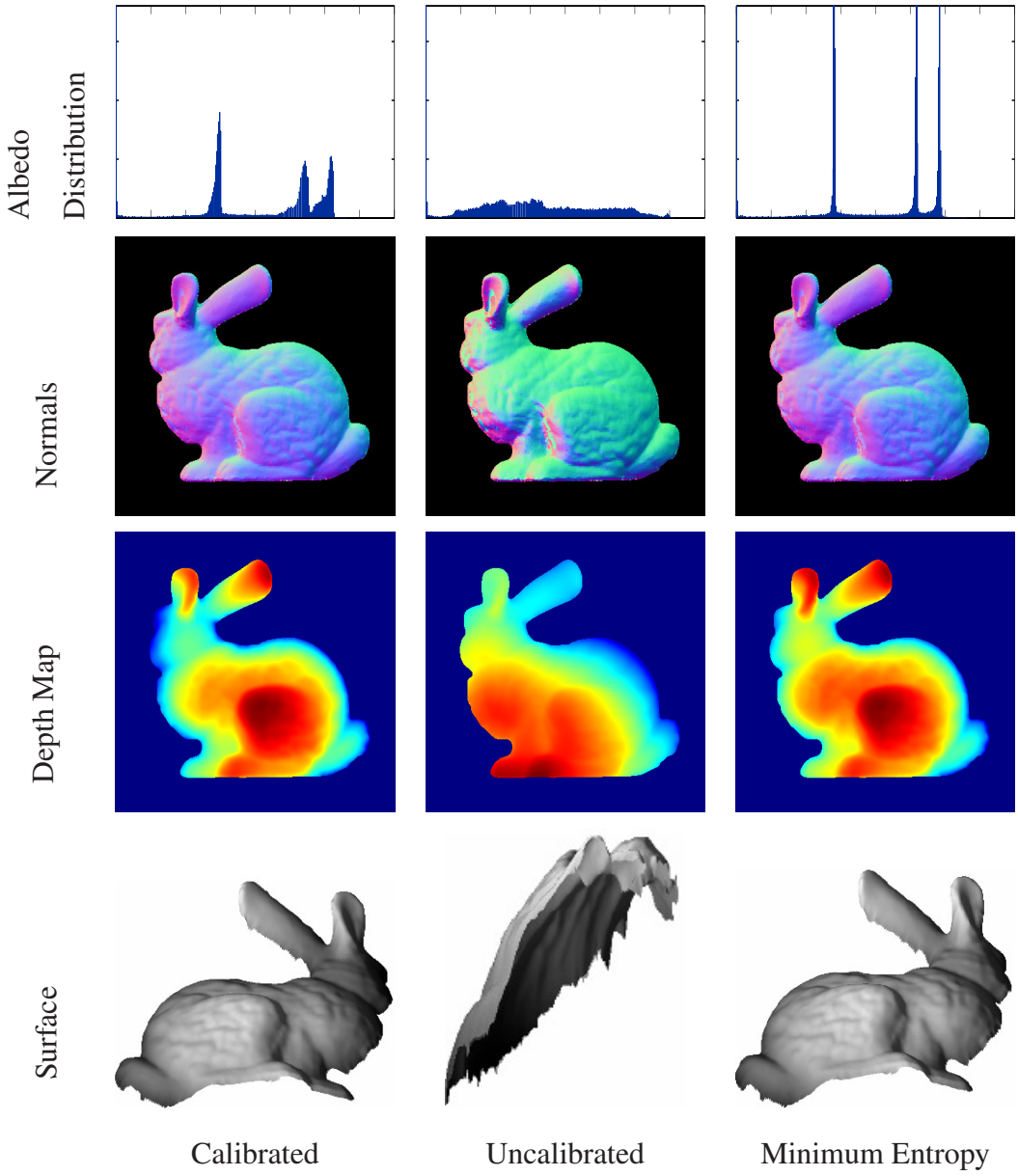


Figure 4.4: Results on synthetic images. (Column 1) Calibrated photometric stereo. (Column 2) Uncalibrated photometric stereo. (Column 3) Minimum entropy solution. (Row 1) Distribution of albedos. (Row 2) Color encoded surfaces normals. (Row 3) Color encoded depth map. (Row 4) Reconstructed surfaces from the same viewpoint.

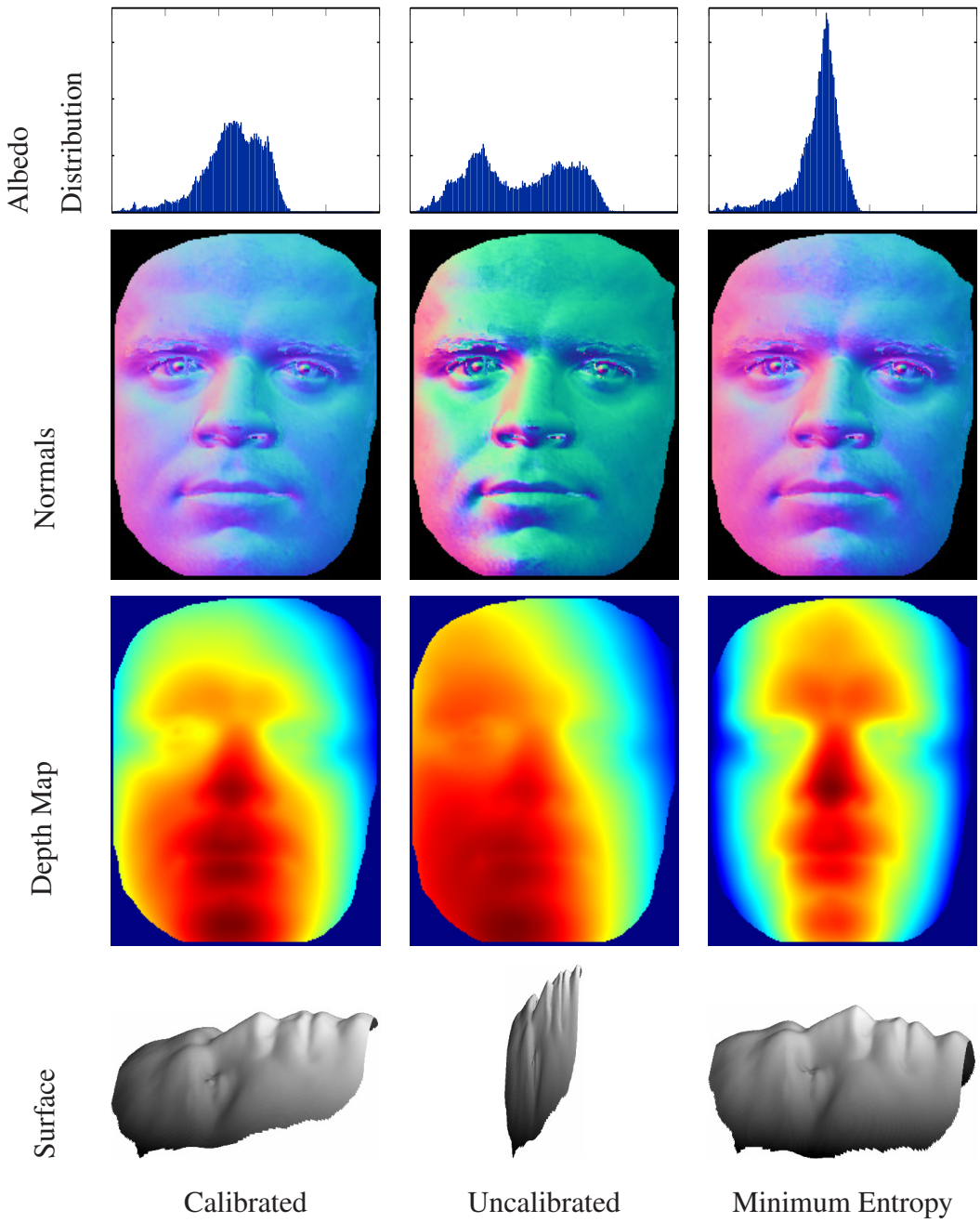


Figure 4.5: Results on a human face (from the Yale face database).

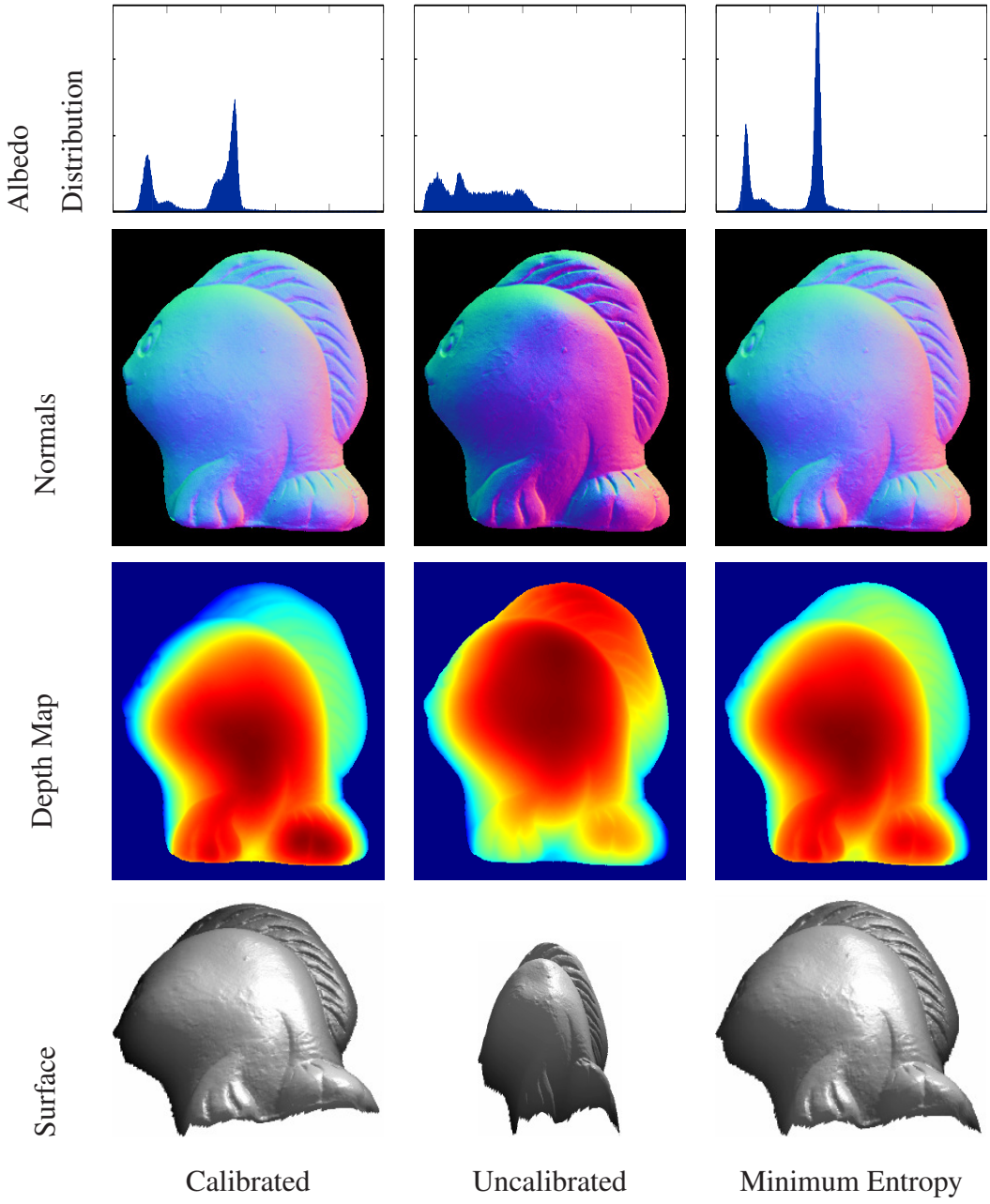


Figure 4.6: Results on (non-synthetic) images of a fish.

4.4 Summary

This chapter presents an algorithm to resolve the Generalized Bas-Relief ambiguity based on a novel prior on the observed albedos. In addition to providing an intuitive reasoning in favor of entropy minimization for estimation of GBR parameters, we show that entropy minimization provides provably correct results under certain assumptions about the surface normal and albedo distribution. Finally, we validate our results on real and synthetic data.

Chapter 5

Reconstructing Surfaces With Arbitrary Isotropic Reflectance

Assuming multiple images with known directional illumination and fixed viewpoint, one might wonder "What is the minimal set of constraints required to recover the shape of a surface?" To answer this question requires systematically defining the set of potential constraints and determining how each one restricts the surface shape. Figure 5.1 shows a hierarchy of possible constraints that could be employed in a photometric stereo setup to reconstruct a surface. In this chapter we examine the constraints of isotropic BRDF and surface smoothness / differentiability and show that just using isotropy, for every image point the surface normal can be constrained to a plane. By also imposing surface smoothness, the isocontour structure (e.g., curves of constant surface height) can be determined. While not a full Euclidean representation, the isocontour structure provides topological information about the surface (such as critical points and the Reeb graph) and could be sufficient for many applications including object recognition and parts inspection (for example, Samir et al. (2006) uses iso-depth contours of human faces for face recognition). Moreover, by imposing any of the additional constraints listed in Figure 5.1 it is possible to recover the true surface barring degenerate cases. For example, at attached shadow boundaries the surface normal is constrained

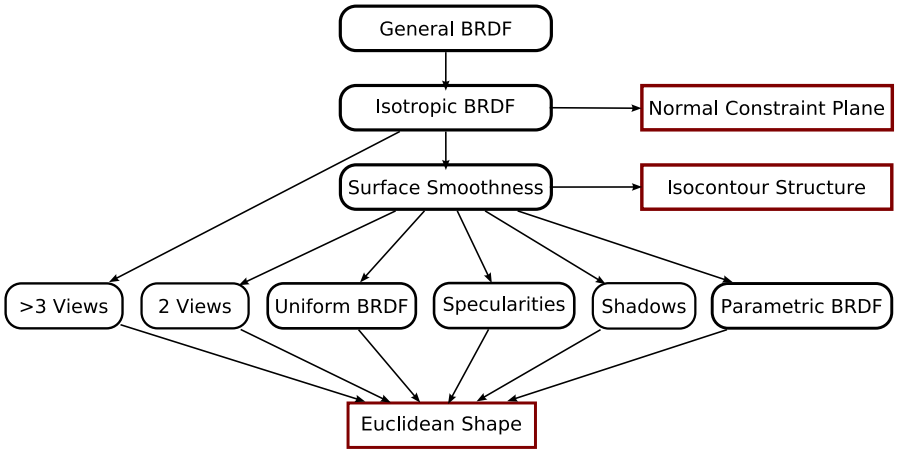


Figure 5.1: A hierarchy of assumptions whose corresponding constraints can be used to recover the surface of an object in a photometric stereo setup.

to lie in the plane orthogonal to the light source direction. Combined with isotropy this constrains the surface normal to lie in the intersection of two planes – a unique solution in general.

One of the main contributions of this chapter is the use of a relatively unexplored physical property that holds for any isotropic BRDF – that isotropic BRDFs are symmetric about the plane spanned by the viewing direction and surface normal. From image data we show how to estimate this plane at each point and thus restrict the set of surface normals to lie in a plane. While we apply this constraint to photometric stereo, it could potentially be useful in other computer vision contexts as well.

5.1 The Bilateral Symmetry Constraint

The crux of our algorithm is based on a known, but relatively unexplored property of isotropic BRDFs that has been previously referred to as *bilateral symmetry*[1] (Marschner, 1998; Drbohlav, 2003). Consider a surface patch with normal $\mathbf{n}$ viewed

[1] Some authors consider isotropy and bilateral symmetry to be distinct phenomenon (i.e., isotropy $\not\Rightarrow$ bilateral symmetry); we do not make such a distinction since all or nearly all physically valid isotropic BRDFs have this property.

from direction $\mathbf{v}$. Bilateral symmetry simply means that the BRDF is symmetric about the plane spanned by $\mathbf{n}$ and $\mathbf{v}$ with respect to the incident lighting direction $\mathbf{s}$ (see Figure 5.2). Isotropic BRDFs are often described by the fact that the exitant radiance emitted from an isotropic surface patch is constant when the surface is rotated about its normal. Bilaterally symmetric BRDFs can similarly be described by the fact that the exitant radiance emitted from a bilaterally symmetric surface patch is constant when the surface is reflected about any plane containing its normal.

Looking again at Figure 5.2, consider some incident light source direction $\mathbf{s}$. Then there exists another direction $\mathbf{s}'$ – obtained by reflecting $\mathbf{s}$ about the plane spanned by $\mathbf{n}$ and $\mathbf{v}$ – that gives rise to the same reflectance. Following the notation of Tan et al. (2007), we call such a pair of points an *isotropic pair*, which we define as,

Definition 1. *Two light source directions* $\mathbf{s}$ *and* $\mathbf{s}'$ *form an* isotropic pair *if they satisfy* $\mathbf{n}^\top\mathbf{s} = \mathbf{n}^\top\mathbf{s}'$ *and* $\mathbf{v}^\top\mathbf{s} = \mathbf{v}^\top\mathbf{s}'$ *where* $\mathbf{n}$ *is the normal of a surface patch and* $\mathbf{v}$ *is the viewing direction.*

We summarize the main consequence of isotropic pairs in the following fact,

Fact 1. *For any isotropic pair of light sources defined relative to the surface normal and viewing direction, the value of an isotropic BRDF is identical.*

Fact 1 follows directly from the bilateral symmetry present in isotropic BRDFs.

5.1.1 Image Formation Model

As is typically done in photometric stereo, we assume distant point light sources and an orthographic camera. Since the BRDF ρ is isotropic, it can be parametrized as a function of $\alpha = \mathbf{n}^\top\mathbf{s}$, $\beta = \mathbf{n}^\top\mathbf{v}$, and $\gamma = \mathbf{v}^\top\mathbf{s}$. Ignoring cast shadows and interreflections we arrive at the following image formation model,

$$E = L\rho(\mathbf{n}^\top\mathbf{s}, \mathbf{n}^\top\mathbf{v}, \mathbf{v}^\top\mathbf{s}) \max\{0, \mathbf{n}^\top\mathbf{s}\} \tag{5.1}$$

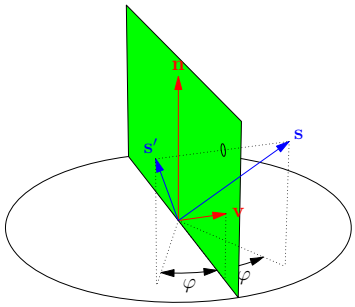


Figure 5.2: Illustration of the bilateral symmetry constraint. Symmetric light source vectors are obtained by reflecting $\mathbf{s}$ about the plane spanned by the surface normal $\mathbf{n}$ and viewing direction $\mathbf{v}$.

where E is the radiance arriving at the camera from a given scene point, L is the radiant intensity of the light source, and $\mathbf{n}$, $\mathbf{v}$, and $\mathbf{s}$ refer to the surface normal, viewing direction and light source direction respectively.

Note that this parameterization makes Fact 1 very explicit : For fixed $\mathbf{n}$ and $\mathbf{v}$ an isotropic pair of light source vectors map to the same BRDF parameters since $\alpha = \mathbf{n}^\top\mathbf{s} = \mathbf{n}^\top\mathbf{s}'$, $\beta = \mathbf{n}^\top\mathbf{v}$, and $\gamma = \mathbf{v}^\top\mathbf{s} = \mathbf{v}^\top\mathbf{s}'$. Moreover, Equation 5.1 also makes apparent the following fact,

Fact 2. *For a given surface normal* $\mathbf{n}$ *and viewing direction* $\mathbf{v}$*, the emitted radiance* $E(\mathbf{s})$ *from an isotropic material is symmetric about the span of* $\mathbf{n}$ *and* $\mathbf{v}$ *(barring non-local illumination effects).*

Consider an isotropic pair of light sources $\mathbf{s}$ and $\mathbf{s}'$. From Equation 5.1 it is clear that $E(\mathbf{s}) = E(\mathbf{s}')$ since $\mathbf{n}^\top\mathbf{s} = \mathbf{n}^\top\mathbf{s}'$ and $\mathbf{v}^\top\mathbf{s} = \mathbf{v}^\top\mathbf{s}'$. Since an isotropic pair can be formed from any light source by reflecting it about the plane spanned by $\mathbf{n}$ and $\mathbf{v}$ it follows that the emitted radiance $E(\mathbf{s})$ is symmetric about the plane spanned by $\mathbf{n}$ and $\mathbf{v}$.

Finally, we note that for all but a few degenerate BRDFs the span of $\mathbf{n}$ and $\mathbf{v}$ is the only plane about which $E(\mathbf{s})$ is symmetric,

Fact 3. *For a given surface normal* $\mathbf{n}$ *and viewing direction* $\mathbf{v}$*, consider the emitted radiance function* $E(\mathbf{s})$ *resulting from isotropic BRDF* ρ*. If we consider the set of planes passing through* $\mathbf{v}$*, then* $E(\mathbf{s})$ *is only symmetric about the plane containing* $\mathbf{n}$*, unless the BRDF is of the form* $\rho = \frac{f(\mathbf{v}^\top \mathbf{s})}{\mathbf{n}^\top \mathbf{s}}$ *for* $\mathbf{n}^\top \mathbf{s} > 0$*.*

Proof.

1. Suppose $E(\mathbf{s})$ is symmetric about some plane defined by unit normal π and that this plane passes through the viewing direction $\mathbf{v}$, but not through the surface normal $\mathbf{n}$ so that $\mathbf{v}^\top \pi = 0$ and $\mathbf{n}^\top \pi \neq 0$.

2. For $E(s)$ to be symmetric about this plane, $E(\mathbf{s}) = E(\mathbf{s}')$ must hold for all $\mathbf{s}$, where $\mathbf{s}' = \mathbf{s} - 2(\mathbf{s}^\top \pi)\pi$ is the reflection of $\mathbf{s}$ about plane π.

3. For fixed $\mathbf{n}$ and $\mathbf{v}$ any isotropic BRDF can be written in the form,

$$\rho(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}) = \begin{cases} \frac{f(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s})}{\mathbf{n}^\top \mathbf{s}} & \text{if } \mathbf{n}^\top \mathbf{s} > 0 \\ 0 & \text{if } \mathbf{n}^\top \mathbf{s} \leq 0 \end{cases} \tag{5.2}$$

4. Re-expressing Equation 5.1 we get,

$$E(\mathbf{s}) = L\rho(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}) \max\{0, \mathbf{n}^\top \mathbf{s}\} \tag{5.3}$$

$$= Lf(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}). \tag{5.4}$$

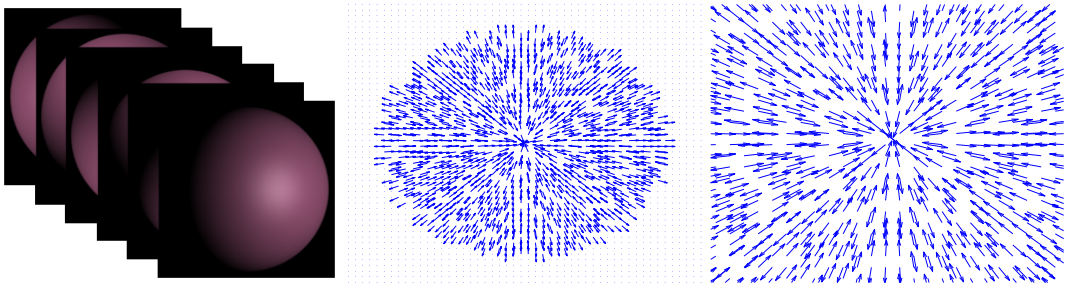
5. From Step 2, $E(\mathbf{s}) = E(\mathbf{s} - 2(\mathbf{s}^\top \pi)\pi)\ \forall \mathbf{s}$. This implies that for all $\mathbf{s}$,

$$f(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}) = f(\mathbf{n}^\top (\mathbf{s} - 2(\mathbf{s}^\top \pi)\pi), \mathbf{v}^\top (\mathbf{s} - 2(\mathbf{s}^\top \pi)\pi)) \tag{5.5}$$

$$= f((\mathbf{n} - 2(\mathbf{n}^\top \pi)\pi)^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}). \text{ [Using fact that } \mathbf{v}^\top \pi = 0] \tag{5.6}$$

6. Since $\mathbf{n}^\top \pi \neq 0$, this holds only if f does not depend on $\mathbf{n}^\top \mathbf{s}$, *i.e.*, $f(\mathbf{n}^\top \mathbf{s}, \mathbf{v}^\top \mathbf{s}) = f(\mathbf{v}^\top \mathbf{s})$. Plugging into Equation 5.2 we see that the BRDF must have the form $\rho = \frac{f(\mathbf{v}^\top \mathbf{s})}{\mathbf{n}^\top \mathbf{s}}$.

□



(a) (b) (c)

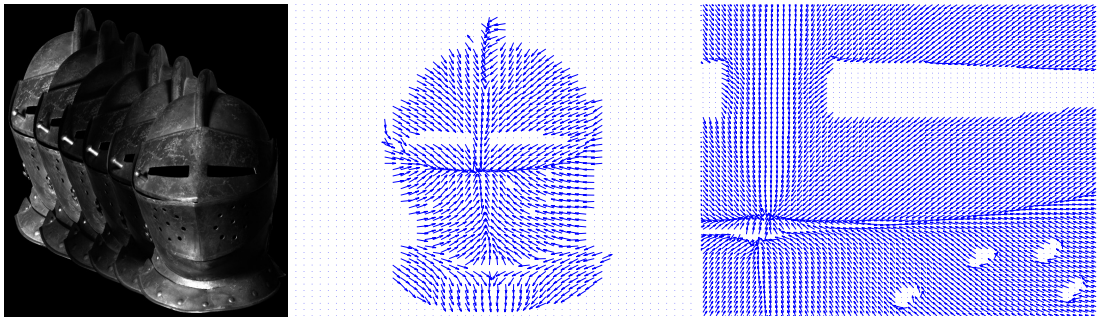
Figure 5.3: Surface gradient directions recovered from 36 images of a synthetic sphere. (a) Input images. (b) Quiver plot of gradient directions. (c) Zoom-in of quiver plot at center of sphere.

As an aside, it is worth discussing the special case when $\mathbf{n}$ and $\mathbf{v}$ are coincident. When this occurs, the span of $\mathbf{n}$ and $\mathbf{v}$ is degenerate; however, our theory still holds if symmetry is defined appropriately. Specifically, we say two points $\mathbf{s}$ and $\mathbf{s}'$ are symmetric about the span of $\mathbf{n}$ and $\mathbf{v}$ if $\mathbf{n}^\top \mathbf{s} = \mathbf{n}^\top \mathbf{s}'$ and $\mathbf{v}^\top \mathbf{s} = \mathbf{v}^\top \mathbf{s}'$.

5.1.2 A Minimal Lighting Configuration for Detecting Symmetry

Based on the theory in Section 5.1.1 we could recover the symmetry plane spanned by $\mathbf{n}$ and $\mathbf{v}$ at each point on the surface by detecting symmetry in the emitted radiance function $E(\mathbf{s})$ as measured over the entire sphere of lighting directions. However, this is quite redundant considering we already know the viewing direction (e.g., $\mathbf{v} = (0, 0, 1)^\top$ in a camera centered coordinate system). In fact, since symmetry planes are coincident with $\mathbf{v}$, it suffices to use a 1D slice of the reflectance field with constant $\mathbf{v}^\top \mathbf{s}$.

In terms of acquisition setup, this implies a circle of light source positions parallel to the image plane and centered about the optical axis. Parametrizing $\mathbf{s}$ in spherical coordinates (θ, ϕ) with pole $\mathbf{v}$, such a circle is obtained by fixing the elevation angle θ



(a) (b) (c)

Figure 5.4: Surface gradient directions recovered from 32 images of a helmet. Note that the gradients have been flipped (if necessary) to point in the direction of maximum image intensity. (a) Input images. (b) Quiver plot of gradient directions. (c) Zoom-in of quiver plot near right eye.

so that,

$$\mathbf{s}_\theta(\phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^\top \tag{5.7}$$

which induces a 1D emitted radiance function,

$$E(\phi) = E(\mathbf{s}_\theta(\phi)). \tag{5.8}$$

This 1D radiance function is guaranteed to be symmetric[2] about angle ϕ_g, the azimuthal angle of the surface normal with respect to pole $\mathbf{v}$ (or equivalently, the azimuthal angle of the surface gradient).

5.1.3 Symmetry Detection

In practice, we sample $E(\phi)$ at N uniformly spaced intervals and minimize the following objective to recover the symmetry angle ϕ_g at each pixel,

$$F(\phi_g) = \sum_{i=0}^{N} \min\left\{\eta; \frac{E(\phi_i)}{E(r(\phi_i,\phi_g))} + \frac{E(r(\phi_i,\phi_g))}{E(\phi_i)}\right\}, \tag{5.9}$$

[2]Ignoring non-local illumination effects; such effects are handled in practice by treating them as outliers.

where η is a threshold to account for outliers and $r(\phi_i, \phi_g)$ is a function mapping angle ϕ_i to its reflected position about angle ϕ_g. In our experiments, we use a threshold of either $\eta = 2.1$ or $\eta = 2.2$. As for the number of samples N, this is clearly related to the angular frequency of $E(\phi)$; for the materials in our experiments we found that 20 to 30 samples are sufficient for accurate reconstruction. It should also be noted that this objective is robust to outliers, caused for example by cast shadows and interreflections. Moreover, attached shadows actually preserve symmetry and thus do not violate our assumptions in Section 5.1.1.

Figures 5.3 and 5.4 demonstrate our ability to recover the gradient direction using this approach. The sphere dataset consists of 36 images rendered in POV-Ray with source directions separated by 10°. The helmet dataset consists of 32 images interpolated from a total set of 252 images taken about the sphere of lighting directions[3]; this corresponds to about 11° between light sources. In Figure 5.3, we see that the recovered gradient directions correctly point either toward or away from the center of the sphere. Figure 5.4 shows the recovered surface gradient directions for a helmet. While we do not have ground truth for this dataset, the gradients certainly seem plausible. It should be noted that symmetry is computed per-pixel, and thus our results are completely local. Finally, a number of additional results were generated on datasets with illumination spanning the entire sphere and/or upper hemisphere of incident directions. As seen in Figures 5.7-5.10, our technique operates robustly on a wide variety of surfaces, including ones with complex reflectance, fine-scale geometry, depth discontinuities, and interreflections.

5.2 Can Surface Constraints Resolve the Surface?

Based on the image acquisition setup described in Section 5.1.3, image measurements do not uniquely determine the surface normal, but constrain the surface normal to

[3]Data obtained from the Light Stage Data Gallery, ICT Graphics Lab, USC (Debevec et al., 2000; Chabert et al., 2006).

a plane. This is similar to shape from shading (Horn, 1970) where an image measurement constrains the normal to a cone, but does not fully determine the normal. In terms of the gradient of the height function, measurements provide the direction of the gradient, but not its magnitude. Here, we show that there is a family of surfaces that give rise to the same gradient direction at each surface point; thus surface constraints alone are insufficient to recover the full surface normal map. Yet, we show that for differentiable surfaces the isocontour structure can be recovered using only knowledge of the gradient direction.

5.2.1 A Class of Ambiguous Surfaces

Fact 4. *Consider two surfaces given by height functions $z = f(x, y)$ and $z' = g(x, y)$ with gradient fields ∇f and ∇g respectively. Then the two surfaces have the same gradient direction if*

$$\nabla g = k(x, y)\nabla f, \tag{5.10}$$

where $k(x, y)$ is some function of x and y that satisfies

$$\frac{\partial k}{\partial x}\frac{\partial f}{\partial x} = \frac{\partial k}{\partial y}\frac{\partial f}{\partial y} \tag{5.11}$$

for integrability to hold.

While we have not directly solved this system of partial differentiable equations, we do show the existence of a family of ambiguous surfaces,

Fact 5. *Two surfaces defined by $z = f(x, y)$ and $z' = h(z)$ will have the same gradient direction at each point if h is a differentiable function of z.*

Proof. This follows directly from Fact 4 and the chain rule,

$$\nabla h = \frac{\partial h}{\partial f}\nabla f \tag{5.12}$$

with $k(x, y) = \frac{\partial h}{\partial f}$. □

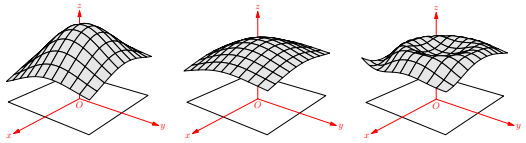


Figure 5.5: Three surfaces with the same gradient direction at each point.

This implies that without additional information or constraints, we can at best recover a surface up to an arbitrary function of the true height. Figure 5.5 shows a set of surfaces with the same gradient direction at each point to illustrate this ambiguity.

5.2.2 Gradient Direction Resolves Isocontour Structure

We now show that two surfaces with the same gradient direction at each point must have height functions that result in the same set of iso-depth contour curves,

Theorem 1. *Two surfaces defined by height functions $z = f(x, y)$ and $z' = g(x, y)$ that have the same gradient direction at each point must have the same set of iso-depth contour curves (i.e., curves of constant height).*

Proof. Suppose $(x(t), y(t))$ corresponds to an iso-depth contour curve of z. Then $z(t) = f(x(t), y(t))$ and the derivative of z with respect to t is,

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}. \tag{5.13}$$

Likewise, the derivative of z' with respect to t is,

$$\frac{\partial z'}{\partial t} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} \tag{5.14}$$

$$\frac{\partial z'}{\partial t} = k(t)\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}\right) \tag{5.15}$$

$$\frac{\partial z'}{\partial t} = k(t)\frac{\partial z}{\partial t}, \tag{5.16}$$

where Equation 5.15 holds from Fact 4 since z and z' have the same gradient direction at each point. Now note that $\frac{\partial z}{\partial t} = 0$ since $(x(t), y(t))$ is an iso-depth curve of z; combined

with Equation 5.16 it is clear that $\frac{\partial z'}{\partial t} = 0$ meaning $(x(t), y(t))$ is an isocontour curve of z' as well. □

Theorem 1 tells us that an isocontour of the true surface height must be an isocontour of any surface height function that has the same set of gradient directions. This is quite significant because it reduces the problem of finding the true height at every surface point to the problem of finding the true surface height of a single point on each isocontour curve. Another consequence of Theorem 1 is that it is possible to recover iso-depth contours of a surface given only the direction of the gradient at each point – the tangent of the iso-depth contour at a given point is orthogonal to the direction of the gradient. Thus, it is possible to obtain iso-depth contour curves by tracing in the direction orthogonal to the gradient.

It is also important to note that Theorem 1 does not imply that the only class of ambiguous surfaces are functions of the true height (e.g., Fact 5). For example, consider two surfaces : one composed of two non-intersecting hemispheres and another composed of the same hemispheres with one of the hemispheres raised higher than the other. The two surfaces have the same gradient direction at each point and share the same iso-depth contours; however, the second surface is not a function of the first surface's height.

5.2.3 Experimental Validation

To validate Theorem 1, we ran experiments on four datasets. For each surface shown in Figure 5.6 we first computed the gradient direction at each pixel as explained in Section 5.1.3. We then hand-selected a set of points on each surface and traced the iso-contour curves for some distance (~ 500 pixels) starting at those points. Clearly the isocontour structure for the synthetic data closely matches the ground-truth. Also, while the true surface for the helmet is unknown, the iso-contour curves look highly plausible, with the exception of regions with depth discontinuities (such as the ridge on the helmet) which violate our assumption of surface differentiability. Figure 5.11 shows results on

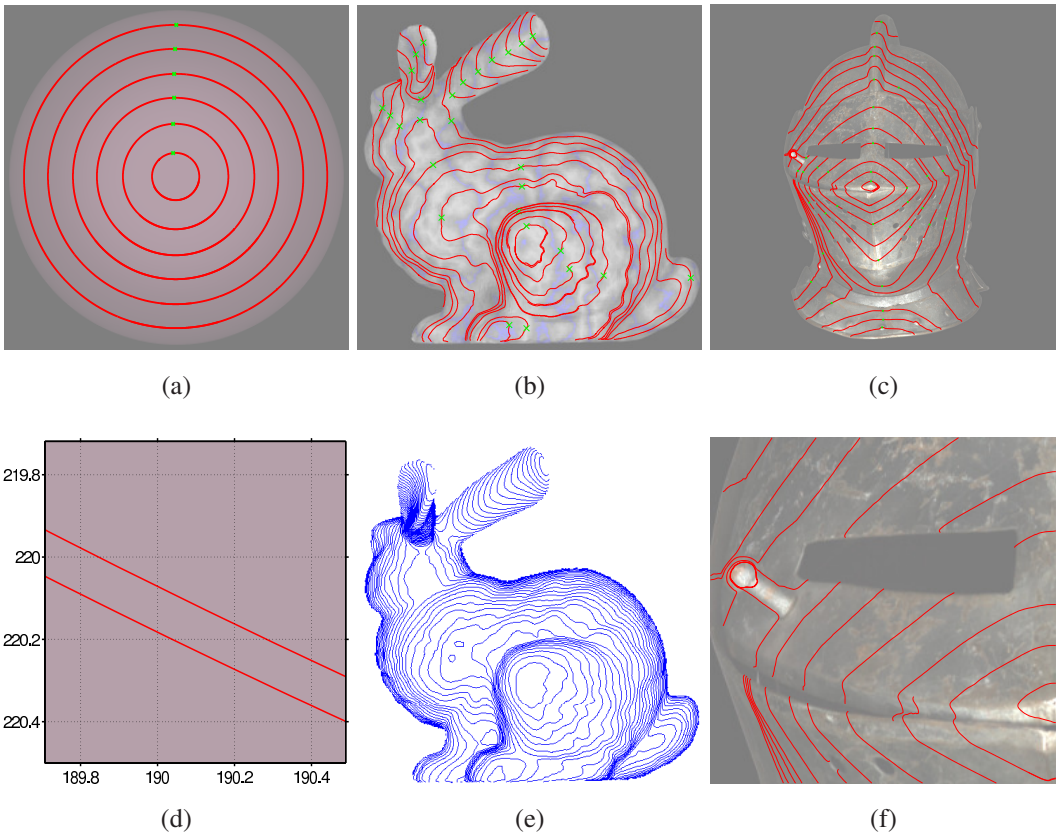


Figure 5.6: Recovered isocontour structure for three different data sets. (a) Synthetic sphere. (b) Synthetic bunny. (c) Helmet. (d) Zoom-in of inner-most isocontour. Notice the accumulated error after one loop is around 1/10th of a pixel. (e) Ground-truth isocontour map for the Stanford bunny. (f) Zoom-in of recovered isocontours on helmet.

a knight. This is a much more challenging dataset, yet we still obtain reasonable results for most of the surface.

5.3 Recovering the Full Euclidean Structure

Suppose we have recovered the direction of the gradient at each point as well as the iso-depth contours of a surface, but do not know the true height of the surface nor the remaining component of the surface normals. As suggested in the chapter open-

ing, to recover Euclidean structure we need to impose additional constraints (see Figure 5.1). Our options include (1) cast and attached shadows, (2) spatially uniform BRDF, (3) specularities, (4) multiple viewpoints, (5) parametric BRDF, (6) additional surface constraints, (7) structured lighting, and (8) heuristics such as in Ecker et al. (2007). Each of these constraints have previously been used in some form or another for surface reconstruction, but we have a distinct advantage since our surface is already highly constrained. In theory we only need to estimate a single value per iso-depth contour. In the following subsections we outline how one might fully constrain the surface by utilizing some of these constraints.

5.3.1 Shadow Constraints

Resolving structure from cast and attached shadows has been studied in some detail in the computer vision literature. Representative works include Shafer and Kanade (1983) who first describe the constraints that shadow boundaries impose on a surface; Kriegman and Belhumeur (2001) who show that the set of shadows produced by distant illumination can resolve the shape of a surface up to a generalized bas-relief transformation when the lighting is unknown, and Savarese et al. (2007) who implement a "shadow carving" algorithm.

The following facts capture the fundamental constraints that shadows provide,

Fact 6. *Consider a surface point that lies on an attached shadow boundary. Then the surface normal at that point must be orthogonal to the light source direction.*

Fact 7. *Consider a surface point* $\mathbf{p}_1$ *that lies on a shadow boundary cast by point* $\mathbf{p}_2$*. Then the difference in height between the two points can be determined from the light source direction.*

Fact 6 suggests that the surface normal can be fully determined at the intersection of attached shadow boundaries and Fact 7 implies that surface height can be determined between cast shadow boundaries and corresponding occluding points (which

are themselves attached shadow boundaries). A major hurdle to utilizing these facts is that detecting attached shadows is difficult to do reliably. However, if the iso-depth contours of the surface are known then constraints from noisy estimates of attached shadow boundaries can be distributed across entire isocontours, making the final surface estimate much more reliable than using shadows alone.

5.3.2 Uniform BRDF

If every surface point has the same BRDF then we can impose at least two additional constraints : constant brightness and reciprocity. The constant brightness constraint simply reflects the fact that points illuminated from the same light source direction will have the same intensity. Since we know the orientation of the surface normals and have measurements over a set of light source positions, we can effectively cluster the surface points according to the angle between the surface normal and viewing direction, $\mathbf{n}^\top \mathbf{v}$.

Helmholtz reciprocity imposes another set of constraints on the surface. Specifically, consider two surface points with surface normals $\mathbf{n}$ and $\mathbf{m}$ respectively. $\mathbf{n}$ and $\mathbf{m}$ are said to be reciprocal pairs under light source positions $\mathbf{s}_n$ and $\mathbf{s}_m$ if $\mathbf{n}^\top \mathbf{v} = \mathbf{m}^\top \mathbf{s}_m$ and $\mathbf{m}^\top \mathbf{v} = \mathbf{n}^\top \mathbf{s}_n$. It is well known that BRDFs which satisfy Helmholtz reciprocity are constant with respect to a reciprocal pair which means the image intensities corresponding to a reciprocal pair must satisfy,

$$E_n \mathbf{n}^\top \mathbf{v} = E_m \mathbf{m}^\top \mathbf{v}. \tag{5.17}$$

The problem then becomes one of isolating reciprocal pairs.

5.3.3 Specular Highlights

If we assume that the BRDF has a relatively tight specular lobe and that the specular lobe points in the idealized reflection direction, then we can directly recover the surface normal at positions corresponding to specular peaks. Consider a surface

point that coincides with a specular peak from light source direction $\mathbf{s}$. Then the surface normal at that point is given by the half angle between the viewing and source directions,

$$\mathbf{n} = (\mathbf{s} + \mathbf{v}) / \|\mathbf{s} + \mathbf{v}\|. \tag{5.18}$$

5.4 Summary

Reconstructing shape from images is of fundamental importance to computer vision, yet is a very challenging problem that requires many constraints to effectively solve in practice. An unfortunate consequence is that many of the constraints used for shape reconstruction are only physically valid for very limited types of objects, or are not physically valid at all (e.g., brightness constancy in structure from motion). In this chapter, we have shown how to utilize a relatively unexplored constraint for photometric stereo that is valid for arbitrary, unknown, and spatially varying isotropic materials. Much like Helmholtz reciprocity, bilateral symmetry is an important physical property of isotropic BRDFs that can and should be utilized when possible. Unlike most competing methods, we do not assume any parametric form for the BRDF and allow arbitrary spatial variation, making our technique the least restrictive to date with respect to assumptions on object reflectance.

While bilateral symmetry is only strong enough to constrain the surface normal at each point to a plane, we show how additional assumptions can be used to recover further structure. A particularly interesting case is surface differentiability which, when combined with bilateral symmetry, constrains the surface up to a set of iso-depth contour curves. This representation, which has been shown to be useful in its own right (Samir et al., 2006), reveals object topology (singular points and saddle points stand out for example) and reduces the surface ambiguity to a single value per iso-depth contour. In the following chapter, a method is presented for recovering both full Euclidean shape as well as the reflectance function at each surface point.

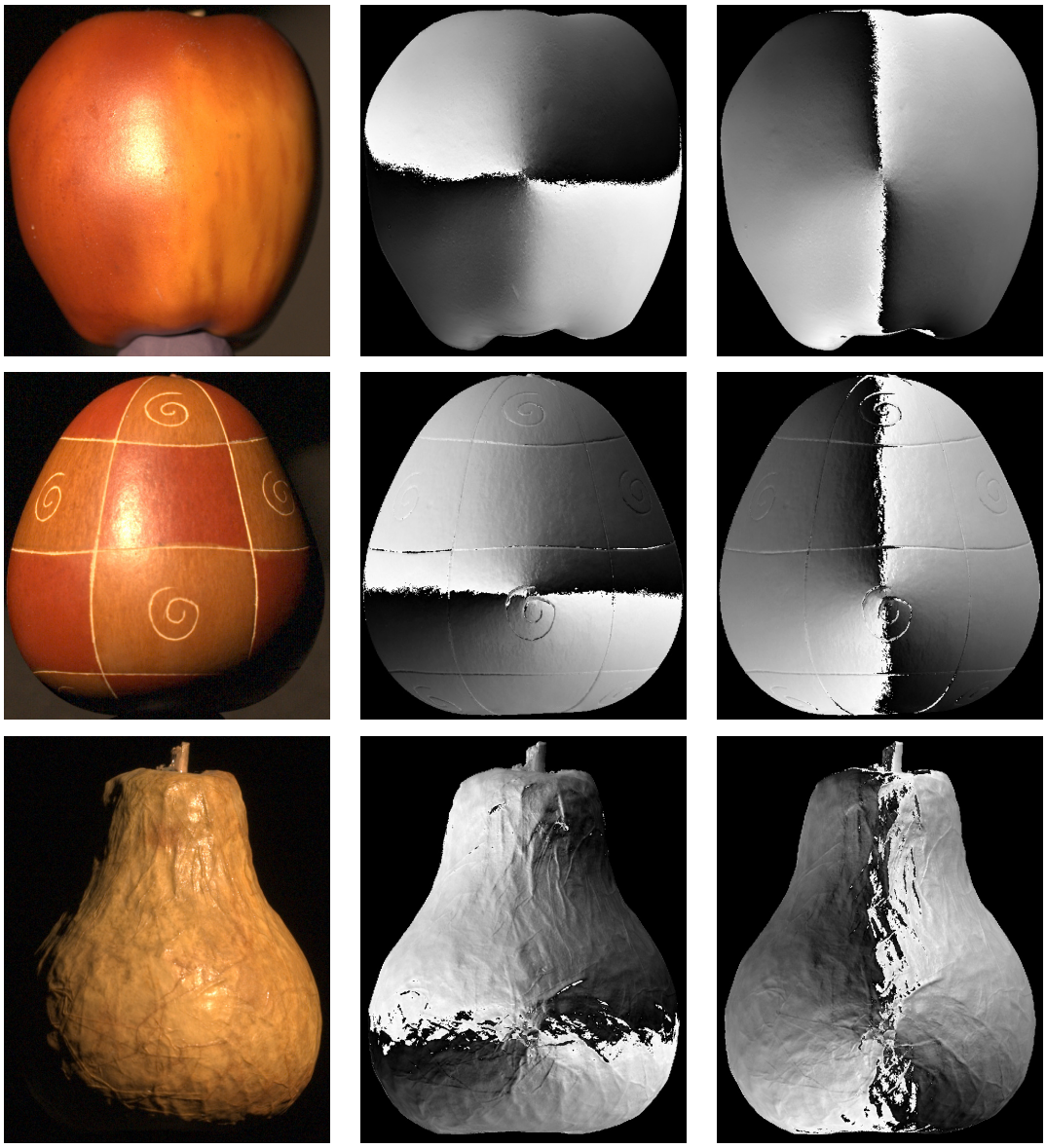


Figure 5.7: Recovered azimuth angles for various objects. (Left) A representative input image. (Center) Phase map of recovered azimuth angle, displayed as ϕ_n/π. (Right) Phase map shifted by an angle of $\pi/2$. Note: incident illumination distributed about the upper hemisphere of possible directions.

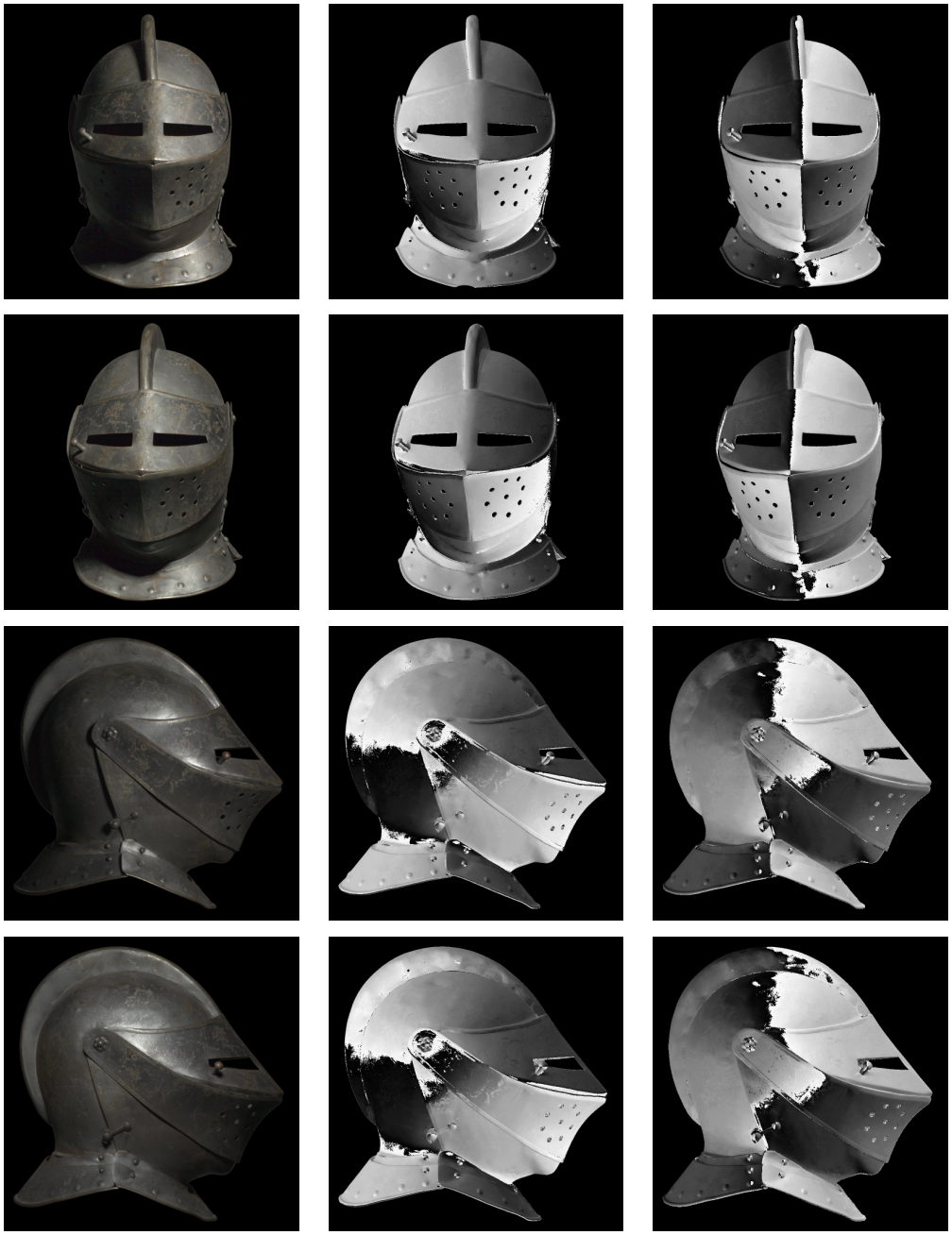


Figure 5.8: Recovered azimuth angles for various objects. (Left) A representative input image. (Center) Phase map of recovered azimuth angle, displayed as ϕ_n/π. (Right) Phase map shifted by an angle of $\pi/2$. Note: incident illumination distributed about the upper hemisphere of possible directions.

Figure 5.9: Recovered azimuth angles for various objects. (Left) A representative input image. (Center) Phase map of recovered azimuth angle, displayed as ϕ_n/π. (Right) Phase map shifted by an angle of $\pi/2$. Note: incident illumination distributed about the upper hemisphere of possible directions.

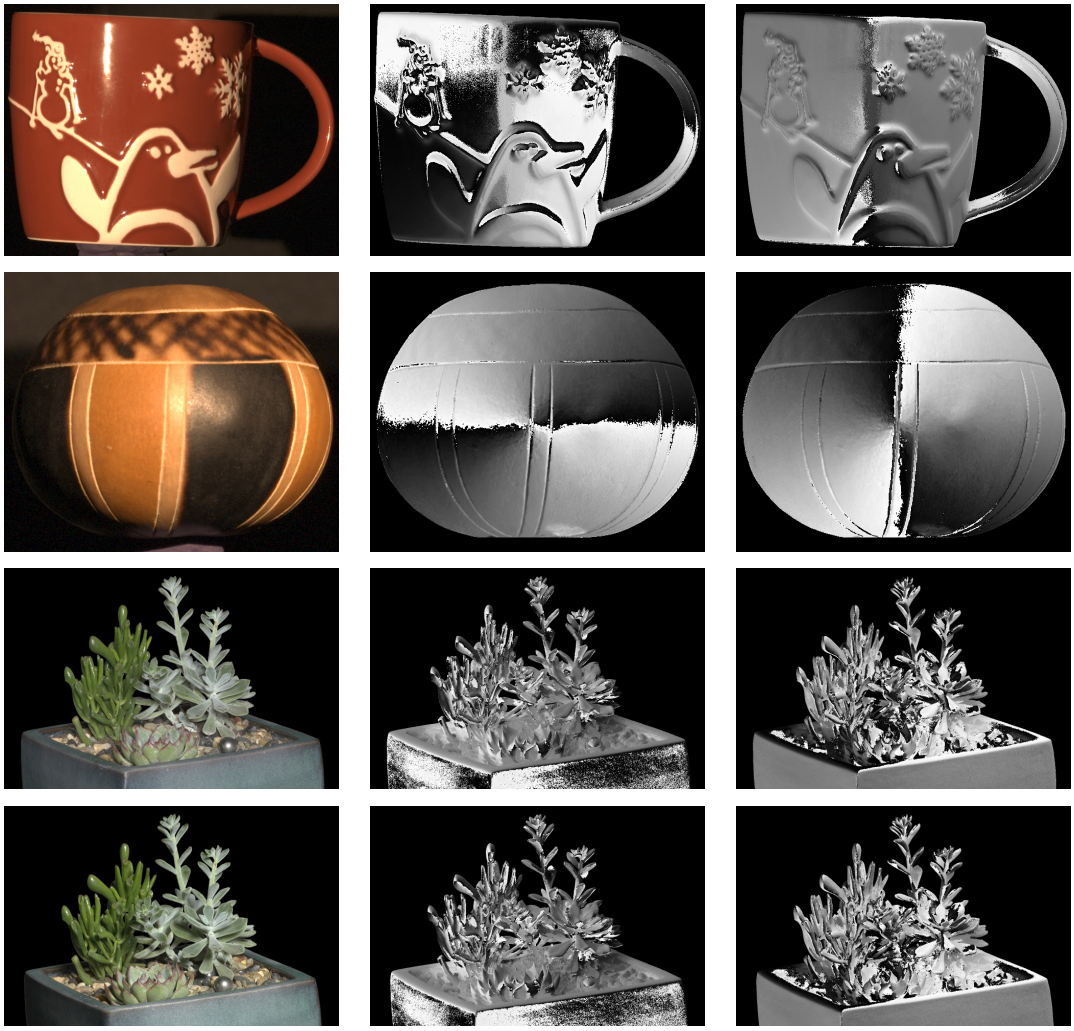


Figure 5.10: Recovered azimuth angles for various objects. (Left) A representative input image. (Center) Phase map of recovered azimuth angle, displayed as ϕ_n/π. (Right) Phase map shifted by an angle of $\pi/2$. Note: incident illumination distributed about the upper hemisphere of possible directions.

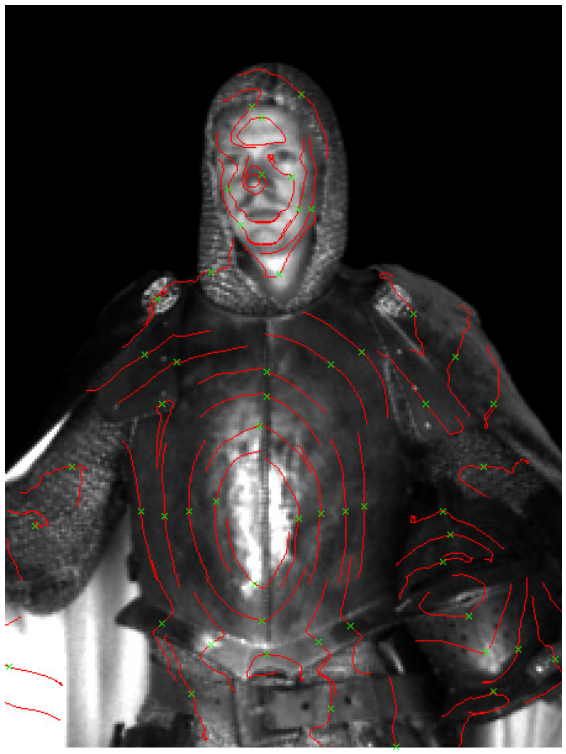


Figure 5.11: Recovered isocontour structure of a knight (Debevec et al., 2000; Chabert et al., 2006). Note that some regions (e.g., chain-mail) violate our assumptions of isotropy and/or surface differentiability.

Chapter 6

Photometric Stereo With Non-Parametric and Spatially-Varying Reflectance

Capturing the "appearance" of objects from images has become increasingly important in recent years, especially as computer graphics applications demand a level of photo-realism unattainable by hand modeling. By "appearance", we mean a model that is able to predict images of the object under all possible view and illumination conditions. To adequately sample the appearance of an object which truly varies arbitrarily in both shape and reflectance would require images from every combination of view and illuminant, which is impractical in most situations. Fortunately, objects in the real world typically exhibit regularity that can be exploited to drastically reduce the number of images required. Choosing constraints that are valid (or close to valid), yet powerful enough to be useful in practical systems is thus essential to appearance capture.

In this chapter, we consider the special case of photometric stereo – recovering an *explicit* appearance model (i.e., separate shape and reflectance models) from images taken at a single viewpoint under varying, known illumination. This is an important special case of appearance capture, since it relies solely on photometric cues, avoids

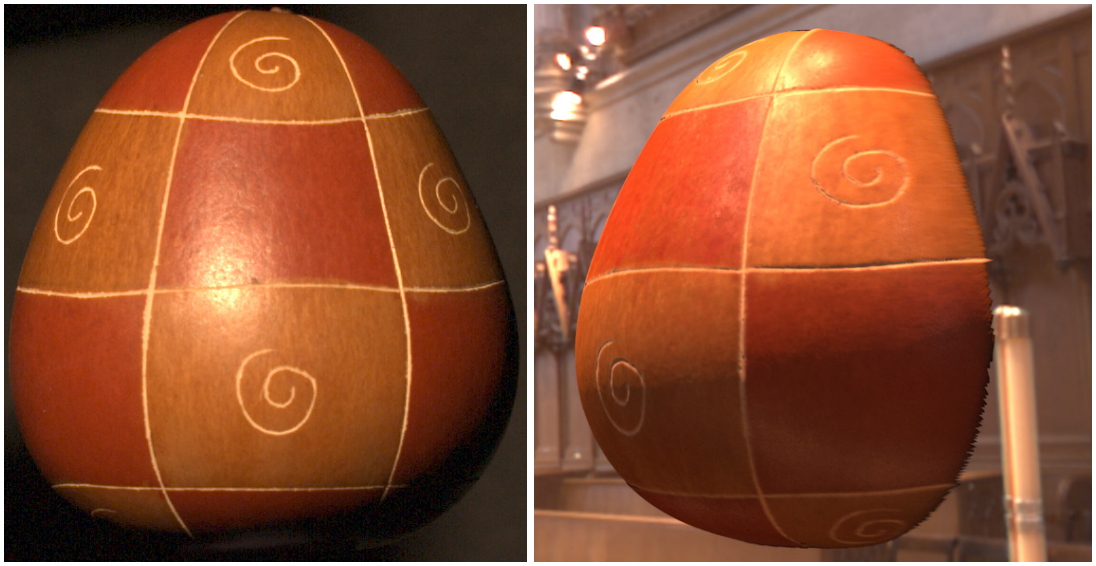


Figure 6.1: (Left) One of 102 single-viewpoint input images. (Right) Rendering from a novel viewpoint, using shape and reflectance acquired by our algorithm.

solving the correspondence problem, and is relatively simple to extend to multiple views if necessary. It is also important because explicit appearance models of this kind have been shown to be useful for visual tasks such as face recognition (Blanz and Vetter, 2003). Simultaneously recovering surface normals and reflectance from such input data remains a challenging problem, however. Typically, the form of the reflectance function is restricted by either assuming a parametric model or by the existence of a set of homogeneous reference objects in the scene. The obvious downside to these methods is reduced generality – if the materials in the object being measured differ from the assumed reflectance models, the accuracy of the recovered appearance model will be poor.

Our technique differs from most previous approaches in that we do not impose a parametric model on the reflectance function. Rather, we restrict the form of the reflectance function to satisfy empirically observed physical properties shared by many materials. These physical properties allow us to reduce the domain of the bi-directional reflectance distribution function (BRDF) from a function of four variables to a func-

tion of two variables without a significant loss in accuracy. This approximation has both theoretical and empirical motivation. Theoretically, Stark et al. (2005) have shown that many of the parametric BRDF models commonly used in computer graphics are in fact bi-variate functions, which suggests that bi-variate approximations can have at least some level of accuracy. They also show impressive empirical results for (an albeit small) set of measured isotropic BRDFs. We provide additional analysis in this chapter, both theoretical and empirical, to further support the validity of bi-variate BRDF approximations.

The method described in this chapter builds upon and improves three recent advances. First, we exploit isotropy to constrain surface normals to a single degree of freedom, as described in Chapter 5. Second, we utilize a non-parametric bi-variate approximation of the BRDF. Finally, we assume that surfaces are composed of a small number of "basis" materials and solve a factorization problem similar to that of Lawrence et al. (2006), but tailored to our differing setup (single viewpoint, recovery of surface geometry, fewer image measurements).

The main contributions of this chapter are (1) to present a technique capable of simultaneously recovering shape and non-parametric reflectance from photometric stereo, and (2) to introduce bi-variate representations of reflectance as a useful tool for vision applications.

6.1 Imaging Setup and Assumptions

Consider a photometric stereo setup with fixed object, fixed orthographic camera, and m images taken under distant point source illumination, with known source positions scattered about the sphere of incident directions. From this set of images, we wish to recover the surface normal and BRDF at each point on the object's surface. In Chapter 5 it was shown how to reliably recover the azimuthal component of each surface normal (relative to the camera coordinate frame), by assuming that the BRDF at each point is isotropic. The primary advantage of this approach is that the BRDF can vary

arbitrarily in both the spatial and angular domain[1], so long as the BRDF is isotropic. In this chapter, we seek to recover the elevation angle of the normal by imposing two additional constraints : (1) that the surface be composed of a small set of fundamental materials, and (2) that the BRDF at each point is well approximated by a bi-variate function.

More specifically, suppose the BRDF at each surface point is a linear combination of a small set of basis BRDFs. Then the BRDF at each point can be compactly represented as the product of two rank-constrained matrices,

$$\mathbf{H} = \mathbf{W}\mathbf{B}^{\top} \tag{6.1}$$

where $\mathbf{H} \in \mathbb{R}^{n \times d}$ is a discretization of the BRDF at each of n surface points, $\mathbf{B} \in \mathbb{R}^{d \times k}$ contains a discretization of k basis BRDFs, and $\mathbf{W} \in \mathbb{R}^{n \times k}$ weights the contribution of each basis BRDF at each surface point. For this decomposition to be physically valid, $\mathbf{W}$ and $\mathbf{B}$ should be non-negative and $\mathbf{B}$ should satisfy BRDF constraints such as energy conservation and reciprocity.

6.1.1 Bivariate BRDF Assumption

A general isotropic BRDF is a function of three dimensions, and is typically written $\rho(\theta_i, \theta_o, |\phi_i - \phi_o|)$, where (θ_i, ϕ_i) and (θ_o, ϕ_o) are the spherical coordinates of the directions of incident and reflected flux relative to a local coordinate system. (The absolute value, $|\phi_i - \phi_o|$, is sometimes discussed as a separate property called bilateral symmetry as in Chapter 5, but we do not do so here.) In what follows, it will also be convenient to represent the incident and exitant directions using unit vectors $\mathbf{s}$ and $\mathbf{v}$ in the same coordinate system.

An alternative parameterization is the halfway/difference parameterization of Rusinkiewicz (1998). Here, an isotropic BRDF is expressed as $\rho(\theta_h, \theta_d, \phi_d)$, where θ_h (the *half-angle*) is the angle between the surface normal and the bisector vector $\mathbf{s} + \mathbf{v}$,

[1] Spatial variation refers to changes in the BRDF across surface points; angular variation refers to changes in the BRDF across incident and exitant angles of illumination.

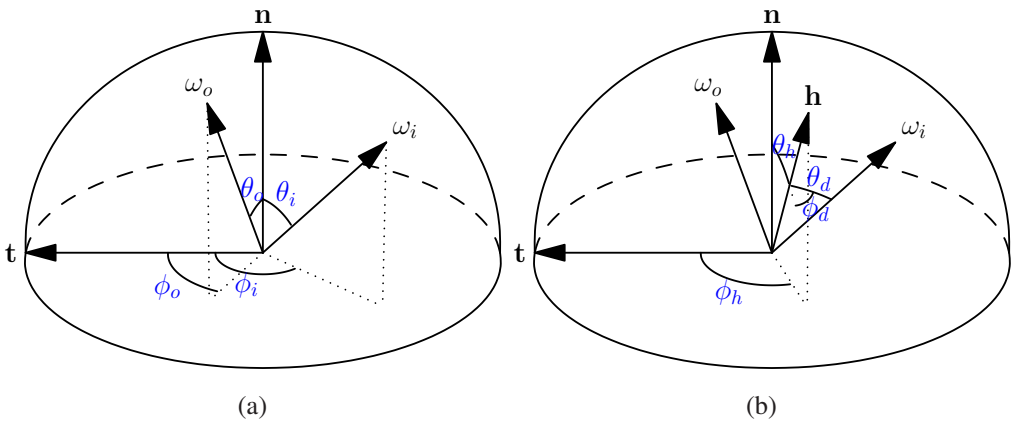


Figure 6.2: (a) The standard BRDF parameterization. (b) Half-angle parameterization.

and (θ_d, ϕ_d) are the spherical coordinates of the source vector computed relative to the bisector vector (see Figure 6.2). In particular, θ_d (the *difference angle*) is the angle between the source vector and the bisector vector.

Both of these parameterizations represent all three dimensions of the isotropic BRDF domain. The possibility that general isotropic BRDFs might be well-represented by simpler bi-variate functions was first formally studied by Stark et al. (2005). Their work is motivated by the observation that a number of parametric BRDF models (Lafortune, Phong, Blinn, and Ward) are inherently bivariate functions. Drawing from a combination of empirical observations and theoretical insights, they propose the '$\alpha\sigma$-parameterization' for bivariate BRDFs and show this to represent a small number of measured BRDFs (Westin, 2003) with high fidelity. In this paper, we use an alternative bivariate parameterization based on the half-way and difference angles, $\rho(\theta_h, \theta_d)$. One can show that there is a bijective mapping between (θ_h, θ_d) and (α, σ).

6.1.2 Image Formation Model

Suppose we know the true surface normal at each surface point. Then this imputes a half-angle for each surface point and light source direction from which we form

a data matrix, $\mathbf{E} \in \mathbb{R}^{n \times m}$. The i, jth entry is simply the image intensity at the ith surface point illuminated by the jth light source. If we assume the BRDF at each point is a linear combination of a set of basis BRDFs, then the BRDF of the ith point can be expressed as $\mathbf{H}_i^\top = \mathbf{w}_i^\top \mathbf{B}^\top \in \mathbb{R}^{1 \times d}$, where $\mathbf{w}_i \in \mathbb{R}^{k \times 1}$ is a set of material weights and $\mathbf{B} \in \mathbb{R}^{d \times k}$ contains a discretization of the basis BRDFs. The image intensity at the ith point under the jth illuminant is then modeled as,

$$\begin{aligned} e_{ij} &= \mathbf{H}_i^\top \mathbf{\Phi}_{ij} \max\{0, \mathbf{n}_i^\top \mathbf{s}_j\} \\ &= \mathbf{H}_i^\top \widetilde{\mathbf{\Phi}}_{ij} \\ &= \mathbf{w}_i^\top \mathbf{B}^\top \widetilde{\mathbf{\Phi}}_{ij} \end{aligned} \tag{6.2}$$

where $\max\{0, \mathbf{n}_i^\top \mathbf{s}_j\}$ accounts for shading and $\mathbf{\Phi}_{ij} \in \mathbb{R}^{d \times 1}$ is an interpolation vector mapping the domain of BRDF $\mathbf{H}_i$ to the half-angle / difference angle of the ijth measurement.

Equation 6.2 is easily extended to multiple color channels by slightly altering the BRDF matrix $\mathbf{B}$ and interpolation matrices, $\widetilde{\mathbf{\Phi}}_{ij}$. Suppose we wish to handle c color channels; then we simply fold each color channel into the BRDF discretization (e.g., $\mathbf{B} \in \mathbb{R}^{dc \times k}$) and modify the interpolation matrices appropriately. Alternatively, color can be encoded in the weight matrix $\mathbf{W}$, which allows arbitrary color scaling per point. This may be useful for surfaces that vary in color, but not in monochromatic reflectance.

6.2 Alternating Constrained Least Squares

If $\mathbf{W} = (\mathbf{w}_1, ..., \mathbf{w}_n)^\top$ and $\mathbf{B}$ are unknown, then we can estimate them using the method of alternating constrained least squares (ACLS), as described by Lawrence et al. (2006). ACLS works by alternately updating $\mathbf{W}$ and $\mathbf{B}$ to minimize the residual between measured intensities and predicted intensities. In each iteration, the material weights $\mathbf{W}$ are updated by fixing $\mathbf{B}$ and solving the resulting constrained convex optimization problem after which $\mathbf{B}$ is updated by fixing $\mathbf{W}$ and solving another constrained convex optimization problem. ACLS is guaranteed to find a local minimum since each update

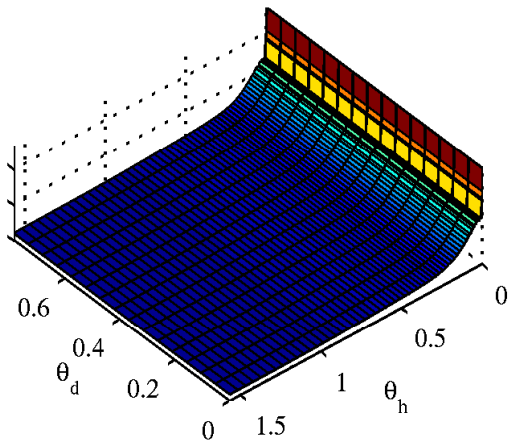


Figure 6.3: Surface plot showing discretization of one color channel of a basis BRDF (red channel of the 2nd basis BRDF recovered from the APPLE dataset).

step is guaranteed to not increase the residual. While this means the algorithm may not converge to a global minimum, in practice one can perform multiple trials with random initialization or use domain knowledge to initialize $\mathbf{W}$ and $\mathbf{B}$ near the optimal solution.

Since the elevation angles of the surface normals are also unknown, we also need to incorporate this into our optimization procedure. The simplest thing to do is to simply alternate between all three sets of parameters. However, since the normals are constrained to a single degree of freedom, it's possible to find a global minimum over material weights and surface normals simultaneously. This vastly improves convergence over three-way alternating optimization. We cover each step of our optimization procedure in the following subsections.

6.2.1 Initialization and Pre-Processing

The first step of our algorithm is to recover the azimuth angle of the surface normals using the technique described in Chapter 5. This step is based on the fact that the 2D reflectance field (image intensity as a function of source direction) is symmetric about the plane spanned by the normal and viewing direction. This plane, which corre-

sponds to the azimuth angle of the surface normal, can be estimated from a cone of light source directions parallel to and centered about the image plane. Thus, our algorithm also requires at least a cone of light source directions centered about the optical axis. Before starting the optimization process we also randomly initialize $\mathbf{W}$, $\mathbf{B}$, and θ_n (the elevation component of the surface normals).

6.2.2 Update $\mathbf{B}$ with Fixed $\mathbf{n}$ and $\mathbf{W}$

In this step, we solve for the BRDF matrix $\mathbf{B}$ that minimizes the L_2 error between image measurements e_{ij} and our image formation model $\mathbf{w}_i^\top \mathbf{B}^\top \widetilde{\mathbf{\Phi}}_{ij}$. From equation 6.2, we set up the following constrained least squares problem,

$$\begin{aligned} &\arg\min_{\mathbf{x}} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \\ &\text{Subject to,} \\ &\mathbf{x} \geq 0 \end{aligned} \tag{6.3}$$

where $\mathbf{x} \in \mathbb{R}^{dk \times 1}$ is a vector encoding the entries of $\mathbf{B}$ in column-major order. $\mathbf{A}$ and $\mathbf{b}$ can be constructed as,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{pmatrix} \tag{6.4}$$

$$\mathbf{A}_i = \mathbf{w}_i^\top \otimes \widetilde{\mathbf{\Phi}}_i^\top \qquad \mathbf{b}_i = \mathbf{E}_i^\top \tag{6.5}$$

where $\otimes$ denotes the Kronecker product.

6.2.3 Update $\mathbf{W}$ and $\mathbf{n}$ with Fixed $\mathbf{B}$

For the moment, suppose both $\mathbf{n}$ and $\mathbf{B}$ are fixed. From equation 6.2, we set up the following constrained least squares problem for each surface point,

$$\begin{aligned} &\arg\min_{\mathbf{w}_i} \quad \|\mathbf{A}_i\mathbf{w}_i - \mathbf{b}_i\|^2 \\ &\text{Subject to,} \\ &\mathbf{w}_i \geq 0 \end{aligned} \tag{6.6}$$

where,

$$\mathbf{A}_i = \widetilde{\mathbf{\Phi}}_i^\top \mathbf{B} \qquad\qquad \mathbf{b}_i = \mathbf{E}_i^\top \tag{6.7}$$

with $\mathbf{E}_i = (e_1, ..., e_m)$ the set of measurements at the ith surface point and $\widetilde{\mathbf{\Phi}}_i \in \mathbb{R}^{d\times m}$ the corresponding interpolation matrix. The solution to this optimization problem is the set of weights that minimizes the L_2 error of image measurements to intensities predicted by the image formation model, subject to non-negativity.

Since the weights for each point are estimated independently, the size of each constrained least squares problem is quite small (k variables). The global minimum with respect to both the weights and surface normal can be obtained by exhaustively searching over all possible $\mathbf{n}$ (tractable since there is only one degree of freedom).

6.3 Additional Constraints

In practice, we found it necessary to impose additional regularization constraints based on domain knowledge of our problem. Specifically, we impose smoothness and monotonicity over the BRDF domain, and we re-weight the constraints in Equations 6.3 and 6.6 to prevent specular highlights from dominating the solution. Empirically, these constraints improved convergence as well as the visual quality of the recovered basis BRDFs.

The need for regularization is caused by a number of factors. First, specularities usually occur in a very compact region of the BRDF domain, and within this region the

BRDF value can vary by orders of magnitude. As a result, these regions of the BRDF domain are very sensitive to misalignment of light sources; a very small misregistration can lead to large changes in predicted intensity. This is exacerbated by the fact that memory constraints prevent us from using all available pixel measurements when updating the BRDF matrix $\mathbf{B}$.

To introduce a bias toward smooth BRDFs, let $\mathbf{D}_1 \in \mathbb{R}^{d \times d}$ be a discrete operator approximating the gradient over the BRDF domain. We add the following quadratic penalty term to our objective function:

$$\lambda_{\mathrm{D1}}(\mathbf{D}_1\mathbf{B}_l)^\top(\mathbf{D}_1\mathbf{B}_l), \text{ for } l = 1...k. \tag{6.8}$$

This can be incorporated into Equation 6.3 by augmenting $\mathbf{A}$ and $\mathbf{b}$ with rows,

$$\mathbf{A}_{D1} = \sqrt{\lambda_{D1}}\left(\mathbf{I}_k \otimes \mathbf{D}_1^\top\right) \qquad\qquad \mathbf{b}_{D1} = \mathbf{0} \tag{6.9}$$

where $\mathbf{I}_k$ is a $k \times k$ identity matrix and $\otimes$ denotes the Kronecker product. In our experiments, we non-linearly weight the smoothness penalty so that specular regions (i.e., near $\theta_h = 0$) are penalized less strongly than non-specular regions.

Monotonicity can be enforced by adding the following inequality constraints:

$$B_{h,l} \geq B_{h+1,l}, \text{ for } l = 1...k. \tag{6.10}$$

It is also quite simple to enforce monotonicity over a portion of the domain (e.g., $\theta_h \in [0, \pi/4]$) by only including inequalities from the desired subset of the domain. Monotonicity is particularly important in specular portions of the BRDF domain, where undersampling and registration errors could otherwise cause unnatural visual artifacts in recovered BRDFs.

6.3.1 Confidence Weights

While there are relatively few measurements of specularities, such measurements carry a lot of weight since specular pixels typically have intensities more than

an order of magnitude stronger than other pixels. To prevent such measurements from overly biasing the final solution, we weight each constraint in Equations 6.3 and 6.6 according to the intensity of the corresponding measurement. In our experiments, we found the following ad-hoc weights to work well,

$$c_{ij} = (\log(1 + e_{ij})/e_{ij})^3. \tag{6.11}$$

6.4 Discussion on ACLS Procedure

While our optimization procedure is computationally similar to that of Lawrence et al. (2006), our methods differ in important ways. At a high level, our primary goal is to recover shape and reflectance in order to extrapolate appearance to novel viewpoints. Lawrence et al., on the other hand, assume they have data from multiple viewpoints as input and seek to obtain compact and separable representations of spatially varying BRDFs for editing purposes. Our data is also very different from Lawrence et al. (2006) in that we consider rather arbitrary geometry instead of focusing on near-planar surfaces.

The two approaches also differ at a more technical level. In their optimization, Lawrence et al. alternate between three sets of variables : BRDF basis, material weights, and surface normals. In this paper, we alternate over only two sets of variables because we find globally optimal material weights and surface normals in each iteration of our optimization algorithm. As a result, our method should be less prone to local minima. In addition, in order to bootstrap their reconstruction, Lawrence et al. use a parametric BRDF model (the Ward model), while in our work we have purposefully avoided the use of parametric BRDF models at any stage of the process. This yields an acquisition system for isotropic surfaces that is as general as possible. Another difference is how scattered data is handled. In Lawrence et al. (2006), measurements are interpolated in the BRDF domain, while in our method, the BRDF domain is interpolated onto the measurements. The effects of this change are twofold : (1) each measurement counts equally in our method, and (2) interpolation of the basis BRDFs is more numerically sta-

ble than interpolation of the measured data. A similar interpolation strategy is described in Weistroffer et al. (2007), although our method was derived independently.

6.5 Experimental Validation

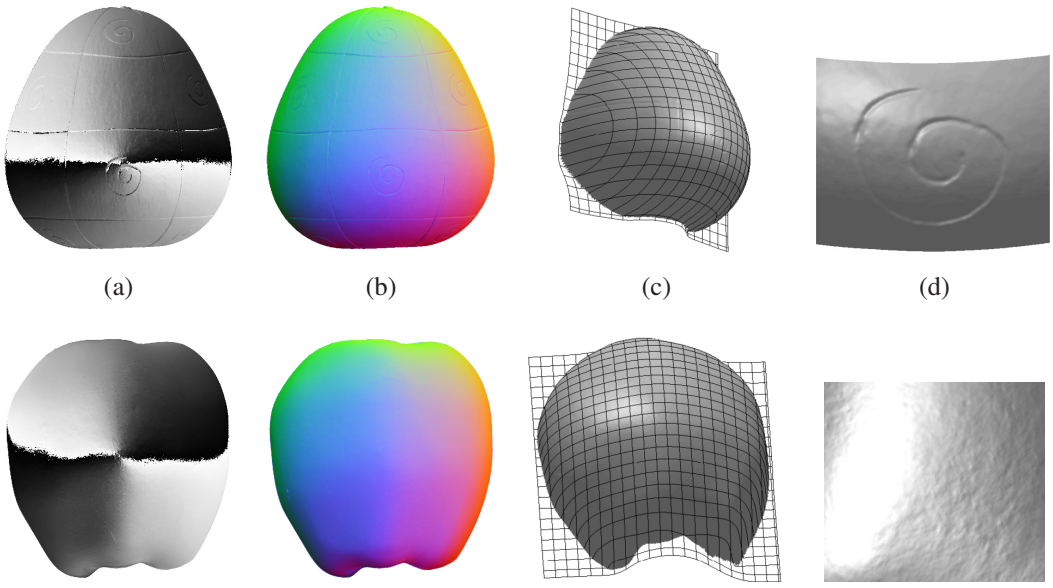


Figure 6.4: GOURD (Top) / APPLE (Bottom) shape reconstruction results. (a,e) Phase map showing the azimuthal components of the surface normal field, recovered as in Chapter 5. (b,f) Recovered normal map, encoded to RGB as $r = (n_x + 1)/2$, $g = (n_y + 1)/2$, $b = n_z$. (c,g) Surface obtained by integrating the recovered normal field. (d,h) Detail of the surface; note the recovered mesostructure.

To validate our approach, we ran experiments on two datasets consisting of images of a gourd and an apple, respectively. For each dataset, we acquired high-dynamic range images in a dark room (see Figure 6.6) with the camera and light sources placed between 1.5 and 2 meters from the test object (both test objects have diameter between 5 and 10 centimeters). Light source directions and intensities were measured from spec-

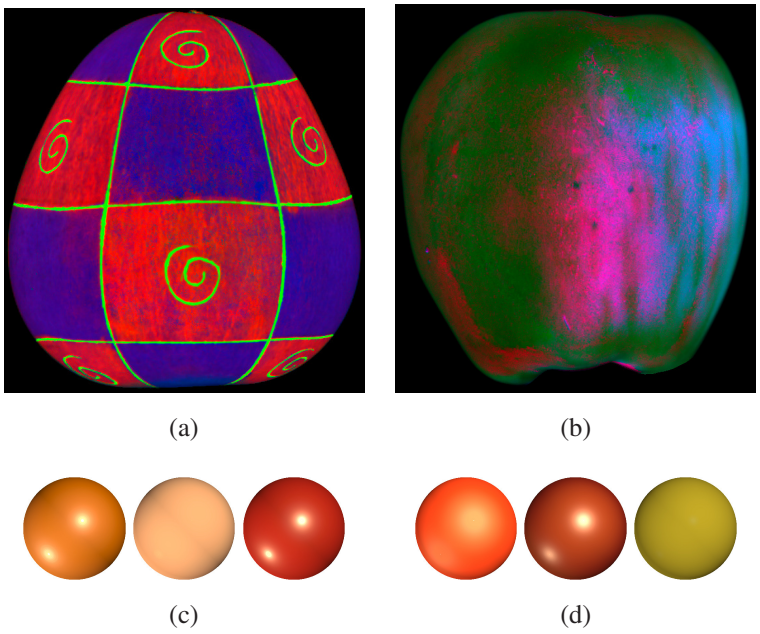


Figure 6.5: (Top) Material weight maps recovered from the GOURD and APPLE datasets. Red, green, and blue channels correspond to (normalized) weights of the first, second, and third basis BRDFs, respectively. (Bottom) Spheres rendered with the first, second, and third basis BRDFs recovered from the GOURD and APPLE datasets.

ular and diffuse spheres placed in the scene with sources spanning much of the upper hemisphere of lighting directions. 102 images were acquired for the GOURD dataset and 112 for the APPLE dataset. For both datasets, we assumed three basis BRDFs during reconstruction.

Figure 6.4 shows the shapes recovered by our algorithm on the GOURD and APPLE datasets. While the overall shape of each surface is simple (we sought to avoid cast shadows and interreflections which are not modeled by our algorithm), note that we accurately recover both the coarse and fine-scale geometric structure (i.e., macrostructure and mesostructure) of the object. In terms of appearance capture, recovery of surface mesostructure plays an important role (observe specular highlights in Figures 6.6 and 6.7).

Figure 6.5 shows the recovered basis BRDFs and material weight maps for the GOURD and APPLE datasets. Note the clear separation of materials visible in the material weight maps as well as the varying shape of specular lobes and body color in the recovered BRDFs.

The most important test of our algorithm is the ability to accurately generate novel views of the test objects. As seen in Figure 6.6, we are capable of rendering novel views that closely match real photographs. In particular, note the accurate reproduction of specular highlights which depend strongly on both the BRDF at each surface point as well as the surface mesostructure. As a final test, we rendered each object from a variety of viewpoints under complex illumination conditions (see Figures 6.1 and 6.7). Most of the illumination conditions in these results were measured from real-world scenes Debevec and Malik (1997) and the viewpoint was varied up to around 80 degrees from the original viewpoint of the training data. Even with extreme changes in viewing angle, the rendered images look natural and visually plausible. Additional results on two additional datasets can also be seen in Figure 6.8. While these are qualitative tests, the resulting images are highly realistic lending credence to the correctness of the reconstructions.

Finally, in Figure 6.9, we compare the output of our algorithm with that of Lambertian photometric stereo. Deviations from the true shape are clearly visible using both standard Lambertian photometric stereo and robust Lambertian photometric stereo. Most obvious is a bias toward non-frontal surface normals which are caused by specular highlights. Our method, on the other hand, faithfully captures the shape of each object.

Acknowledgments

The following projects were utilized in the creation of this chapter : PBRT Renderer, USC High-Resolution Light Probe Image Gallery, CVX, SDPT3, and UCSD FWGrid Project.

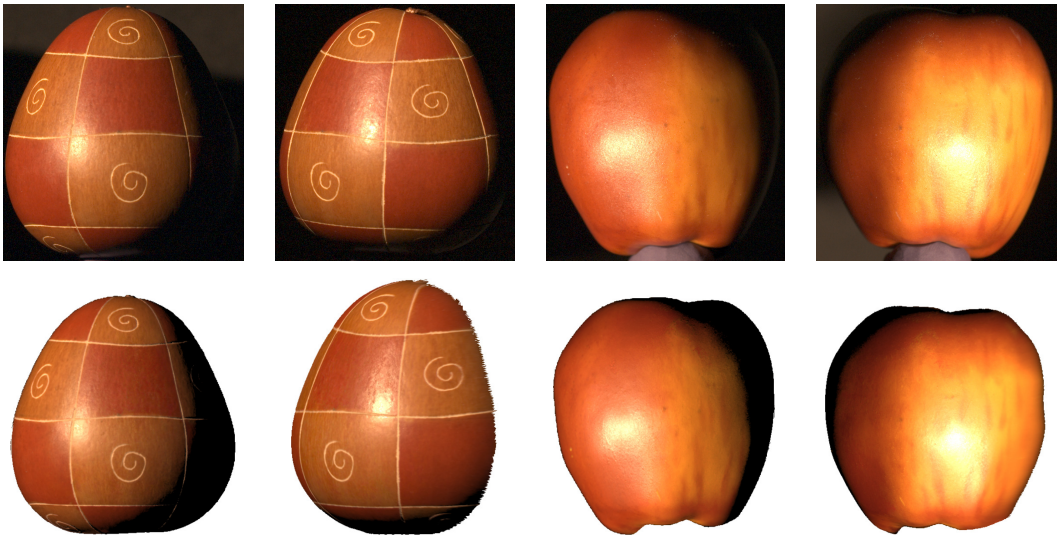


Figure 6.6: (Top) Real images of the GOURD and APPLE test objects. (Bottom) Images rendered using recovered shapes and BRDFs. Images in columns 1 and 3 are taken from the training data. Images in columns 2 and 4 are from novel viewpoints.

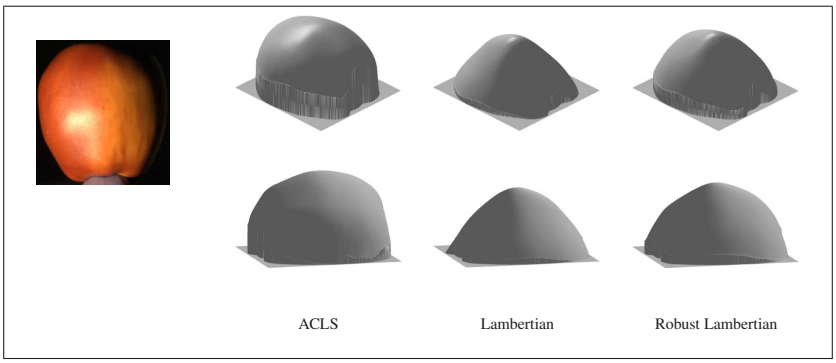
Figure 6.7: Images rendered in novel view and illumination conditions using shape and reflectance acquired by our algorithm.

(a) Real images from the HELMET and GOURD2 datasets

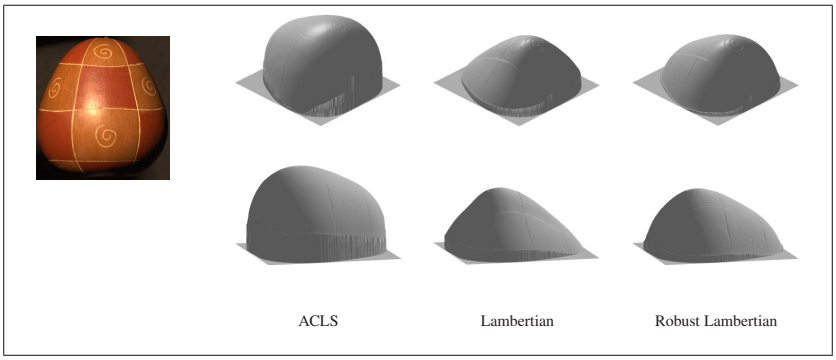
(b) Synthetic images, HELMET

(c) Synthetic images, GOURD2

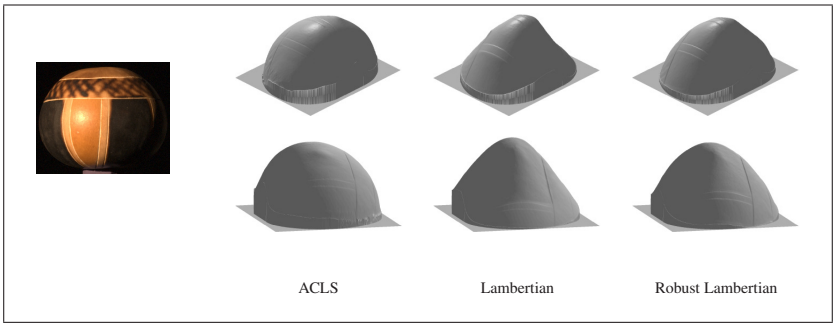
Figure 6.8: Results on two additional datasets. Images rendered in novel view and illumination conditions using shape and reflectance acquired by our algorithm.



(a) APPLE dataset



(b) GOURD1 dataset



(c) GOURD2 dataset

Figure 6.9: Comparison of shape recovered by our method to that of Lambertian photometric stereo. Significant distortions are noticeable in the shape recovered by standard Lambertian P.S., whereas our method faithfully recovers the shape. Even robust Lambertian P.S., which discounts specular pixels, leads to biased results.

Chapter 7

Multi-View Reconstruction With Non-Parametric and Spatially-Varying Reflectance

In this chapter, we consider multi-view reconstruction of objects with arbitrary isotropic reflectance. In particular, we consider the set of images of an object as it is rotated about a particular axis of rotation (on a turn-table, for example), with fixed camera and illumination. When the lighting is symmetrically distributed about the plane spanned by the viewing direction and the axis of object rotation, we show that it is possible to recover both the depth and one degree of freedom of the surface normal at each scene point, by utilizing photometric information (i.e., pixel intensities). Our underlying matching constraint is based on bilateral symmetry of the bi-directional reflectance function (BRDF), a property present in any physically valid isotropic material. Moreover, we do not impose any parametric form on the BRDF, and our method works for both uniform and spatially varying reflectance.

While shape-from-silhouette algorithms are also invariant to the bi-directional reflectance function (BRDF) of a surface, these methods are only capable of recovering the *visual hull* of an object. This means that any point on the surface that does not lie

on a silhouette boundary in some image will not be faithfully recovered. Another issue with shape-from-silhouette methods is that they yield volumetric reconstructions which typically do not accurately capture surface orientations across the surface. This can be problematic for when rendering shapes acquired using these methods. Our approach is similar to that of Hernandez et al. (2008) in that we initialize the surface shape using shape-from-silhouettes and then refine the surface using a photometric consistency measure. The primary difference is that Hernandez et al. assume Lambertian reflectance, whereas our photo-consistency measure is much more general.

7.0.1 Related Work

Because 3D reconstruction is such a long standing problem, many solutions have been proposed. These methods can be grouped according to the image features utilized and the camera and lighting configurations. Most similar to this chapter are multi-view reconstruction algorithms that incorporate photometric constraints.

In Lu and Little (1999), it is assumed that the object has a smooth surface with uniform and monotonic reflectance. They show that when the illumination is collinear with the viewing direction (for example, a point light source collocated with the camera), correspondence can be obtained for certain points on the surface lying on a silhouette boundary in some image. Because surface normals are known for points on a silhouette boundary, they search points with normals that, when rotated 90 degrees, are coincident with the viewing direction. Correspondence is made by noting that normals coincident with the viewing direction correspond to pixels of maximal brightness. Once correspondence is established for a few points, the BRDF (a uni-variate function of the incident/exitant angle) can be recovered, which is then used to recover surface normals across the entire surface. Our method differs in that we handle spatially varying BRDF and we we do not require collinear illumination (although that is a valid illumination for our method).

Related are so-called multi-view photometric stereo algorithms. Hernandez et al.

(2008) obtain images from multiple viewpoints under varying lighting conditions. From silhouette boundaries, they construct the visual hull of the object as well as calibrate the camera and lighting. They then refine the surface shape using Lambertian photometric stereo constraints. Zhang et al. (2003) and Lim et al. (2005) initialize a depth map using tracked feature points, which is then refined according to the Lambertian reflectance model. Joshi and Kriegman (2007) also utilize Lambertian reflectance, but in a novel way, by noting that a matrix of corresponding points with Lambertian reflectance has rank 3.

There has also been some work analyzing the particular constraints arising from circular motion. For example, Fitzgibbon et al. (1998) present a method for simultaneously recovering 3D structure, motion, and camera calibration parameters for images of an object rotated on a turn-table. Their method assumes feature points can be tracked across the object. Hernandez et al. (2007) show how to recover camera motion parameters from images of a circularly rotating object using silhouette information. This can be useful if camera calibration is unknown.

7.1 Image Formation Model

Consider a point on a surface at position $\mathbf{X}$ with surface normal $\mathbf{n}$ defined relative to some canonical coordinate system. Further suppose that the surface is illuminated by a set of distant light sources $\{\mathbf{s}_j\}, j = 1 \dots M$, and viewed by an orthographic camera in direction $\mathbf{v}$. Ignoring global illumination effects (such as cast shadows, interreflections, subsurface scattering, etc.), the exitant radiance at this point in direction $\mathbf{v}$ is given by,

$$e = \sum_{j=1}^{M} \rho(\mathbf{n}, \mathbf{s}_j) \max\{0, \mathbf{n}^\top \mathbf{s}_j\}, \tag{7.1}$$

where ρ is the BRDF at point $\mathbf{X}$ (dependence on viewing direction has been notationally excluded as it is constant in our setup).

Now suppose the object is rotated about the origin along the y-axis (without loss

of generality). Then the point $\mathbf{X}$ and normal $\mathbf{n}$ are transformed as,

$$\mathbf{X}_\phi = \mathbf{R}_\phi \mathbf{X} \tag{7.2}$$

$$\mathbf{n}_\phi = \mathbf{R}_\phi \mathbf{n} \tag{7.3}$$

with

$$\mathbf{R}_\phi = \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{pmatrix} \tag{7.4}$$

where $\mathbf{R}_\phi$ is a rotation matrix encoding a rotation of angle ϕ about the y-axis. The actual image coordinates are given by the first two components of vector,

$$\mathbf{x}_\phi = \mathbf{R}_c \mathbf{X}_\phi + \mathbf{t}_c \tag{7.5}$$

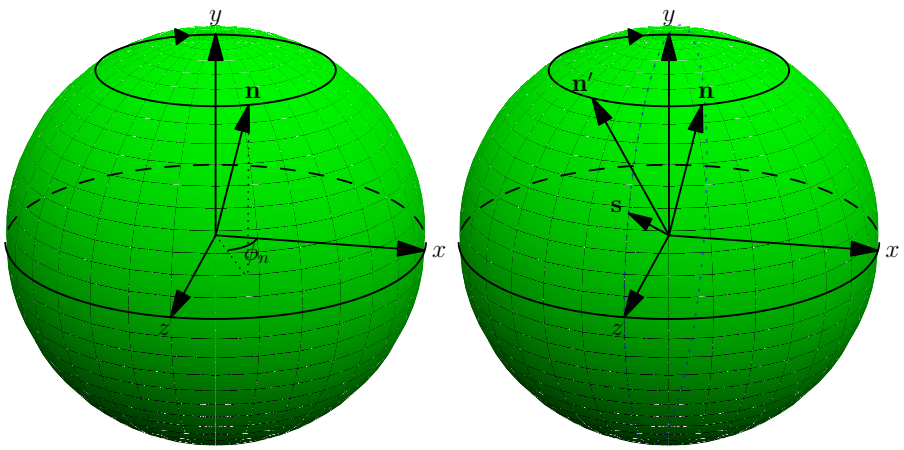
$$= \mathbf{R}_c \mathbf{R}_\phi \mathbf{X} + \mathbf{t}_c \tag{7.6}$$

where $\mathbf{R}_c$ and $\mathbf{t}_c$ map the object coordinate system to the camera coordinate system.

As ϕ varies from 0 to 2π, the surface normal sweeps a circle on the Gauss sphere (see Figure 7.1a). Notationally, we refer to the set of resulting pixel intensities as a function $e(\phi)$. In practice, we sample $e(\phi)$ (modelled by Equation 7.1) by capturing N images of the object as it is rotated. We also assume known rotations and camera extrinsic parameters (*i.e.*, $\mathbf{R}_\phi \; \forall\phi$, $\mathbf{R}_c$, and $\mathbf{t}_c$). These can be obtained through various calibration processes.

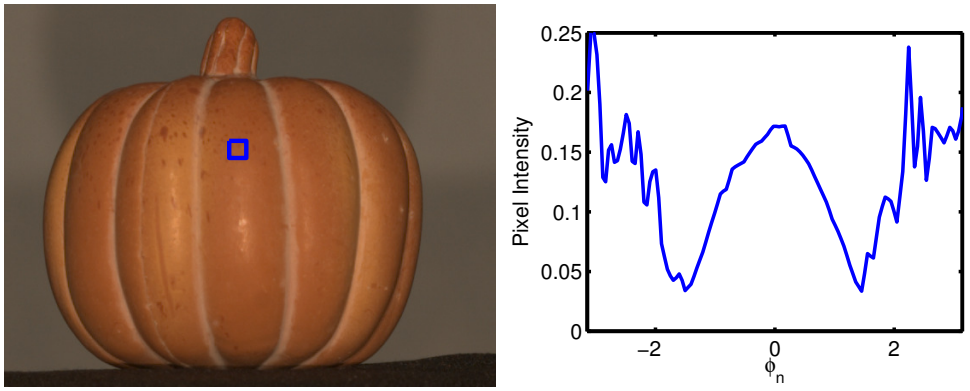
7.1.1 Isotropy / Bilateral Symmetry

The underlying constraint exploited by our algorithm is a property of isotropic BRDFs known as *bilateral symmetry*, which was also utilized in Chapter 5. Consider a particular point light source $\mathbf{s}$ and viewing direction $\mathbf{v}$ as unit vectors on the Gauss sphere (see Figure 7.1b). We partition the Gauss sphere into two halves, according to the plane spanned by $\mathbf{s}$ and $\mathbf{v}$. Further, consider each point on the Gauss sphere as a mapping of all possible (unit) surface normal vectors $\mathbf{n}$, and a BRDF as a function of $\mathbf{n}$. If the



(a) (b)

Figure 7.1: (a) The path swept by the surface normal of a point after rotation about the y-axis. (b) Illustration of surface normals with identical intensity, due to isotropic reflectance. The image intensity resulting from surface normal $\mathbf{n}$ is identical to the intensity resulting from rotated normal $\mathbf{n}'$, where $\mathbf{n}'$ is obtained by rotating $\mathbf{n}$ to a position of equal angle from the plane spanned by $\mathbf{s}$ and the rotation axis. Omitted from the figure for clarity is the viewing direction $\mathbf{v}$ which is coplanar with $\mathbf{s}$ and the rotation axis y.



(a) (b)

Figure 7.2: Example showing effect of bilateral symmetry on image intensity. (a) 1 of 64 input images rotated on a turntable. (b) Plot of image intensity at a single "frontal" surface point after rotations of ϕ_n radians. Notice that the image intensity is symmetric between angles of $-\pi/2$ and $\pi/2$.

BRDF is symmetric about the plane spanned by $\mathbf{s}$ and $\mathbf{v}$, then it is bilaterally symmetric. In other words, the BRDF has identical value for normals mirrored about the dividing plane. Moreover, because the shading term $\max\{0, \mathbf{n}^\top \mathbf{s}\}$ is also symmetric about this plane, the exitant radiance (proportional to image intensity) is also symmetric about this plane. This is analogous to the argument presented in Chapter 5, except in this case the light source(s) are fixed and the surface normal changes. The symmetry argument can be further extended by noting that any lighting configuration symmetrically distributed about the dividing plane will result in the exitant radiance function being symmetric about the dividing plane. Thus, the exitant radiance from a scene point will contain a symmetry for any distribution of lighting that is symmetric about the plane defined by viewing direction $\mathbf{v}$ and the axis of rotation.

A consequence of bilateral symmetry is that the image intensity of a surface point will have identical intensity after rotations that map the surface normal to opposite sides of the Gauss sphere. In terms of our imaging setup, consider a surface point with a normal lying in the plane spanned by the rotation axis and the viewing direction. Then

the pixel intensities for this point after rotations of angles ϕ and $-\phi$ will be identical (see Figure 7.2). This is the underlying photometric property we utilize in this work.

7.2 Matching Constraint

The argument presented in the previous section implicitly assumes known point correspondence. That is, it is necessary to know which pixel a given surface point maps to for each rotation, or equivalently it is necessary to know the 3D location of each surface point. Knowing only the rotation and camera parameters, however, constrains correspondence (or 3D position) only up to a one parameter family of solutions, according to the epipolar geometry. If we consider the canonical coordinate frame (*i.e.*, $\mathbf{R}_0 = \mathbf{I}$), then we can re-express Equation 7.6 as,

$$\mathbf{X} = \mathbf{R}_c^\top(\mathbf{x} - \mathbf{t}_c). \tag{7.7}$$

The x and y components of $\mathbf{x}$ are known (they are simply the pixel coordinates of the point in the canonical reference frame) and the z component defines the family of potential 3D positions for that point. Given z, the 3D position $\mathbf{X}$ can be computed and then reprojected onto each image.

Previously, we defined $e(\phi)$ to be the set of pixel intensities arising from each possible rotation. We can augment this function to also incorporate the unknown depth z; that is, $e(z, \phi)$. For correct values of z, $e(z, \phi)$ will result in pixel intensities drawn from a single point on the object surface. In this case, we expect the function to be symmetric as described in Section 7.1.1. When z does not correspond to a surface point, intensities will be drawn from multiple surface points as the object is rotated and, in general, we expect the resulting intensities will not be symmetric. Figure 7.3 shows an example of the intensity function $e(z, \phi)$ resulting from real images.

To quantify symmetry, we use the following robust cost function, defined for

each pixel in the canonical reference frame,

$$c(z,\phi_n) = \int_{\alpha=0}^{\pi/2} \min\left\{\eta, \frac{e(z,\phi_n-\alpha)}{e(z,\phi_n+\alpha)} + \frac{e(z,\phi_n+\alpha)}{e(z,\phi_n-\alpha)}\right\} d\alpha. \tag{7.8}$$

Here, ϕ_n is the angle about which $e(z,\phi)$ is symmetric, or equivalently the azimuthal component of the surface normal $\mathbf{n}$ when encoded in spherical coordinates relative to the canonical coordinate system (with pole y). η is a threshold intended to minimize the effects of outliers. To facilitate evaluation of the integral, we discretize $e(z,\phi)$ and approximate the integral with a summation. In the absence of outliers, our cost function has global minima at true surface points / symmetry angles. These minima will also be unique in the absence of noise and if sufficient geometric and/or texture variation exists across the surface.

7.2.1 Uniqueness Conditions

It is useful to characterize the uniqueness of minimizing Equation 7.8. First, there are generally at least two depth values that correspond to real surface points (*i.e.*, the front and back surface of the object relative to the canonical frame). In the absence of noise and outliers, $c(z,\phi_n)$ will have a global minima at each depth corresponding to a true surface boundary.

We now consider the cost at depths not lying on a surface boundary. Consider the surface shown in Figure 7.4. We see that when depth is correct, the set of pixel intensities will all come from a single point on the surface; however, if the depth is incorrect the set of pixel intensities will be drawn from a set of points lying on a curve along the surface boundary. If this set of intensities is symmetric, then the cost will be minimized. Generally this will occur quite rarely, but there is a special case that can happen more frequently. Consider a surface point $\mathbf{X}$ whose local neighborhood is both textureless and geometrically symmetric. As seen in Figure 7.4b, a non-surface point will yield symmetric intensity values in this case. In fact, there is a family of such points lying on the line passing through $\mathbf{X}$ in the direction defined by the intersection of the surface normal with the rotation plane. Since most surfaces are both textureless and geometrically

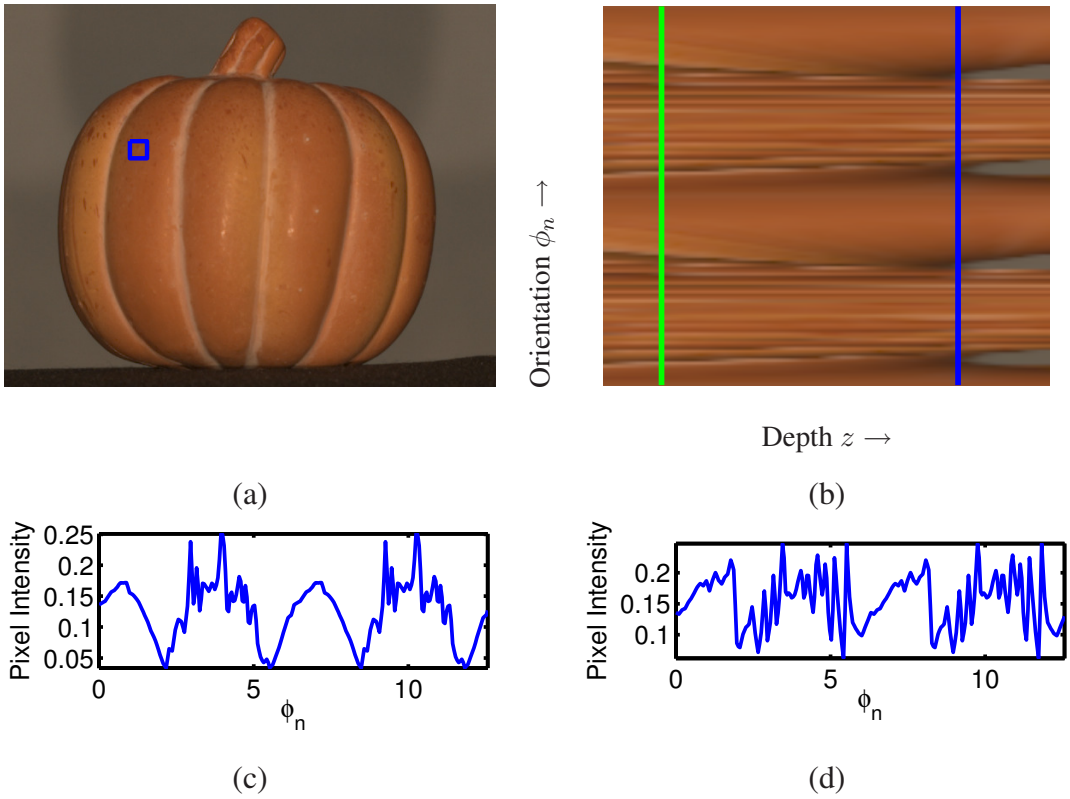
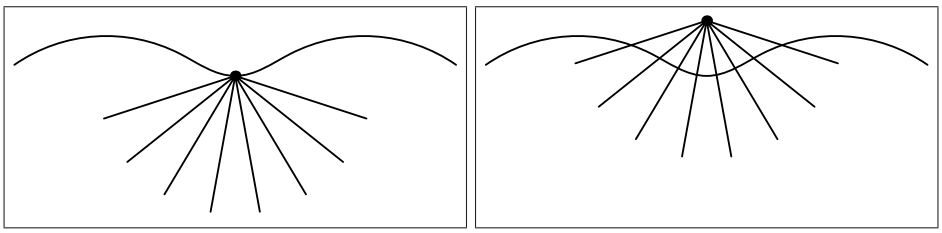


Figure 7.3: Example showing pixel intensity as a function of surface orientation and depth. (a) One of 64 input images. The blue box surrounds a candidate surface point. (b) The set of intensities $e(z, \phi)$ corresponding to the candidate point. Note: ϕ_n varies from 0 to 4π instead of 0 to 2π for visual clarity at the boundary. (c) Plot of intensity values at the true depth value (blue line). Note there is a clear symmetry present near $\phi_n = 7$; symmetry is only present about a range of $\pm\pi/2$ which corresponds to frontal poses of the surface point. (d) Plot of intensity values at an incorrect depth value (green line). There is no obvious symmetry in this case.



(a) (b)

Figure 7.4: Illustration of pixel intensities resulting from incorrect correspondence. (a) When correspondence is correct, all pixel intensities are drawn from a single surface point. (b) When correspondence is incorrect, pixel intensities are drawn from a 1D slice of neighboring surface points. If the shape and reflectance of the local neighborhood is symmetric, then so will the profile of pixel intensities.

symmetric[1] at small scales, this suggests that $c(z, \phi)$ will not be very discriminative with respect to z. Indeed, this has been empirically observed in experimental data.

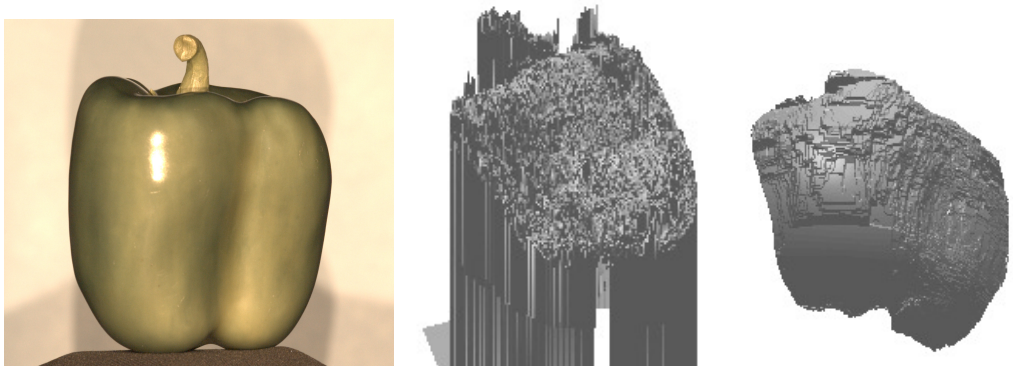
While this is clearly non-ideal, not all is lost. While the cost function may not be highly discriminative with respect to depth, it is discriminative with respect to surface orientation ϕ_n. Even at incorrect depths, only the true surface orientation yields a minimum. This suggests a scheme whereby surface depth is regularized by surface orientation values.

7.3 Experimental Validation

7.3.1 Imaging Setup

We acquired images in a dark room with a Canon EOS-1Ds camera. Prior to image capture, the camera was calibrated to account for both radiometric and geometric distortions. Since the camera has a linear response curve, radiometric calibration only involved estimating the vignetting falloff. We performed this step by imaging a diffuse,

[1]Relative to a frontal viewing angle.



(a) (b) (c)

Figure 7.5: Results with and without surface smoothness. (a) Input image from the reference viewpoint. (b) Heightmap obtained by minimizing $c(z, \phi)$ independently at each point. (c) Heightmap obtained by enforcing smoothness using graph-cuts.

planar surface, which should result in uniform image intensity if no vignetting is present; thus, the image provides a direct, per-pixel estimate of the vignetting falloff. We reduce the effects of noise by imaging the plane multiple times under various distant lighting configurations. For geometric calibration, we imaged a planar checkerboard pattern (4-6x) and used (Bouguet, 2008) to compute the intrinsic camera parameters. Object rotation was achieved using a turntable; the image of the axis of rotation was estimated by imaging a calibration object under a set of rotations.

7.3.2 Results

Ideally, we would be able to obtain a good surface by minimizing Equation 7.8 independently for each pixel in the reference view. However, because of noise, outliers, and the possibility of multiple minima (as described in Section 7.2.1) this does not yield a consistent result. As seen in Figure 7.5b, the surface obtained is extremely noisy, although the overall shape (including concavities) can still be discerned.

One way to overcome the weak discriminativeness of Equation 7.8 is to impose surface smoothness. A popular method for combining per-node data costs with a local

smoothness penalty is to use graph-cuts (Boykov et al., 2001; Kolmogorov and Zabin, 2004; Boykov and Kolmogorov, 2004). In figure 7.5c, we see that the surface smoothness constraint regularizes the solution and the resulting surface is much more clearly defined. It should be mentioned that the surface orientation ϕ_n was not incorporated into this solution. The data cost term used in our graph-cuts algorithm is,

$$d(z) = c(z, \phi'), \tag{7.9}$$

where

$$\phi' = \arg\min_{\phi} c(z, \phi).$$

That is, for each depth value, the data cost is the minimum cost over all orientation values.

For technical reasons it is not possible to incorporate an *orientation consistency* constraint directly into the graph-cuts minimization. However, once a surface is obtained from graph-cuts, it can be refined by adjusting the heightmap to be consistent with the orientation ϕ_n at each point. At each point, ϕ_n specifies the component of the tangent plane orthogonal to the rotation axis. In the special case where the rotation axis is the y-axis, this is equivalent to knowing the x component of the surface gradient (p) at each point. A refined surface can thus be obtained by minimizing the following expression,

$$z' = \arg\min_{z'} \int\int \left[\lambda(z - z')^2 + (p - z'_x)^2\right] dxdy. \tag{7.10}$$

Here, λ controls how far the final solution should deviate from the original surface. Figure 7.6 shows the surface after refinement. Note there are still a few "streaks" in the reconstruction, which correspond to outliers in the detected surface orientation. Overall, the result is quite reasonable.

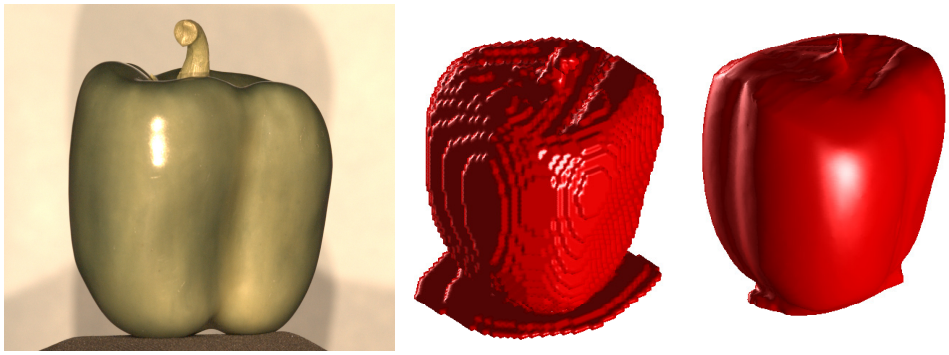
Another way to recover the surface shape is to utilize silhouette information. By intersecting the volumes obtained by back-projecting the silhouette of each image into the scene it is possible to recover the *visual hull* of a surface. Any surface point lying on a silhouette boundary in some image (such points are called *contour generators*) also

Figure 7.6: Refined surface obtained by enforcing orientation consistency.

lie on the boundary of the visual hull. However, points which are not contour generators lie strictly in the interior of the visual hull. Such points correspond to concavities on the true surface.

While the visual hull does not provide the true surface, it does provide a good approximation which can be refined using our photometric constraint. In particular, our photometric constraint provides information about both depth and surface orientation at each point. With the visual hull initialization it is possible to iteratively refine the surface to find a locally optimal surface. We operate over a mesh initialized to the visual hull and optimize over both the photometric cost $c(z, \phi)$, as well as surface smoothness and surface *orientation* constraints. The surface orientation constraint minimizes the difference between photometric orientation ϕ_n and the surface normal defined by the mesh at each vertex.

In Figure 7.7, we see the input images, visual hull initializations, and refined surface for the PEPPER dataset. Note that concave regions of the surface are not recovered by the visual hull. However, by utilizing our photometric constraint we are able to recover the concavities.



(a) (b) (c)

Figure 7.7: Results using visual hull inititialization. (a) An input image. (b) Visual hull obtained by back-projecting silhouette boundaries. (b) Refined surface obtained by enforcing photometric constraint.

7.4 Summary

In this chapter, we have presented a novel constraint for multi-view reconstruction that is applicable to objects composed of arbitrary isotropic materials. In contrast to previous work, our method is applicable to both textured and textureless surfaces and can handle complex, non-Lambertian reflectance.

Chapter 8

A Planar Light Probe

Images of a scene are formed by the interaction of the camera viewpoint, the scene geometry and reflectance, and also upon the way the scene is illuminated. Each component can have a large impact on the final image that is formed. In this chapter, we focus on the illumination or lighting component of the image formation process. Specifically, we consider the problem of measuring or estimating the lighting present in a scene. The output of our method is a frequency space approximation of the 2D environment map for a given scene.

The ability to measure or estimate the illumination in a scene has a wide range of applications. Knowledge of the lighting in a scene is explicitly required in many computer vision techniques such as shape from shading, photometric stereo, shadow carving, and shape from specularities. It can also be utilized in tasks such as face recognition or object detection. In computer graphics, measured illumination environments are commonly utilized to render photo-realistic scenes, and in augmented or mixed reality applications knowledge of the illumination environment is usually necessary to convincingly insert virtual objects into a scene.

A common approach to measuring the illumination in a scene is to acquire images of a reference object, or light probe, placed into the scene. We follow this paradigm, but with a significant twist: rather than exploit varying geometry across the reference

object, we instead utilize spatially varying *reflectance*. While both geometry and reflectance could be varied, we restrict our attention to the case of a *planar* light probe in order to study what is possible when the geometry does not vary.

Our approach is highly novel and will hopefully inspire further research in related areas of computer vision. Additionally, the design of our light probe confers an advantage over competing methods: namely, the illumination can be estimated from a single low-dynamic range image. This is significant because typical scenes are illuminated by a combination of extremely bright, compact light sources as well as darker, but much larger area sources. To faithfully capture the illumination, both types of illumination need to be measured.

8.1 Background and Motivation

The output of an illumination estimation process is some representation of the lighting. In general, lighting is a 5-D function specifying the intensity of light traveling in each direction at every position in the scene. However, it is common and often effective to treat light sources as being infinitely distant from the objects in a scene in which case the illumination can be fully specified as a 2D function defined over a sphere of directions; this function is referred to as an *environment map*.

A popular approach to lighting estimation is to directly sample the 5-D or 2-D illumination function, as in Debevec and Malik (1997). Other work assumes that the lighting is composed of a small number of point light sources located nearby or at a distance (Pentland, 1982; Zheng and Chellappa, 1991; Kim et al., 1998; Miyazaki et al., 2003; Zhang and Yang, 2001; Wang and Samaras, 2002). These approaches return the coordinates and strengths of the sources. In general, these methods can be viewed as assuming a parametrized generative model of lighting and attempt to estimate the parameters. Others take a non-parametric approach by representing lighting as a linear superposition of some set of basis functions, and lighting estimation amounts to estimating the coefficients for each basis function as in Marschner and Greenberg (1997).

Lighting and reflectance have also been shown to be well characterized by spherical harmonic basis functions, and theoretical (Basri and Jacobs, 2003; Ramamoorthi, 2002) as well as empirical evidence (Epstein et al., 1995; Lee et al., 2005) support the idea that a low order expansion of lighting is sufficient for rendering Lambertian scenes. More recently, Haar wavelets have also been used to estimate lighting from cast shadows (Okabe et al., 2004).

A wide variety of techniques have emerged for measuring lighting, and to understand their relative advantages, it is helpful to first consider the different attributes of illumination capture. (1) Does the technique require inserting a physical probe in the scene, require knowledge of the scene geometry, or can it passively infer lighting directly from images of an unknown scene? (2) Does it provide lighting as a 2-D or 5-D function? (3) Does it produce a low or high dynamic range (LDR or HDR) illumination map? (4) Does it require a single image (implying applicability to video with dynamic lighting) or does it use multiple images, perhaps to construct an HDR image? (5) What is the size, bulk, and cost of the probe? (6) What is the resolution (spatial frequency response) of the output? An ideal technique would passively infer lighting from a single image, would provide a high resolution 5-D light field, would produce HDR output, would be applicable to video, and would be low cost.

No technique meets this ideal. Without a probe, the problem is ill-posed and requires some sort of prior to arrive at a solution such as in Pentland (1982). Therefore, we will consider techniques that require insertion of a probe into the scene and that treat lighting as infinitely distant (*i.e.*, a function on a sphere). The most straight forward technique for distant lighting is simply to use a camera to directly measure the light, for example by using a fish eye lens (Greene, 1986; Haeberli and Segal, 1993) or a catadioptric omni-directional camera (Nayar, 1997). Alternatively, a standard camera can observe a mirrored sphere placed in the scene, which reflects light from all directions (though at a lower resolution toward the occluding contour). While these techniques provide high spatial resolution, they require high-dynamic range (HDR) imaging to fully characterize dynamic range of most lighting environments (Debevec and Malik, 1997;

Nayar and Mitsunaga, 2000); this is accomplished by capturing multiple LDR images, and is therefore unsuitable for video. In addition, cameras or spherical probes in the scene can be relatively expensive and/or bulky.

A second approach is to introduce a matte light probe into the scene which can be a sphere as in Zheng and Chellappa (1991), Zhang and Yang (2001), and Wang and Samaras (2002), or an object with known shape as in Weber and Cipolla (2001). When considered in terms of the spherical harmonic expansion of the BRDF, mirrored and Lambertian spheres couldn't be more different. A mirrored sphere, whose impulse response is akin to a delta function, passes all spatial frequencies whereas the Lambertian sphere acts as a low pass filter and severely attenuates high frequencies. It is argued that only an expansion to 3rd order can be recovered from a Lambertian sphere. However, the advantage of a Lambertian probe is that the dynamic range of the images of a sphere under any lighting condition is low, and so lighting can be estimated from a single image, making the technique suitable for video. A slightly more sophisticated approach is to image a set of spheres with varying reflectance (matte, glossy, and specular) to recover lighting (Debevec et al., 2004). This approach can still encounter problems with limited frequency response and / or dynamic range.

A third way of constructing a probe is to take advantage of non-convex geometry and the resulting shadows (Sato et al., 1999, 2001, 2003; Okabe et al., 2004). Consider a sundial. The irradiance arriving at a particular point on the underlying surface is the product of the incident lighting with the visibility function induced by the geometry of the sundial. In Okabe et al. (2004), the relation between lighting and cast shadows are analyzed in the frequency domain in terms of spherical harmonics and Haar wavelet bases. In Ramamoorthi et al. (2005) an analysis of cast shadows is provided using the Fourier domain.

In the following sections, we introduce a fourth mechanism for measuring lighting which utilizes a light probe consisting of a planar surface with spatially varying *reflectance*. The light probe is constructed using a multi-layered, transparent medium which differentially absorbs and reflects light. In particular, the homogeneous middle

layer of the planar probe is a transparent medium (e.g., air or glass), the top layer is patterned to partially absorb and transmit light differently at different locations (e.g., a transparency sheet). The bottom layer is also patterned to reflect light differently at different locations (e.g., a piece of printed paper). The key idea is that the design of the two patterns leads to a spatially varying effective BRDF. That is, when image plane irradiance is averaged over an area (within a single pixel or over multiple pixels), the manufactured probe can be treated as a meso-scopic geometric structure akin to the micro-facet or pit models used in constructing models of BRDFs (Torrance and Sparrow, 1967; Oren and Nayar, 1994; Koenderink et al., 1999). Given a set of these BRDFs distributed over the plane, we can recover lighting over the upper hemisphere by treating each BRDF as a basis function. A special case of the analysis is when the upper and lower patterns are binary (completely opaque or transparent) in which case the effective BRDF is solely the result of shadowing and masking (Oren and Nayar, 1994).

The advantage of such a light probe is that its capabilities lie between a mirrored and Lambertian probe in that it can measure higher order frequencies than a Lambertian sphere, yet high dynamic range imaging is not needed. As a consequence, it is suitable for capturing lighting in video. Furthermore, a particular application of this probe is augmented reality where it is common to include a planar geometric probe with fiducial markers (see for example ARToolkit) for determining relative orientation, and our illumination probe could be readily integrated for lighting estimation and photo-consistent rendering.

In the rest of this chapter, we first introduce a design for the proposed light probe, and theoretically characterize the effective BRDF of the probe as a function of the patterned upper and lower layers. We then discuss probe construction, and explain how lighting can be estimated from the probe. Finally, we report experimental results that validate the potential of the planar light probe.

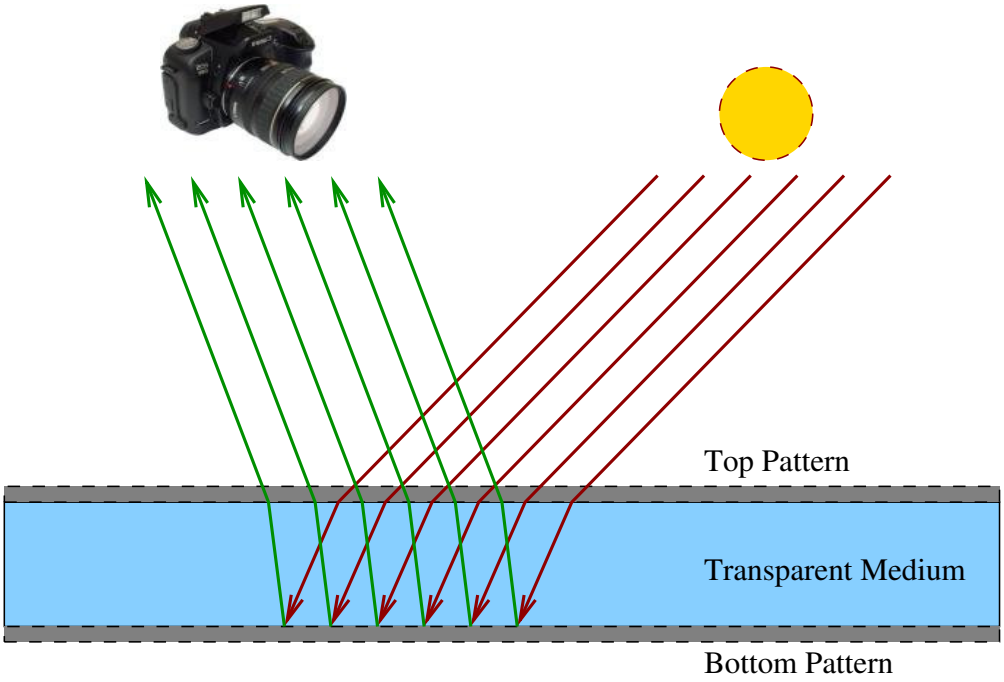


Figure 8.1: Imaging setup for the planar light probe.

8.2 Designing a BRDF for Lighting Recovery

Suppose we have a material consisting of two parallel layers, referred to as the top and bottom layers, separated by a transparent medium (see Figure 8.1). Further suppose both layers are spatially varying so that the top layer reflects and transmits light as a function of position and the bottom layer reflects light as a function of position. We assume that the wavelength of light is much smaller than both the thickness of the transparent layer and the spatially varying patterns so that no diffraction effects occur. When light hits the top layer, some is directly reflected and some is transmitted through the transparent medium where it is reflected by the bottom layer, travels back through the transparent medium, and finally passes out through the top layer. We seek to analyze the reflectance properties of such a material with the ultimate goal of recovering the lighting from a set of distinct BRDFs constructed in this way.

We denote the position on the plane by $\vec{x}$, the incident radiance from direction $\vec{\omega}_i$ arriving at position $\vec{x}$ as $l_i(\vec{x}, \vec{\omega}_i)$, and the reflected radiance in direction $\vec{\omega}_r$ exiting at

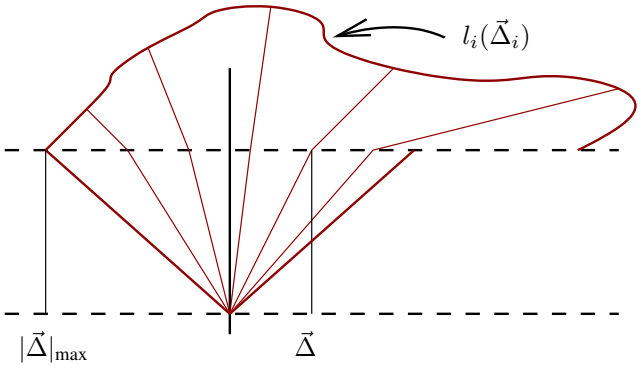


Figure 8.2: Mapping of the incident lighting onto the plane via refraction.

position $\vec{x}$ as $l_r(\vec{x}, \vec{\omega}_r)$. We relate the incident radiance from differential solid angle $d\vec{\omega}_i$ at position $\vec{x}$ to the reflected radiance at position $\vec{x}'$ in direction $\vec{\omega}_r$ through the BSSRDF $S(\vec{x}, d\vec{\omega}_i \rightarrow \vec{x}', \vec{\omega}_r)$ so that the reflected radiance at position $\vec{x}$ in direction $\vec{\omega}_r$ is

$$l_r(\vec{x}, \vec{\omega}_r) = \int_{\vec{x}' \in A} \int_{d\vec{\omega}_i \in \Omega} l_i(\vec{x}', d\vec{\omega}_i) S(\vec{x}', d\vec{\omega}_i; \vec{x}, \vec{\omega}_r) d\vec{\omega}_i^{\mathbf{N}} dA \tag{8.1}$$

where $\vec{\omega}_i^{\mathbf{N}}$ is the projected solid angle onto the plane with surface normal $\mathbf{N}$. This expression is stating that all light arriving in some area A contributes to the reflected radiance at position $\vec{x}$ according to BSSRDF S.

We now begin to specify S. First, we split it into a reflective term and a scattering term,

$$S(\vec{x}, d\vec{\omega}_i \rightarrow \vec{x}', \vec{\omega}_r) = f_r(\vec{x}, d\vec{\omega}_i \rightarrow \vec{\omega}_r) + f_s(\vec{x}, d\vec{\omega}_i \rightarrow \vec{x}', \vec{\omega}_r). \tag{8.2}$$

The reflective term, $f_r(\vec{x}, d\vec{\omega}_i \rightarrow \vec{\omega}_r)$, is just a standard spatially varying BRDF and represents the light directly reflected off the top surface. The scattering term, $f_s(\vec{x}, d\vec{\omega}_i \rightarrow \vec{x}', \vec{\omega}_r)$ represents light passing through the top layer at position $\vec{x}'$ with solid angle $d\vec{\omega}_i$ and re-emerging at position $\vec{x}$ in direction $\vec{\omega}_r$.

Assuming no absorption as light travels through the transparent medium, specular transmission through the top layer, and that scattering beyond the initial reflection at the bottom layer is negligible, we can split the scattering term f_s into the product of

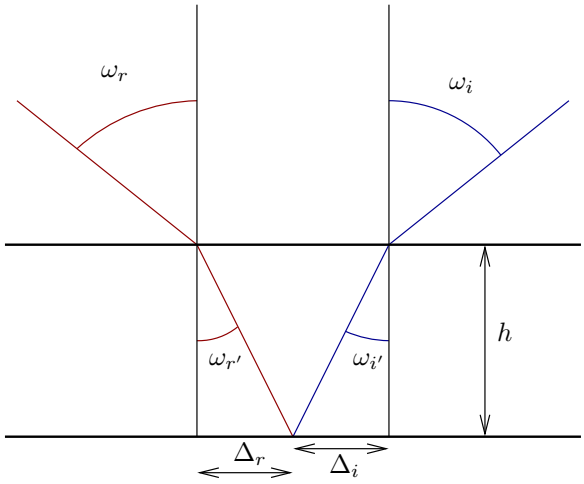


Figure 8.3: Relationships between various angles and distances.

three parts,

$$\begin{aligned} f_s(\vec{x}, d\vec{\omega}_i \to \vec{x}', \vec{\omega}_r) = f_t(\vec{x}, d\vec{\omega}_i) F_i(\vec{\omega}_i) \ldots \\ \cdot\, g_r(\vec{x} + \vec{\Delta}_i, d\vec{\omega}_{i'}, \vec{\omega}_{r'}) d\vec{\omega}_{i'}^{\mathbf{N}} \ldots \\ \cdot\, F_r(\vec{\omega}_r) f_t(\vec{x}', d\vec{\omega}_r) d\vec{\omega}_{r'}^{\mathbf{N}} \end{aligned} \tag{8.3}$$

where $f_t(\vec{x}, d\vec{\omega}_i)$ represents the initial transmission through the top layer, g_r is the reflection from the bottom layer, $f_t(\vec{x}', d\vec{\omega}_r)$ is the transmission out through the top layer, and $F_i(\vec{\omega}_i)$ and $F_r(\vec{\omega}_r)$ are Fresnel transmission terms for entering and leaving the transparent medium. $\vec{x} + \vec{\Delta}(\vec{\omega}_i)$ is the position where a given ray of light hits the bottom layer.

Now observe that if we fix the exitant angle, then only a single incident angle contributes to the scattering term. To understand this, note that only a single position $\vec{x}'$ contributes to the exitant radiance, which is completely specified by $\vec{\omega}_i$, $\vec{\omega}_r$, the thickness of the transparent medium h, and Snell's law (specified with the indices of refraction η_1 and η_2 for the outside medium and transparent medium respectively). We can take this idea one step further by parametrizing the incident and exitant hemisphere in terms of $\vec{\Delta}_i$ and $\vec{\Delta}_r$ respectively, which denote the distance along the plane between where a ray intersects the top layer and bottom layer. We can now specify the scattering term as a

function of $\vec{x}$, $\vec{\Delta}_i$, and $\vec{\Delta}_r$,

$$\begin{aligned} f_s(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) = f_t(\vec{x}, \vec{\Delta}_i) F_i(\vec{\Delta}_i) \ldots \\ \cdot g_r(\vec{x} + \vec{\Delta}_i, \vec{\Delta}_i, \vec{\Delta}_r) d\vec{\omega}_{i'}^{\mathbf{N}} \ldots \\ \cdot F_r(\vec{\Delta}_r) f_t(\vec{x} + \vec{\Delta}_r + \vec{\Delta}_i, \vec{\Delta}_r) d\vec{\omega}_r'^{\mathbf{N}}. \end{aligned} \tag{8.4}$$

If the incident lighting is distant, then it remains constant across the plane and we can parametrize it entirely in terms of $\vec{\Delta}_i$. Furthermore, if the refractive index of the transparent layer is higher than that of the outside environment, then because of Snell's law, $|\vec{\Delta}_i|$ will be bounded by some finite value dictated by the critical angle and the thickness of the medium. Without loss of generality, we scale the coordinate system along the plane so that $|\vec{\Delta}_i|_{\max} = \frac{1}{2}$. To simplify subsequent integrals, we define the incident lighting so that $l(\vec{\Delta}) = 0$ for all $|\vec{\Delta}| > \frac{1}{2}$ (where it was previously undefined).

Putting everything together, we can rewrite equation 8.1 as,

$$l_r(\vec{x}, \vec{\Delta}_r) = \int_{\vec{\Delta}_i \in A} l_i(\vec{\Delta}_i) S(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) \frac{d\vec{\omega}_i^{\mathbf{N}}}{dA} dA \tag{8.5}$$

where $A = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$.

8.2.1 Fourier Analysis

In the previous section we formulated the exitant radiance as the integral of the incident lighting $l_i(\vec{\Delta}_i)$ and simplified BSSRDF $S(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r)$. We now turn our attention to the *average* behavior of the system. To simplify the math, we fold the projected solid angle terms and the Fresnel transmission terms into modified versions of f_r, f_t, and g_r so that,

$$\widetilde{f}_r(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) = f_r(\vec{x}, \vec{\Delta}, \vec{\omega}_r) \frac{d\vec{\omega}_i^{\mathbf{N}}}{dA} \tag{8.6}$$

$$\widetilde{f}_t(\vec{x}, \vec{\Delta}) = f_t(\vec{x}, \vec{\Delta}) \tag{8.7}$$

$$\widetilde{g}_r(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) = g_r(\cdots) F_i(\vec{\Delta}_i) F_r(\vec{\Delta}_r) \frac{d\vec{\omega}_i^{\mathbf{N}}}{dA}. \tag{8.8}$$

We can then write equation 8.5 as

$$l_r(\vec{x}, \vec{\Delta}_r) = \int_{\vec{\Delta}_i \in A} l_i(\vec{\Delta}_i) \widetilde{S}(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) dA \tag{8.9}$$

where

$$\begin{aligned} \widetilde{S}(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) = \widetilde{f}_r(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) + \ldots \\ \widetilde{f}_t(\vec{x}, \vec{\Delta}_i) \cdot \widetilde{g}_r(\vec{x} + \vec{\Delta}_i, \vec{\Delta}_i, \vec{\Delta}_r) \ldots \\ \cdot \widetilde{f}_t(\vec{x} + \vec{\Delta}_i + \vec{\Delta_r}, \vec{\Delta}_r). \end{aligned} \tag{8.10}$$

If $\widetilde{S}$ varies spatially with period 1 in the x and y directions, then the average exitant radiance is

$$\bar{l}_r(\vec{\Delta}_r) = \int_{\vec{x} \in A} \int_{\vec{\Delta}_i \in A} l_i(\vec{\Delta}_i) \widetilde{S}(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) d\vec{\Delta}_i \, d\vec{x} \tag{8.11}$$

where $A = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$.

An Ideal Case

Suppose we are free to choose any form for $\widetilde{f_r}$, $\widetilde{f_t}$, and $\widetilde{g}_r$ so long as it varies in x and y with period 1. To simplify things, let $\widetilde{f_r} = 0$, $\vec{\Delta}_r = \vec{0}$ and suppose $\widetilde{g}_r(\vec{x}, \vec{\Delta}_i, \vec{\Delta}_r) = \widetilde{g}_r(\vec{x}, \vec{\Delta}_r)$ and $\widetilde{f}_t(\vec{x}, \vec{\Delta}_i) = \widetilde{f}_t(\vec{x})$ do not depend on $\vec{\Delta}_i$. Then the average exitant radiance is

$$\begin{aligned} l_r(\vec{0}) = \int_{\vec{x} \in A} \widetilde{f}_t(\vec{x}) \widetilde{g}_r(\vec{x}, \vec{0}) \ldots \\ \cdot \int_{\vec{\Delta}_i \in A} l_i(\vec{\Delta}_i) \widetilde{f}_t(\vec{x} - \vec{\Delta}_i) d\vec{\Delta}_i \, d\vec{x} \end{aligned} \tag{8.12}$$

Recalling our goal of recovering l_r, if we choose $\widetilde{f}_t(\vec{x}) = \delta(\vec{x}) + \delta(\vec{x} - \vec{x}')$ and $\widetilde{g}_r(\vec{x}) = \delta(\vec{x} - \vec{x}')$, where δ is the Kronecker delta function, then we get

$$l_r(\vec{0}) = l_i(\vec{x}') + l_i(\vec{0}). \tag{8.13}$$

Thus, setting $\widetilde{f}_t$ and $\widetilde{g}_r$ to delta functions enables point sampling of the incident radiance. Assuming we could measure the effects of a delta function, this would allow full

recovery of the lighting (or a sampled version if $\vec{x}'$ is restricted to a finite set of values). However, this is an unrealistic solution in practice because delta-like functions would result in very subtle changes in image intensity which would be hard to recover with an image sensor.

Another possible choice for $\widetilde{f}_t$ is

$$\begin{aligned}\widetilde{f}_t(\vec{x}, \vec{\Delta}_i) &= e^{-i2\pi\vec{n}\cdot\vec{x}} \\ &= e^{-i2\pi ux - i2\pi vy}\end{aligned} \tag{8.14}$$

where $\vec{n} = \{u, v\}$ and u,v are integers. In this case the average exitant radiance is

$$\begin{aligned}l_r(\vec{0}) = \int_{\vec{x}\in A} e^{-i2\pi\vec{n}\cdot\vec{x}}\widetilde{g}_r(\vec{x}, \vec{0}) \ldots \\ \cdot \int_{\vec{\Delta}_i\in A} l_i(\vec{\Delta}_i)e^{-i2\pi\vec{n}\cdot(\vec{x}-\vec{\Delta}_i)}d\vec{\Delta}_i\, d\vec{x} \\ l_r(\vec{0}) = \int_{\vec{x}\in A} e^{-i2\pi 2\vec{n}\cdot\vec{x}}\widetilde{g}_r(\vec{x}, \vec{0}) \ldots \\ \cdot \int_{\vec{\Delta}_i\in A} l_i(\vec{\Delta}_i)e^{i2\pi\vec{n}\cdot\vec{\Delta}_i}d\vec{\Delta}_i\, d\vec{x} \\ l_r(\vec{0}) = \widetilde{G}_{2\vec{n}}L^*_{\vec{n}}\end{aligned} \tag{8.15}$$

where $L_{\vec{n}} = L(u, v)$ is the u,vth 2D Fourier series coefficient of the lighting, $\widetilde{G}_{2\vec{n}} = \widetilde{G}(2u, 2v)$ is the $(2u, 2v)$th Fourier series coefficient of $\widetilde{g}_r$, and $*$ denotes the conjugate operator. To recover $L_{\vec{n}}$ we simply need to choose $\widetilde{g}_r$ so that it contains frequencies of $2\vec{n}$: The most logical choice is to set $\widetilde{g}_r(\vec{x}, \vec{0}) = e^{i4\pi\vec{n}\cdot\vec{x}}$, yielding $G_{2n} = 1$ and

$$l_r(\vec{0}) = L^*_{\vec{n}}. \tag{8.16}$$

From this equation we directly obtain $L_{\vec{n}}$. While the assumptions used to reach this result are unrealistic[1], it does provide hope that one can construct a BRDF that directly outputs frequency components of the lighting. Since low frequency lighting is often sufficient for rendering purposes, we should be able to obtain a useful lighting representation using only a small set of such BRDFs.

[1]Not only is positivity violated, but imaginary numbers are used!

A More Realistic Case

To satisfy the laws of physics, we must modify the above formulation in a number of ways:

1. Positivity and conservation of energy must be enforced. A valid BRDF or BTDF is greater than or equal to zero for all inputs and the integral of a BRDF or BTDF over all incident directions must sum to 1 or less.

2. Fresnel reflectance varies with incident angle, so our assumption that the top and bottom layers are constant across incident angles is violated.

3. We must depend exclusively on non-imaginary numbers.

Assumption 1 can be met simply by adding a constant term to $\widetilde{f_t}$ or $\widetilde{g_r}$ and then scaling the signal with a multiplicative factor. Thus, we will get a new signal $k_m(k_a + f(\vec{x}))$ that satisfies positivity and conservation of energy. Assumption 2 implies that we can no longer factor $\widetilde{f_t}$ and $\widetilde{g_r}$ out of the inner integral. However, in many cases we can factor these terms into a spatially varying component that doesn't vary with incident angle and a non-spatially varying component that remains inside the integral. If the spatially varying components have appropriate signals we can recover the product of the lighting with the non-spatially varying components of $\widetilde{f_t}$ and $\widetilde{g_r}$. Once this is recovered we can divide out the undesired terms and recover the original lighting. Assumption 3 is easily overcome by using sinusoids instead of complex exponentials (*i.e.*, the real-valued form of the Fourier series).

8.3 Experimental Validation

8.3.1 Setup

To validate our theory, we printed a set of sinusoidal patterns on a transparency sheet and on a sheet of matte paper and separated the two patterns with a sheet of glass

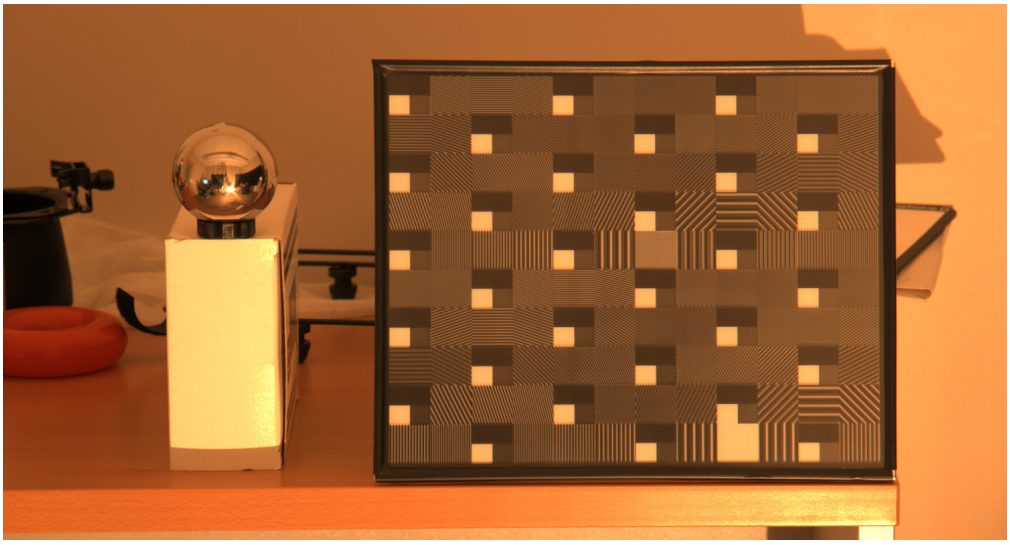


Figure 8.4: Experimental setup. We have our planar light probe next to a mirrored ball, which is used to capture the baseline lighting.

(see figure 8.4). The thickness of the glass was measured to be 0.096 inches and the refractive index assumed to be 1.52. To flatten the transparency sheet we also placed a sheet of glass above the transparency. Thus, our light probe consists of two sheets of glass, a transparency sheet and a piece of matte paper. For each frequency[2] (u, v) we devote two regions where the top layer is a sinusoid of the form $\frac{1}{2}(1+\sin(-2\pi(ux+vy)))$ and the bottom is a sinusoid of the form $\frac{1}{2}(1+\sin(4\pi(ux+vy)+\tau))$. If we assume the bottom layer is Lambertian then it can be shown that the reflected radiance is of the form $aL_0 + bL_{a\vec{n}} + cL_{b\vec{n}} + s(x)$, where L_0 is the average incident radiance, $L_{a\vec{n}}$ and $L_{b\vec{n}}$ are the the even and odd portions of the Fourier coefficients, and $s(x)$ represents the specularities that occur at the surface of the glass. We add a spatial dependence on the surface reflection because while we assume the light and camera are distant, in practice this assumption is violated and surface reflections vary spatially across the surface (albeit slowly). To counteract the effect of the spatially varying specularity term, we sample the specular reflection by placing unpatterned regions at uniform intervals across the light probe. There are four types of unpatterned region: (top clear, bottom

[2]Minus redundant frequencies caused by conjugate symmetry.

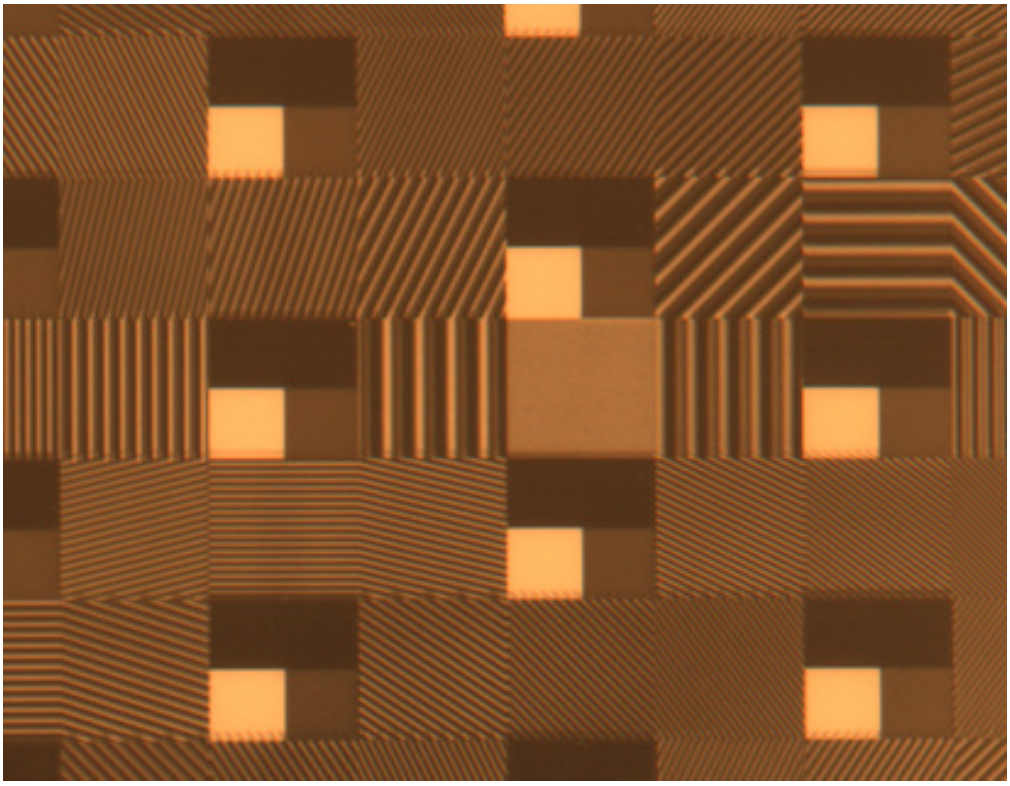


Figure 8.5: Closeup of the patterns on our planar light probe.

white), (top clear, bottom dark), (top dark, bottom white), (top dark, bottom dark). Because Fresnel reflection occurs at the top surface of the glass, by subtracting the average intensity in a constant region from some other type of region, we effectively cancel the specular reflection term.

We layout the planar light probe in terms of blocks, where each block consists of a pattern designed to extract a single frequency component of the incident lighting. To calibrate the light probe, suppose we have K blocks and N images of the probe under different known lighting conditions. Then we can form a feature vector $V \in [KxN]$ of our observations and a corresponding linear system $L = MV$ where L contains the known lighting coefficients. After solving for the matrix M we can then obtain unknown lighting coefficients from images of our light probe using MV.

8.3.2 Results

Figures 8.6,8.7, and 8.8 show some results from our light probe. For very low frequencies the results seem relatively stable; however, the stability quickly declines for higher frequencies[3]. There are many potential causes for this. One likely possibility is that high frequency terms are present in many of the regions we observe as a result of discretization in the printing process. Since we are trying to fit only low order terms, if higher frequencies are present it will interfere with the fitting process. A potential solution would be to use a better printing process such as dye-sublimination[4] to get more continuous half-tones. Another possible workaround would be to directly measure the BRDFs present on the light probe and use these directly as basis functions for the lighting. This is the direction we are currently working towards as it should yield the best results; however it requires tedious and error prone BRDF measurements that can be difficult to get right.

8.4 Summary

We have shown theory that suggests BRDFs can be manufactured that are sensitive to specific frequencies of the lighting. Based on this theory, we have constructed a planar light probe capable of estimating low frequencies of the lighting. Such a probe would be useful for many applications, particularly applications requiring lighting estimation from low-dynamic range images. For most materials, a fifth order frequency approximation of the lighting is enough to photo-realistically render it, so if we can push our probe a little further it will be a truly useful device.

[3]Note: In these figures, redundant conjugate symmetric coefficients have been omitted.
[4]A 1200dpi laser printer was used in our experiments.

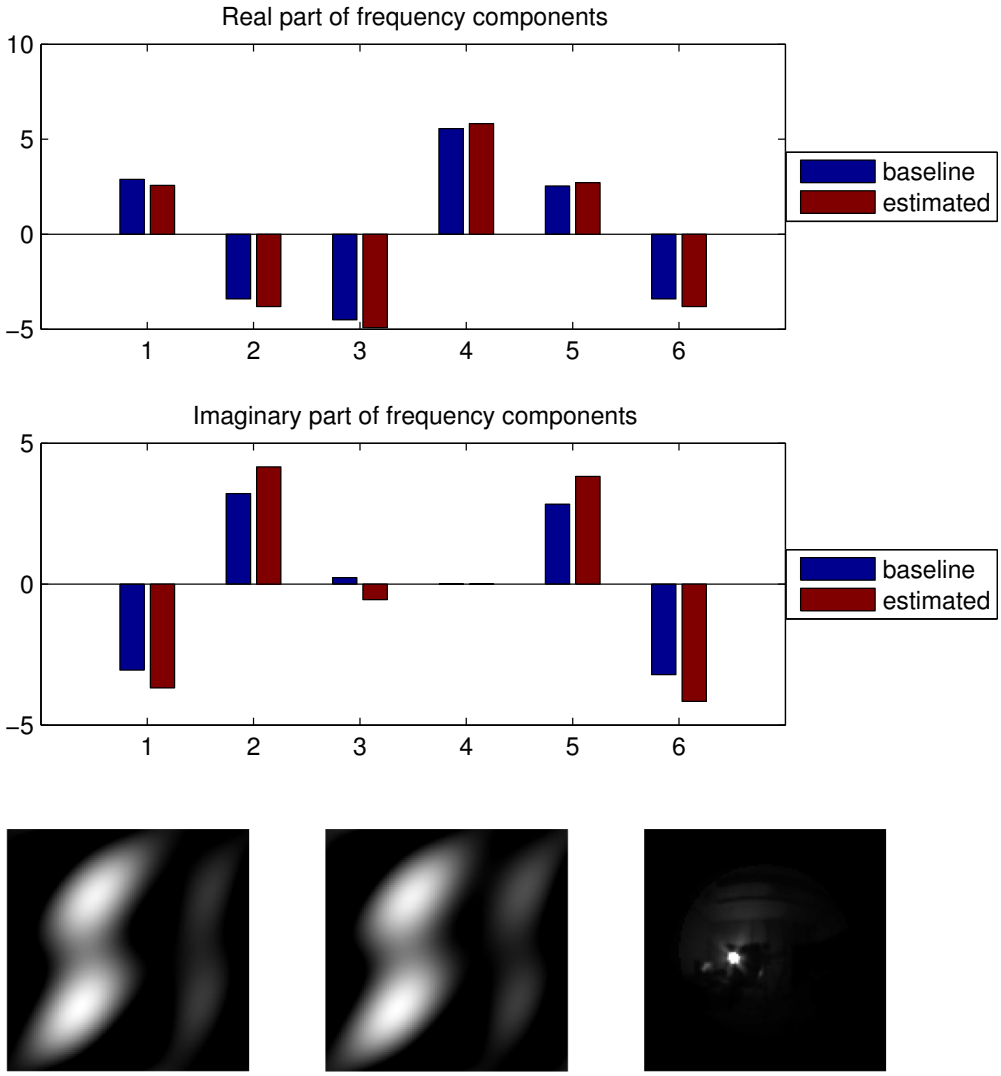


Figure 8.6: Low order frequencies recovered from our light probe vs. the baseline provided by a mirrored ball. (Top) Real and imaginary estimated frequency components compared to the baseline frequencies. (Bottom,Right) Actual lighting, (Bottom,Middle) Actual lighting resulting from an order 1 Fourier series approximation, (Bottom,Left) Estimated lighting.

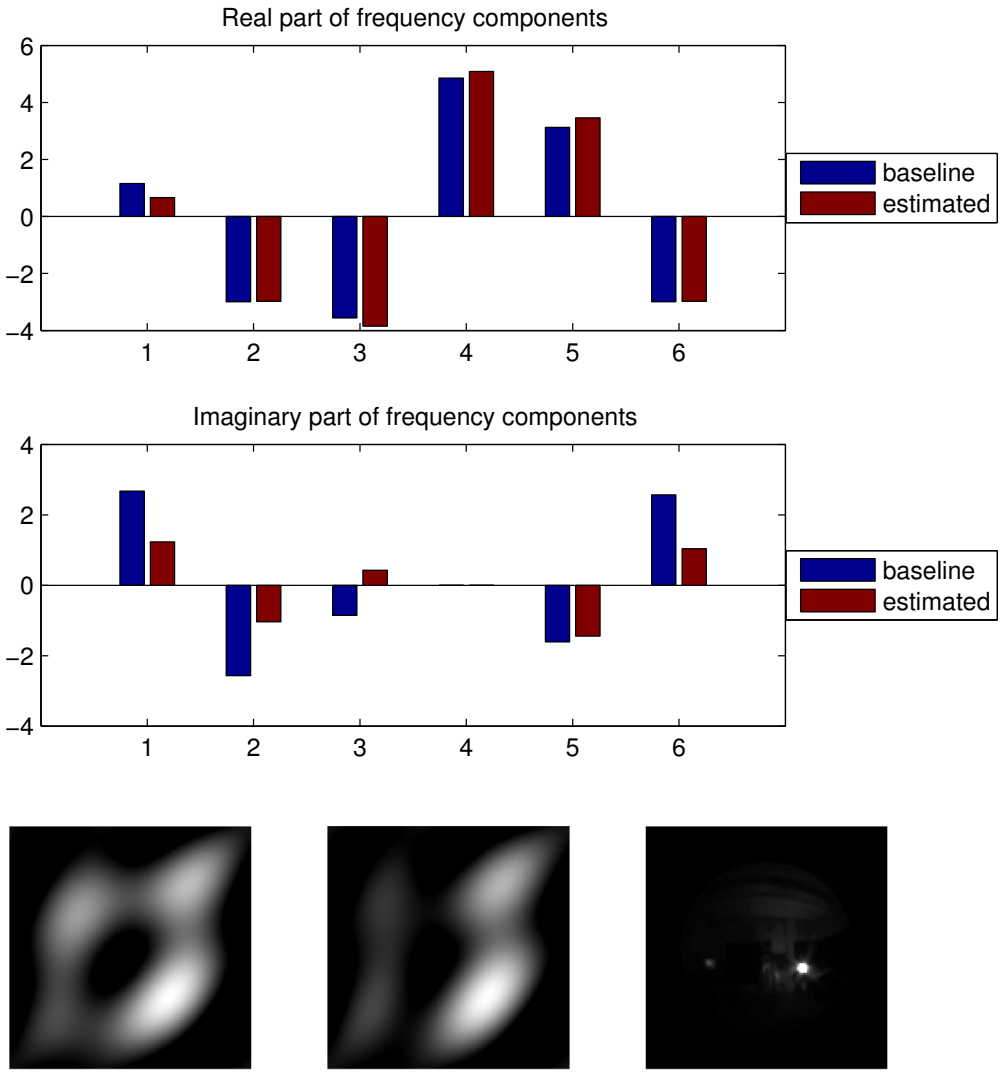


Figure 8.7: Low order frequencies for a different environment recovered from our light probe vs. the baseline provided by a mirrored ball. (Top) Real and imaginary estimated frequency components compared to the baseline frequencies. (Bottom,Right) Actual lighting, (Bottom,Middle) Actual lighting resulting from an order 1 Fourier series approximation, (Bottom,Left) Estimated lighting.

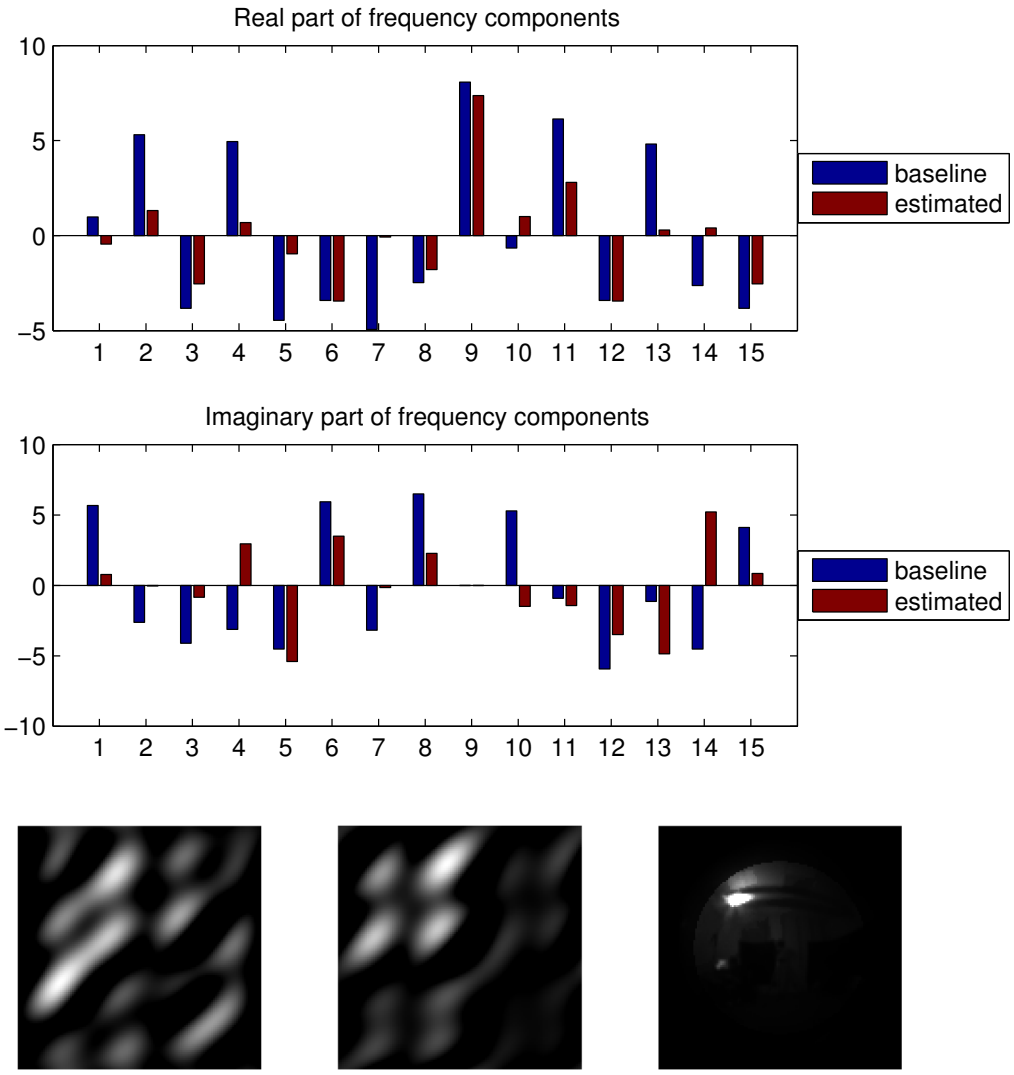


Figure 8.8: Higher frequency approximations. Current estimation techniques are only working well for very low frequencies. Notice all the noise in the estimation; we hope to fix this in future work.

Chapter 9

Conclusion

In this dissertation, we have examined three important aspects of scene reconstruction: recovery of shape, reflectance, and illumination. Each component plays an essential role in the formation of images. In each chapter a different element of the reconstruction problem is addressed.

In Chapter 4 we presented a method for resolving the GBR ambiguity in uncalibrated photometric stereo by utilizing the spatial distribution of surface reflectance. In Chapters 5 and 6 we presented photometric stereo methods which are valid for nearly arbitrary surface reflectance; specifically, we correctly handle spatially varying reflectance and do not impose parametric forms on the BRDF. In Chapter 7 we showed how to utilize similar constraints for use in multi-view reconstruction. Finally, in Chapter 8 we introduced a novel method for measuring the illumination in a scene using a light probe with spatially varying reflectance.

The underlying theme of this research has been to extend existing methods to handle and utilize complex reflectance. This is an important issue because most of the models used in computer vision are incapable of handling the diversity of materials found in the real world. When the underlying models are insufficient, this leads to incorrect and biased inference of scene properties. On the other hand, flexible models can be problematic from a data fitting perspective. For this reason, we have striven to

utilize properties of reflectance which are both physically valid for a wide range of materials and constraining enough for inference. Identifying such properties and determining appropriate methods for utilizing them is non-trivial and there are often trade-offs involved. For example, the method presented in Chapter 5 is valid for objects with arbitrary isotropic reflectance, but requires a specific set of input images to work correctly. In shape from shading, where only a single input image is provided, reconstruction is extremely difficult even with very restrictive models.

It is our hope that the research presented in this dissertation will inspire future methods that correctly operate on more general scenes and / or utilize less restrictive imaging setups. In this respect, there remain a number of open problems. For example, a valid criticism of the techniques in Chapters 5 and 6 is that a large number of images are required. This is mostly due to the flexibility of the underlying reflectance models; however, it is not clear how much flexibility is required in practice. A promising approach may be to learn reflectance models from material databases such as that of Matusik et al. (2003). Using a suitable database it should be theoretically possible to construct reflectance models that are optimally flexible for a given application.

Another challenge is how to properly handle global illumination effects, which include phenomenon such as cast shadows, interreflections, and subsurface scattering. These effects are particularly difficult because the shape and reflectance at each point in the scene is potentially affected by every other point in the scene. Apart from a few notable exceptions (Kender and Smith, 1987; Nayar et al., 1991; Chandraker et al., 2005, 2007), most state of the art methods including those presented in this dissertation, either treat global illumination effects as outliers or simply ignore them as negligible. These are dangerous assumptions to make, however, as global illumination often has a significant influence on the appearance of objects. To properly solve this problem will require new analysis techniques capable of handling the complex interdependencies of each point in the scene.

Reconstruction of scenes viewed under complex illumination conditions is another difficult problem. In most of this dissertation, with the exception of Chapter 8,

illumination was assumed to arise from a single directional light source. However, it is very uncommon to encounter this type of illumination in the real-world, where objects typically receive light from all directions. A promising way to handle such illumination is to use frequency domain representations of the illumination and reflectance as in Ramamoorthi (2002) and Basri and Jacobs (2003); however, these approaches have difficulty handling high-frequencies in either the reflectance or the illumination. Another issue is spatially varying illumination, which means the light arriving at one point may be different from the light arriving at another point. If the spatial variation is known, it can be utilized to constrain the shape of the scene; this is the basis for techniques such as shape from structured illumination. How to properly handle unknown spatial variation is an open problem.

A more specific direction for future work is to further analyze the shape ambiguity that arises in Chapter 5. As shown in that chapter, the surface can be recovered by photometric stereo up to a set of iso-depth contours by imposing bilateral symmetry and surface integrability. At least two questions immediately arise: (a) What additional constraints can be used to recover the full surface structure? and (b) In what applications can the isocontour structure be directly utilized? The technique in Chapter 6 presents one possible way to address the former question, but there are numerous other ways to fully constrain the surface shape that remain to be explored. Potential applications, including recognition, industrial parts inspection, topological analysis, *etc*. are completely unexplored.

Another area of future work is to apply bilateral symmetry, isotropy, and the bi-variate BRDF approximation of Chapter 6 to other problems in computer vision. These properties are valid for a wide range of real-world materials making them ideal constraints for scenes with complex reflectance. The main problem in utilizing these properties is that they require samples from specific regions of the BRDF domain to be useful. This means that incorporating such constraints in existing algorithms will often be non-trivial. To overcome these difficulties, it may be necessary to acquire images from multiple lighting conditions and / or viewpoints, or to formulate constraints that

link multiple surface points together.

References

Agrawal, A., Chellappa, R., and Raskar, R., 2005: An algebraic approach to surface reconstruction from gradient fields. In *ICCV '05*, 174–181. IEEE Computer Society, Washington, DC, USA.

Agrawal, A. K., Raskar, R., and Chellappa, R., 2006: What is the range of surface reconstructions from a gradient field? In *ECCV '06*, 578–591. Springer.

Alldrin, N. G., and Kriegman, D. J., 2006: A planar light probe. In *CVPR '06*, 2324–2330. IEEE Computer Society, Washington, DC, USA.

Alldrin, N. G., and Kriegman, D. J., 2007: Toward reconstructing surfaces with arbitrary isotropic reflectance : A stratified photometric stereo approach. In *ICCV '07*. IEEE Computer Society, Washington, DC, USA.

Alldrin, N. G., Mallick, S. P., and Kriegman, D. J., 2007: Resolving the generalized bas-relief ambiguity by entropy minimization. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Alldrin, N. G., Zickler, T., and Kriegman, D. J., 2008: Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR '08*. IEEE Computer Society, Washington, DC, USA.

ARToolkit, 2006: A software library for building Augmented Reality (AR) applications. http://www.hitl.washington.edu/artoolkit/.

Astor, M., 2008: Image of the original Utah teapot at the Computer History Museum, Boston. http://www.flickr.com.

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

Basri, R., and Jacobs, D. W., 2003: Lambertian reflectance and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **25**(2), 218–233.

Belhumeur, P., Kriegman, D., and Yuille, A., 1999: The bas-relief ambiguity. *International Journal of Computer Vision (IJCV)*, **35**(1), 33–44.

Bichsel, M., and Pentland, A., 1992: A simple algorithm for shape from shading. In *CVPR '92*, volume 92, 459–465. IEEE Computer Society, Washington, DC, USA.

Blanz, V., and Vetter, T., 2003: Face recognition based on fitting a 3D morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **25**(9), 1063–1074. ISSN 0162-8828.

Blinn, J. F., and Newell, M. E., 1976: Texture and reflection in computer generated images. *Communications of the ACM*, **19**(10), 542–547.

Bouguet, J.-Y., 2008: Camera calibration toolbox for Matlab. http://www.vision.caltech.edu/bouguetj.

Boykov, Y., and Kolmogorov, V., 2004: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **26**(9), 1124–1137.

Boykov, Y., Veksler, O., and Zabih, R., 2001: Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **23**(11), 1222–1239.

Brand, M., 2002: Incremental singular value decomposition of uncertain data with missing values. In *ECCV '02*, 707–720. Springer.

Chabert, C.-F., Einarsson, P., Jones, A., Lamond, B., Ma, W.-C., Sylwan, S., Hawkins, T., and Debevec, P., 2006: Relighting human locomotion with flowed reflectance fields. In *SIGGRAPH '06 (Sketches)*, 76.

Chandraker, M., Agarwal, S., and Kriegman, D., 2007: Shadowcuts: Photometric stereo with shadows. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Chandraker, M., Kahl, F., and Kriegman, D., 2005: Reflections on the generalized bas-relief ambiguity. In *CVPR '05*. IEEE Computer Society, Washington, DC, USA.

Coleman, J. E., and Jain, R., 1982: Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *CGIP*, **18**(4), 309–328.

CVX, 2008: Matlab software for disciplined convex programming. http://www.stanford.edu/~boyd/cvx/.

Debevec, P., 1998: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH '98*, 189–198.

Debevec, P., 2004a: The Parthenon. In *SIGGRAPH '04 (Computer Animation Festival)*, 188.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M., 2000: Acquiring the reflectance field of a human face. In *SIGGRAPH '00*, 145–156.

Debevec, P., Tchou, C., Gardner, A., Hawkins, T., Poullis, C., Stumpfel, J., Jones, A., Yun, N., Einarsson, P., Lundgren, T., and Fajardo, M., 2004: Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical Report ICT-TR-06.2004, University of Southern California ICT.

Debevec, P. E., 2004b: Digitizing the Parthenon: Estimating surface reflectance properties of a complex scene under captured natural illumination. In *VMV*, 99.

Debevec, P. E., and Malik, J., 1997: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97*, 369–378.

Drbohlav, O., 2003: *Towards Uncalibrated Photometric Stereo for Non-Lambertian Surfaces*. Ph.D. thesis, Czech Technical University.

Drbohlav, O., and Sara, R., 2002: Specularities reduce ambiguity of uncalibrated photometric stereo. In *ECCV '02*, 46–60. Springer.

Ecker, A., Kutulakos, K., and Jepson, A., 2007: Shape from planar curves: A linear escape from flatland. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Epstein, R., Hallinan, P., and Yuille, A., 1995: 5+/-2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *PBMCV*.

Finlayson, G., Drew, M., and Lu, C., 2004: Intrinsic images by entropy minimization. In *ECCV '04*, 582–595. Springer.

Fitzgibbon, A. W., Cross, G., and Zisserman, A., 1998: Automatic 3D model construction for turn-table sequences. In *SMILE '98*, 155–170.

Foo, S.-C., 1996: *A Gonioreflectometer for Measuring the Bidirectional Reflectance of Material for use in Illumination Computation*. Master's thesis, Cornell University.

Forsyth, D., and Ponce, J., 2003: *Computer Vision: A Modern Approach*. Prentice Hall.

Georghiades, A. S., 2003: Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *ICCV '03*. IEEE Computer Society, Washington, DC, USA.

Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J., 2001: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **23**(6), 643–660.

Glassner, A. S., 1994: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 1558602763.

Goldman, D. B., Curless, B., Hertzmann, A., and Seitz, S. M., 2005: Shape and spatially-varying BRDFs from photometric stereo. In *ICCV '05*. IEEE Computer Society, Washington, DC, USA.

Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F., 1996: The lumigraph. In *SIGGRAPH '96*, 43–54.

Greene, N., 1986: Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications*, **6**(11), 21–29. ISSN 0272-1716.

Haeberli, P., and Segal, M., 1993: Texture mapping as a fundamental drawing primitive. In *EGSR '93*, 259–266. Eurographics Association, Grenoble, France.

Hayakawa, H., 1994: Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America (JOSA) A*, **11**(11), 3079–3089.

Hernandez, C., Schmitt, F., and Cipolla, R., 2007: Silhouette coherence for camera calibration under circular motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **29**(2), 343–349.

Hernandez, C. E., Vogiatzis, G., and Cipolla, R., 2008: Multi-view photometric stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **30**(3), 548–554.

Hertzmann, A., and Seitz, S. M., 2003: Shape and materials by example: A photometric stereo approach. In *CVPR '03*. IEEE Computer Society, Washington, DC, USA.

Hertzmann, A., and Seitz, S. M., 2005: Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **27**(8), 1254–1264.

Hirschmüller, H., and Scharstein, D., 2007: Evaluation of cost functions for stereo matching. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Horn, B., 1970: *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. Ph.D. thesis, MIT.

Horn, B. K. P., 1986: *Robot Vision*. The MIT Press. ISBN 0262081598.

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

Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M., and Okamoto, Y., 2007: The great Buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *International Journal of Computer Vision (IJCV)*, **75**(1), 189–208.

Joshi, N., and Kriegman, D. J., 2007: Shape from varying illumination and viewpoint. In *ICCV '07*. IEEE Computer Society, Washington, DC, USA.

Ke, Y., Tang, X., and Jing, F., 2006: The design of high-level features for photo quality assessment. In *CVPR '06*, 419–426. IEEE Computer Society, Washington, DC, USA.

Kender, J., and Smith, E., 1987: Shape from darkness. In *ICCV '87*, 539–546. IEEE Computer Society, Washington, DC, USA.

Kim, C., Petrov, A. P., Choh, H., Seo, Y., and Kweon, I., 1998: Illuminant direction and shape of a bump. *Journal of the Optical Society of America (JOSA) A*, **15**, 2341–2350.

Koenderink, J. J., Doorn, A. J. V., Dana, K. J., and Nayar, S., 1999: Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision (IJCV)*, **31**(2-3), 129–144.

Kolmogorov, V., and Zabin, R., 2004: What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **26**(2), 147–159.

Kriegman, D., and Belhumeur, P., 1998: What shadows reveal about object structure. In *ECCV '98*, volume 2, 399. Springer.

Kriegman, D., and Belhumeur, P., 2001: What shadows reveal about object structure. *Journal of the Optical Society of America (JOSA) A*, **18**(8), 1804–1813.

Kutulakos, K. N., and Seitz, S. M., 2000: A theory of shape by space carving. *International Journal of Computer Vision (IJCV)*, **38**(3), 199–218.

Laurentini, A., 1994: The visual hull concept for silhouette-based image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **16**(2), 150–162.

Lawrence, J., Ben-Artzi, A., DeCoro, C., Matusik, W., Pfister, H., Ramamoorthi, R., and Rusinkiewicz, S., 2006: Inverse shade trees for non-parametric material representation and editing. In *SIGGRAPH '06*.

Lee, K., Ho, J., and Kriegman, D., 2005: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 684–698.

Lee, K., and Kuo, C., 1993: Shape from shading with a linear triangular element surface model. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **15**(8), 815–822.

Lensch, H. P. A., Kautz, J., Goesele, M., Heidrich, W., and Seidel, H.-P., 2001: Image-based reconstruction of spatially varying materials. In *EGSR '01*. Eurographics Association, Grenoble, France.

Levoy, M., 2008: Image of the original Stanford bunny. http://www.cc.gatech.edu/~turk/bunny/bunny.html.

Levoy, M., and Hanrahan, P., 1996: Light field rendering. In *SIGGRAPH '96*, 31–42.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., and Fulk, D., 2000: The digital Michelangelo project: 3D scanning of large statues. In *SIGGRAPH '00*, 131–144.

Lim, J., Ho, J., Yang, M.-H., and Kriegman, D., 2005: Passive photometric stereo from motion. In *ICCV '05*, 1635–1642. IEEE Computer Society, Washington, DC, USA.

Lu, J., and Little, J. J., 1999: Reflectance and shape from images using a collinear light source. *International Journal of Computer Vision (IJCV)*, **32**(3), 213–240.

Ma, Y., Soatto, S., Kosecka, J., and Sastry, S., 2005: *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer-Verlag.

Mallick, S. P., Zickler, T., Kriegman, D. J., and Belhumeur, P. N., 2005: Beyond Lambert: Reconstructing specular surfaces using color. In *CVPR '05*. IEEE Computer Society, Washington, DC, USA.

Marschner, S. R., 1998: *Inverse Rendering for Computer Graphics*. Ph.D. thesis, Cornell.

Marschner, S. R., and Greenberg, D. P., 1997: Inverse lighting for photography. In *CIP '97*.

Matusik, W., Pfister, H., Brand, M., and McMillan, L., 2003: A data-driven reflectance model. *ACM Transactions on Graphics*, **22**(3), 759–769.

Miyazaki, D., Tan, R. T., Hara, K., and Ikeuchi, K., 2003: Polarization-based inverse rendering from a single view. In *ICCV '03*, 982–987. IEEE Computer Society, Washington, DC, USA.

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

Nayar, S., and Mitsunaga, T., 2000: High dynamic range imaging: Spatially varying pixel exposures. In *CVPR '00*, 472–479. IEEE Computer Society, Washington, DC, USA.

Nayar, S. K., 1997: Catadioptric omnidirectional camera. In *CVPR '97*, 482. IEEE Computer Society, Washington, DC, USA.

Nayar, S. K., Ikeuchi, K., and Kanade, T., 1991: Shape from interreflections. *International Journal of Computer Vision (IJCV)*, **6**(3), 173–195.

Ngan, A., Durand, F., and Matusik, W., 2005: Experimental analysis of BRDF models. *EGSR '05*.

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T., 1922: *Geometrical considerations and nomenclature for reflectance*. Jones and Bartlett Publishers, Inc., USA. ISBN 0-86720-294-7.

Okabe, T., Sato, I., and Sato, Y., 2004: Spherical harmonics vs. Haar wavelets: Basis for recovering illumination from cast shadows. In *CVPR '04*, volume 1, 50–57. IEEE Computer Society, Washington, DC, USA.

Oren, M., and Nayar, S. K., 1994: Generalization of Lambert's reflectance model. In *SIGGRAPH '94*, 239–246.

Palubinskas, G., Descombes, X., and Kruggel, F., 1998: An unsupervised clustering method using the entropy minimization. In *International Conference on Pattern Recognition (ICPR)*, volume 2, 1816–1818.

Paninski, L., 2003: Estimation of entropy and mutual information. *Neural Computation*, **15**(6), 1191–1253.

PBRT Renderer, 2008: http://www.pbrt.org.

Pentland, A., 1982: Finding the illuminant direction. *Journal of the Optical Society of America (JOSA) A*, **72**, 448–455.

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

POV-Ray, 2008: The persistence of vision raytracer. http://www.povray.org.

Ramamoorthi, R., 2002: *A signal-processing framework for forward and inverse rendering*. Ph.D. thesis, Stanford. Adviser-Pat Hanrahan.

Ramamoorthi, R., Koudelka, M., and Belhumeur, P., 2005: A fourier theory for cast shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **27**(2), 288–295.

Rohatgi, V. K., 1976: *"An Introduction to Probability Theory and Mathematical Statistics"*. John Wiley & Sons Inc., New York.

Rusinkiewicz, S., 1998: A new change of variables for efficient BRDF representation. In *EGSR '98*. Eurographics Association, Grenoble, France.

Samir, C., Srivastava, A., and Daoudi, M., 2006: Three-dimensional face recognition using shapes of facial curves. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **28**(11), 1858–1862.

Sato, I., Sato, Y., and Ikeuchi, K., 1999: Illumination distribution from brightness in shadows: adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *ICCV '99*, volume 2, 875–882. IEEE Computer Society, Washington, DC, USA.

Sato, I., Sato, Y., and Ikeuchi, K., 2001: *Illumination distribution from shadows*. Kluwer Academic Publishers. ISBN 0-7923-7515-7.

Sato, I., Sato, Y., and Ikeuchi, K., 2003: Illumination from shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **25**(3), 290–300.

Sato, Y., and Ikeuchi, K., 1994: Temporal-color space analysis of reflection. *Journal of the Optical Society of America (JOSA) A*, **11**(11), 2990–3002.

Savarese, S., Andreetto, M., Rushmeier, H., Bernardini, F., and Perona, P., 2007: 3D reconstruction by shadow carving: Theory and practical evaluation. *International Journal of Computer Vision (IJCV)*, **71**(3), 305–336.

Scharstein, D., and Pal, C., 2007: Learning conditional random fields for stereo. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Schlüns, K., and Wittig, O., 1993: Photometric stereo for non-lambertian surfaces using color information. In *International Conference on Image Analysis and Processing (ICIAP)*.

SDPT3, 2008: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html.

Shafer, S., and Kanade, T., 1983: Using shadows in finding surface orientations. *CVGIP*, **22**, 145–176.

Silver, W. M., 1980: *Determining Shape and Reflectance Using Multiple Images*. Master's thesis, MIT.

Stark, M. M., Arvo, J., and Smits, B., 2005: Barycentric parameterizations for isotropic BRDFs. *IEEE Transactions on Visualization and Computer Graphics*, **11**(2), 126–138.

Stumpfel, J., Tchou, C., Yun, N., Martinez, P., Hawkins, T., Jones, A., Emerson, B., and Debevec, P., 2003: Digital reunification of the Parthenon and its sculptures. In *VAST '03*, 41–50.

Tan, P., Mallick, S. P., Quan, L., Kriegman, D. J., and Zickler, T., 2007: Isotropy, reciprocity and the generalized bas-relief ambiguity. In *CVPR '07*. IEEE Computer Society, Washington, DC, USA.

Torrance, K. E., and Sparrow, E. M., 1967: Theory for off-specular reflection from rough surfaces. *Journal of the Optical Society of America (JOSA) A*, **57**(9), 1105–1114.

Trucco, E., and Verri, A., 1998: *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.

UCSD FWGrid Project, 2008: http://fwgrid.ucsd.edu.

USC High-Resolution Light Probe Image Gallery, 2008: http://gl.ict.usc.edu/Data/HighResProbes/.

Wang, Y., and Samaras, D., 2002: Estimation of multiple illuminants from a single image of arbitrary known geometry. In *ECCV '02*, volume 3, 272–288. Springer.

Weber, M., and Cipolla, R., 2001: A practical method for estimation of point light-sources. In *BMVC '01*.

Weistroffer, R. P., Walcott, K. R., Humphreys, G., and Lawrence, J., 2007: Efficient basis decomposition for scattered reflectance data. In *EGSR '07*. Eurographics Association, Grenoble, France.

Westin, S. H., 2003: Measurement data, Cornell university program of computer graphics. http://www.graphics.cornell.edu/online/measurements/.

Woodham, R., 1980: Photometric method for determining surface orientation from multiple images. *Optical Engineering*, **19**(1), 139–144.

Yuille, A., and Snow, D., 1997: Shape and albedo from multiple images using integrability. In *CVPR '97*, 158–164. IEEE Computer Society, Washington, DC, USA.

Zhang, L., Curless, B., Hertzmann, A., and Seitz, S. M., 2003: Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV '03*, 618. IEEE Computer Society, Washington, DC, USA.

Zhang, R., Tsai, P.-S., Cryer, J. E., and Shah, M., 1999: Shape from shading: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **21**(8), 690–706.

Zhang, Y., and Yang, Y.-H., 2001: Multiple illuminant direction detection with application to image synthesis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, **23**(8), 915–920.

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

Zickler, T., Belhumeur, P. N., and Kriegman, D. J., 2002a: Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. In *ECCV '02*. Springer.

Zickler, T. E., Belhumeur, P. N., and Kriegman, D. J., 2002b: Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision (IJCV)*, **49**(2-3), 215–227.