

Abstract

IMAGE-BASED MODELING WITH COMPLEX REFLECTANCE

Todd Ernest Zickler

2004

The appearance of natural surfaces is determined by the interaction of surface structure with scene illumination, and this appearance can be very complex, especially when the illumination and viewpoint are allowed to vary. Our understanding of visual appearance is benefiting from the accelerating development of inexpensive, high-resolution imaging technology, since these devices provide access to vast amounts of visual data. In order to fully exploit this information, however, we must answer open questions about how to efficiently acquire, represent and apply this data.

This dissertation describes work on two aspects of the general problem of creating appearance models from image data: recovering both the shape and reflectance (i.e., material properties) of a scene. First, *Helmholtz stereopsis* is introduced as a reconstruction process that enables the decoupling of shape and reflectance in images, allowing the accurate recovery of shape for a far greater class of surfaces than is possible using existing techniques. Second, this dissertation presents an image-based method for recovering reflectance when the scene shape is known. By exploiting the fact that reflectance varies slowly from point to point over much of a typical object's surface, this reflectometry method enables the accurate estimation of spatially-varying reflectance from a drastically reduced set of images.

What is common to these two techniques is that they are designed to handle surfaces with complex and arbitrary reflectance, meaning reflectance that is not necessarily well-represented by a pre-chosen, low-dimensional (i.e., parametric) model. By eliminating the need for low-dimensional approximations of reflectance, these techniques seek to provide accurate appearance models for general scenes.

IMAGE-BASED MODELING WITH COMPLEX REFLECTANCE

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Todd Ernest Zickler

Dissertation Director: Prof. Peter N. Belhumeur

December, 2004

©2004 by Todd Ernest Zickler.

All rights reserved.

for Lianne

Acknowledgements

This dissertation would not exist without the guidance and support of my advisor, Peter Belhumeur. Working with Peter has taught me what questions to ask and how to ask them. I have also been fortunate to have close contact with David Kriegman, who in many senses has been my second advisor. I am indebted to both Peter and David for their amazing insight that Helmholtz reciprocity might be exploited for surface reconstruction.

I would like to thank Steve Zucker, whose Vision class at McGill University abruptly and completely pulled me into this field. Shree Nayar also deserves credit for always being willing to share his wisdom, and Ravi Ramamoorthi for his tireless and insightful comments.

I have benefited enormously from conversations with peers at two different institutions. At Yale this included Melissa Koudelka, Athos Georghiades, “Archer” Jie Lin, Patrick Huggins, Ohad Ben-Shahar, and Gang Li. At Columbia University this included Michael Grossberg, Srinivasa Narasimhan, Rahul Swaminathan and Ko Nishino.

Others who contributed directly to this work include Jeffrey Ho, who implemented the integration for Figs. 4.3 and 4.4, and Sebastian Enrique and Aner Ben-Artzi, who created the real-time rendering system that produced the results in Fig. 6.10.

I am forever grateful for having the opportunity to do this research—an opportunity created by the devotion and guidance of my wonderful parents and family and the education they enabled.

Finally, the greatest thanks go to my wife, Lianne, who continues to bring happiness, love and balance to my life.

Contents

1	Introduction	7
2	Background	10
2.1	Bidirectional Reflectance Distribution Function (BRDF)	10
2.2	Defining a suitable scale	22
3	Helmholtz Stereopsis	24
3.1	Related Work	25
3.2	Reciprocal Image Pairs	28
3.3	Helmholtz Stereopsis	32
3.4	Comparison to Existing Methods	32
3.5	A Multinocular Helmholtz Stereo Method	36
3.6	Summary	48
4	Binocular Helmholtz Stereopsis	49
4.1	The Reciprocity Constraint as a PDE	49
4.2	Surface Reconstruction	53
4.3	Summary	60
5	Stratified Helmholtz Stereopsis	61
5.1	An Uncalibrated Reciprocity Constraint	62
5.2	Distant Sources	65
5.3	Affine Reconstruction	68

<i>CONTENTS</i>	3
5.4 Metric Reconstruction	69
5.5 Implementation and Results	71
5.6 Summary	77
6 Reflectance Sharing	78
6.1 Assumptions and Related Work	79
6.2 Notation and BRDF Parameterization	82
6.3 Homogeneous Surfaces	85
6.4 Inhomogeneous Surfaces	90
6.5 Generalized Spatial Variation	94
6.6 Summary	101
7 Conclusion	103
8 Future Work	105
8.1 Helmholtz Stereopsis	105
8.2 Reflectance Sharing	106
A Uncalibrated reciprocity constraint	108

List of Figures

2.1	Structure at a continuum of scales	11
2.2	Geometry of incident and reflected beams for the definition of the BRDF .	12
2.3	BRDF measurement of a planar sample	14
2.4	BRDF parameterizations	21
3.1	Acquisition of a reciprocal pair of images	29
3.2	A reciprocal pair of images	30
3.3	Surface properties required for Lambertian surface reconstruction by typical stereo and Helmholtz stereo techniques	35
3.4	Proposed module for co-location of camera and light source	36
3.5	A wheel for capturing multiple reciprocal pairs with a single camera and a single light source.	38
3.6	Six reciprocal image pairs captured using the wheel in Fig. 3.5	39
3.7	Recovered depth and surface normals for a wax cube with no texture	41
3.8	Recovered depth and surface normals for a specular plastic figure	42
3.9	Recovered depth and surface normals for a rough ceramic figure	44
3.10	Recovered depth and surface normals for a surface with a spatially-varying BRDF	45
3.11	Helmholtz stereo reconstruction of a wax cube with no texture	46
3.12	Helmholtz stereo reconstruction of a surface with spatially-varying BRDF .	47
4.1	Rectified reciprocal pair of a painted, plastic mannequin head	50

4.2	A cyclopean coordinate system for binocular Helmholtz stereopsis	51
4.3	A family of reconstructions for one epipolar line of a specular cylinder	53
4.4	Binocular reconstruction of three cylinders with different material types	54
4.5	Two views of a reconstructed plastic mannequin created using a binocular Helmholtz stereo technique.	58
4.6	Recovered depth map and texture-mapped surface	59
5.1	Reciprocal images from eight camera/source positions	63
5.2	Matrix columns displayed as images	73
5.3	Recovered affine and metric normal fields	74
5.4	Affine and metric reconstructed surfaces	75
5.5	Metric reconstruction, texture-mapped with one input image	75
6.1	Isotropic BRDF parameterization suitable for general interpolation in \mathbb{R}^3	84
6.2	Accuracy of the RBF representation of reflectance as the number of RBF centers is increased	90
6.3	Accuracy of the RBF reflectance representation for a sphere with measured metallic-blue BRDF	91
6.4	Accuracy of the RBF reflectance representation for a sphere with synthesized Oren-Nayar BRDF	92
6.5	Accuracy of the RBF representation of spatially-varying reflectance as the number of RBF centers is increased	95
6.6	Evaluation of reflectance sharing using synthetic images of a hemisphere with spatially-varying reflectance	96
6.7	Another evaluation of reflectance sharing using synthetic images of a hemisphere with spatially-varying reflectance	97
6.8	Input images and shape for estimating the reflectance of a human face	98
6.9	Rendering of human face with reflectance function recovered from four input images	100

6.10 Rendering of human face with reflectance function recovered from a single
input image 101

Chapter 1

Introduction

An emerging paradigm in computer vision is generative model-based vision. According to this paradigm, statistical class- or task-specific appearance models represent high-level knowledge that can be integrated with low-level vision cues to disambiguate the inversion of the image formation process and reliably recover the 3D world from its 2D projections. An obvious example is recent work in face recognition, where models of the human face are used for recognition under widely varying pose and illumination. The success of generative model-based methods increases demand for accurate statistical models, which in turn increases demand for systems that capture the data required to build them.

Image-based modeling (IBM) is the area of computer vision dedicated to recovering accurate appearance models (usually in the form of *shape* and *reflectance*) from one or more images of a scene. Since the primary concern is the accuracy of the model, we usually assume these images can be captured under controlled view and illumination conditions. (It is this assumption that differentiates IBM from many other areas of vision.) In addition to providing the necessary data for generative model-based vision, IBM (and the shape and reflectance models it recovers) finds direct application in visual metrology and visual inspection, and is the basis for many image-based rendering systems in computer graphics.

This dissertation considers the image-based modeling problem in two parts. First, Chapters 3–5 address the problem of recovering shape from image data. An image E provides only indirect information about scene shape through the measurement of scene radiance,

since the radiance emitted from a scene is determined by the interaction of shape, reflectance and illumination:

$$E = g(\textit{shape}, \textit{reflectance}, \textit{illumination}). \quad (1.1)$$

Our goal is to invert the image formation process to recover shape. The majority of existing approaches to image-based shape recovery (e.g., stereo and photometric stereo methods) deal with this coupling by making assumptions about surface reflectance. For example, most stereo systems rely on the brightness-constancy constraint, which assumes that scene radiance is independent of viewpoint. Assumptions such as this necessarily limit the generality and accuracy of existing techniques. In the first part of this dissertation, we present a reconstruction process—termed Helmholtz stereopsis—that enables the decoupling of shape and reflectance in a set of images, enabling accurate image-based shape estimation for surfaces with arbitrary and complex reflectance.

In contrast to conventional stereopsis (reconstruction from images taken from multiple views under fixed illumination) or photometric stereopsis (reconstruction from images with multiple illuminations under fixed viewpoint), Helmholtz stereopsis is the process of recovering shape from *reciprocal image pairs*. The notion of a reciprocal pair of images is introduced in Chapter 3, and an associated constraint equation is derived that provides access to reflectance-independent shape information from these images. Following this, Chapters 3–5 present three different Helmholtz stereo methods that use variations of this constraint for the recovery of shape in the multinocular, binocular, and uncalibrated cases.

The second part of this dissertation (Chapter 6) deals with the task of estimating reflectance from a set of images when the shape of the scene is known. Given the shape and a set of images of the scene under known illumination conditions, each pixel in an image represents a sample of the surface reflectance. Without further assumptions, recovering an accurate description of reflectance requires an input set of images large enough to observe high-frequency radiometric events, such as sharp specular highlights, at each point on the surface. This set consists of a near exhaustive sampling of images of the scene from all viewpoints and lighting directions, which can be tens-of-thousands of images or more. While

existing techniques make this problem tractable using low-dimensional approximations of reflectance, this dissertation presents an alternative approach. By exploiting spatial coherence, we develop *reflectance sharing* as a technique for the estimation of non-parametric reflectance from a sparse set of images.

This dissertation is organized as follows. The first chapter provides some necessary background in radiometry, and discusses the representation of scene structure in terms of shape and reflectance. Chapters 3–5 deal with Helmholtz stereopsis, and Chapter 6 discusses the second problem of recovering reflectance when the scene shape is known.

Chapter 2

Background

The appearance of an object is determined by structure at multiple scales. As an example, Fig. 2.1 shows how the structure of an orange can be stratified from megastructure (“it’s roughly a sphere”) to microstructure (molecular and cellular composition.)

For image-based modeling, it is convenient to divide structure into *shape*, or resolved structure, and *reflectance*, or unresolved structure. In this context, we can view reflectance as a summary of the light-scattering effects of microstructure within a small (planar) surface patch. This chapter discusses some of the many existing representations for reflectance. The basic tools for this discussion are provided by the field of radiometry (the science of measuring electromagnetic radiation), so we begin with a summary of some relevant definitions.

2.1 Bidirectional Reflectance Distribution Function (BRDF)

The development in this section summarizes relevant material from the NBS monograph assembled by Nicodemus et al. [81]. It is assumed that the reader is familiar with radiometric terms such as radiant flux, radiance and irradiance, since this material is now common in standard vision and graphics texts [37, 32].

When a surface is illuminated, the incident radiant flux is partially absorbed and partially reflected over the output directions. Consider a surface patch A irradiated with monochromatic, randomly polarized light from direction (θ_i, ϕ_i) over an element of solid

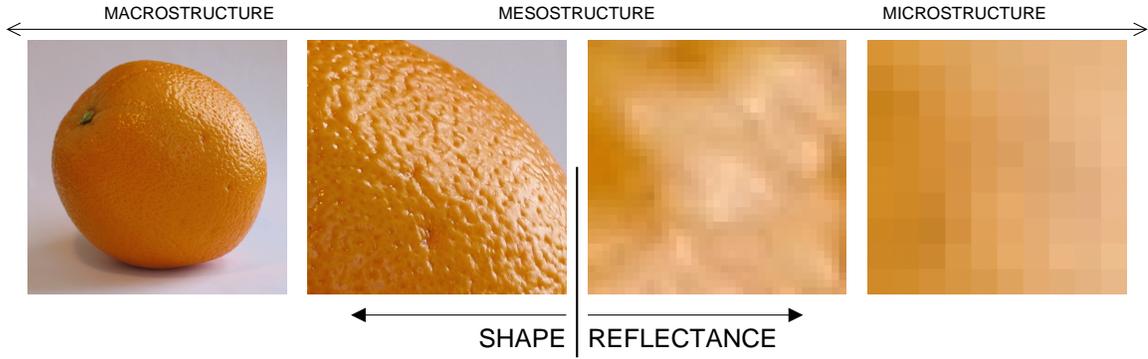


Figure 2.1: Surface structure exists at a continuum of scales. For image-based rendering, it is convenient to divide this structure into resolved structure (shape) and unresolved structure (reflectance).

angle $d\omega_i$, and let $d\Phi_i$ denote the resulting element of incident flux [W] that strikes an element of surface dA centered at position (x_i, y_i) . This incident flux is scattered by the surface, and some is emitted in direction (θ_o, ϕ_o) from another surface point (x_o, y_o) . (See Fig. 2.2.) Letting $dL_o(x_o, y_o, \theta_o, \phi_o)$ denote this reflected radiance [$\text{W}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$], we can write

$$dL_o = S d\Phi_i,$$

which simply says that the reflected radiance is directly proportional to the incident radiant flux. The constant of proportionality is a function of the location of the input and output points of the surface as well as the incident and reflected directions, i.e., $S = S(\theta_i, \phi_i, x_i, y_i; \theta_o, \phi_o, x_o, y_o)$, and its units are [$\text{m}^{-2}\cdot\text{sr}^{-1}$]. The function S (referred to as the bidirectional scattering-surface reflectance distribution function, or BSSRDF) is derived without making any assumptions about the mechanism behind the reflection process, and it provides a very general description of reflectance. The BRDF provides a more tractable expression, and it is derived by making a few assumptions that are valid for a wide range of interesting cases.

Instead of describing the microstructure in detail, as shown in Fig. 2.2, a reference plane is chosen (intersecting the highest point on the surface) to represent the reflecting surface. The normal to this reference plane, along with an arbitrary tangent vector, defines the

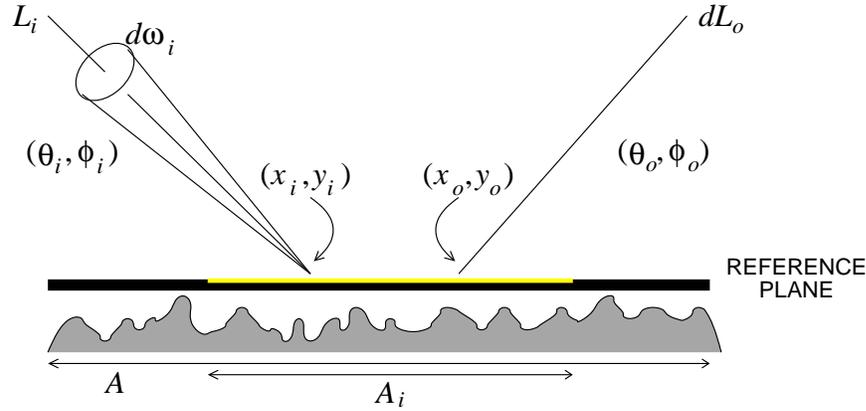


Figure 2.2: Geometry of incident and reflected beams for the definition of the BRDF. A reference plane is chosen that lies above the surface structure. (The scale of the surface structure is exaggerated in this figure.) In deriving the BRDF from the BSSRDF, it is assumed that the illumination is uniform over A_i , and that the scattering properties are uniform and isotropic *across the reference plane*.

coordinate system for the description of angular directions. The following two assumptions are made:

1. The surface is uniformly irradiated over area A_i , which completely contains the part of area A from which there is a significant contribution to the reflected radiance at (x_o, y_o) .
2. The scattering properties of the surface are uniform and isotropic¹ across the reference plane. In this case, the scattering function S does not depend on the incident and reflected points, but only on the distance between them. That is, $S = S(\theta_i, \phi_i; \theta_o, \phi_o; r)$ where $r = [(x_o - x_i)^2 + (y_o - y_i)^2]^{\frac{1}{2}}$.

When these conditions are satisfied, we can add up the contributions to the reflected radiance at (x_o, y_o) from the entire incident flux that is incident on A from direction (θ_i, ϕ_i) over the element of solid angle $d\omega_i$. Defining the BRDF as

$$f_r(\theta_i, \phi_i; \theta_o, \phi_o) = \int_A S(\theta_i, \phi_i; \theta_o, \phi_o; r) dA, \quad (2.1)$$

¹This refers to spatial isotropy of the surface and is different from the angular BRDF isotropy that is commonly considered in vision and graphics.

gives a simple relation between the incident and reflected radiance,

$$dL_o(\theta_o, \phi_o) = f_r(\theta_i, \phi_i; \theta_o, \phi_o) dL_i(\theta_i, \phi_i) \cos \theta_i d\omega_i.$$

Equation 2.1 shows that the BRDF depends only on the angular input and reflection directions, a property that follows directly from the assumption of uniform and isotropic surface scattering properties. The BRDF has units $[\text{sr}^{-1}]$.

It is often convenient to express the input (θ_i, ϕ_i) and reflected (θ_o, ϕ_o) directions using unit vectors instead of spherical coordinates, and here we use the notation $f_r(\theta_i, \phi_i; \theta_o, \phi_o)$ and $f_r(\hat{\mathbf{i}}, \hat{\mathbf{e}})$ interchangeably. Additionally, when a surface is such that the BRDF is the same at each point, it is referred to as an *homogeneous* surface; an inhomogeneous surface is said to have spatially-varying reflectance, or a *spatially-varying BRDF*.

Wavelength and polarization

In addition to being a function of input and output directions, the BRDF can be written as a function of wavelength and polarization. Nicodemus et al. [81] propose the use of a matrix of sixteen separate BRDFs to represent polarization effects. For the majority of vision tasks, however, randomly polarized light and unpolarized sensors are used, and polarization can be ignored.

In a grayscale image, we observe a weighted spectral average of the BRDF, where the weight function represents both the spectral composition of the illuminant and the spectral sensitivity of the sensor. As is customary, for the remainder of this dissertation the term BRDF will refer to this spectrally averaged BRDF. To handle colour, three separate BRDFs are used—one per channel.

2.1.1 BRDF Measurement

The BRDF is a differential quantity, so in practice, we can only observe its average over finite incident and reflected solid angles. In addition, although the BRDF is derived in terms of the reflected radiance at a single point (x_o, y_o) , any measurement device will observe the sum of the reflected radiance from a finite surface area. These are usually not large concerns

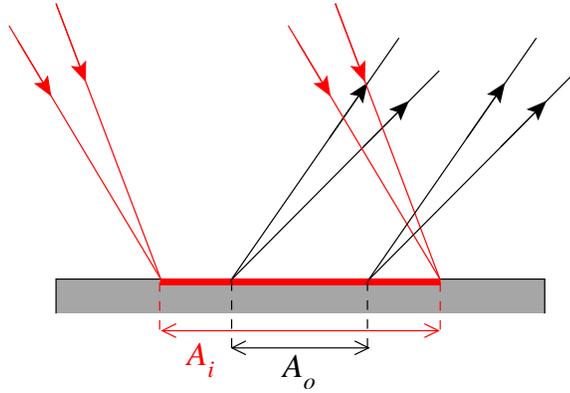


Figure 2.3: BRDF measurement of a planar sample. An area on the surface A_o is observed over a finite solid angle by the sensor (e.g., a single element of a CCD array.) This area must be uniformly illuminated over a larger area A_i , and the scattering properties of the sample must be uniform and isotropic across the reference plane. (Only extreme rays are drawn.)

provided that the solid angles are small relative to the angular variation of the BRDF, and that over its entire extent, the observed surface satisfies the assumptions underlying the BRDF (i.e., uniform and isotropic scattering properties.)

As an example, consider the measurement geometry of Fig. 2.3. A portion of planar sample is observed by a sensor through an optical system (e.g., by a single element of a CCD array.) The finite area of the sensor defines a finite area on the surface A_o . In order for this system to provide an accurate measurement of the BRDF, we require that both the illumination and the scattering effects are uniform across the surface over a larger area $A_i \supset A_o$. Here, A_i must be large enough that incident flux outside of A_i does not contribute significantly to the radiance reflected within A_o .

Traditionally, BRDF measurement is performed using uniform planar samples and a gonireflectometer. This has been greatly improved by exploiting the high spatial resolution available in CCD cameras. By using curved mirrors with a planar sample [111, 22] or simply by imaging a curved, homogeneous sample [67, 64, 71], one can capture thousands of BRDF measurements in a single image. A natural question is how best to interpolate and represent this data, which is the topic of Sect. 2.1.3. First we discuss some useful properties and special cases of BRDFs.

2.1.2 Properties of BRDFs

The BRDF is a positive function defined on the four-dimensional domain given by the cross product of the input and output hemispheres, i.e., $f_r : \Omega \times \Omega \rightarrow \mathbb{R}^+$. It is not quite an arbitrary 4D positive function, however, and is generally constrained in a few ways.

Conservation of energy. Since we assume that no energy is generated by the surface, the total reflected flux must be less than or equal to the total incident flux. This gives,

$$\int_{\Omega} f_r(\theta_i, \phi_i; \theta_o, \phi_o) d\omega_o < 1 \quad \text{for all } (\theta_i, \phi_i),$$

which is satisfied by all BRDFs.

Helmholtz Reciprocity. The property that is of greatest importance to this dissertation is a symmetry property commonly referred to as the principle of Helmholtz reciprocity. According to this principle, the BRDF is symmetric about the incident and reflected directions, i.e.,

$$f_r(\theta_i, \phi_i; \theta_o, \phi_o) = f_r(\theta_o, \phi_o; \theta_i, \phi_i). \quad (2.2)$$

Historically, it is unclear how the term ‘Helmholtz reciprocity’ came to describe this equation. In the document that defines the BRDF, the term is used without reference, and it is stated that Helmholtz reciprocity (Eq. 2.2), “holds in the absence of polarization and magnetic fields.” ([81], p40.) Although it is unclear where the term was first used, it is clear that it originates from a statement made by Helmholtz in (among other places) his *Treatise on Physiological Optics* ([41] p231). A corrected translation of his statement is given by Clarke and Parry [17]:

“Suppose light proceeds by any path whatever from a point A to another point B, undergoing any number of reflections or refractions en route. Consider a pair of mutually perpendicular planes a_1 and a_2 whose line of intersection is along the initial path of the ray at A, and a corresponding [equivalent] pair of planes b_1 and b_2 intersecting along the path of the ray at B when it comes to B. The components of the vibrations of the aether particles in these two pairs of

planes may be imagined. Now suppose that a certain amount of light J leaving the point A in the given direction is polarized in the plane a_1 ; and that of this light the amount K arrives at the point B polarized in the plane b_1 ; then it can be proved that, when the light returns over the same path, and the quantity of light J polarized in the plane b_1 proceeds from point B , the amount of this light that arrives at the point A polarized in the plane a_1 will be equal to K .

“Apparently the above proposition is true no matter what happens to the light in the way of single or double refraction, reflection, absorption, ordinary dispersion, and diffraction, provided there is no change of its refrangibility, and provided it does not traverse any magnetic medium that affects the position of the plane of polarization, as Faraday found to be the case.”

Helmholtz reciprocity (i.e., Eq. 2.2) follows directly from symmetry in the BSSRDF,

$$S(\theta_i, \phi_i, x_i, y_i; \theta_o, \phi_o, x_o, y_o) = S(\theta_o, \phi_o, x_o, y_o; \theta_i, \phi_i, x_i, y_i),$$

which represents a significant generalization of the statement by Helmholtz. (See the discussion of Clarke and Parry [17] and Veach [109] for more details.) Nevertheless, the term ‘Helmholtz reciprocity’ has come to be associated with Eq. 2.2 in photometry, graphics and vision, and we will continue to use it here.

There is no general proof of Eq. 2.2 because there are known (constructed) exceptions. The most notable involves the Faraday isolator, which was acknowledged by Helmholtz himself. For common materials, however, reciprocity is not violated, and there have been a number of empirical validations [51, 67, 23]. In a sense, the results in Chapters 3–5 of this dissertation can be viewed as further validation, since Helmholtz reciprocity is used directly for shape recovery. (See Fig. 4.4, for example.)

Isotropy. An important special case is when the BRDF is symmetric about the surface normal. In this case, it reduces to a function of three dimensions, commonly written as $f_r(\theta_i, \theta_o, \phi_i - \phi_o)$.

Bilateral Symmetry. Another, less commonly considered special case is when the BRDF is symmetric across the plane of incidence (the plane defined by the surface normal and the input direction.) This special case does not follow from isotropy and reciprocity. It was exploited by Marschner [67] to create a convenient parameterization for interpolating sampled BRDFs.

2.1.3 Representing BRDFs

Even when we account for the properties described in the previous section, the BRDF is very impractical as an image analysis tool. Instead, in computer vision and graphics we generally rely on low-dimensional BRDF models. Many of these representations are discussed in this section, with two applications in mind: 1) modeling the image formation process for vision tasks such as stereo, and 2) interpolating and representing measured BRDF data.

The simplest model of reflectance is the Lambertian model in which the BRDF is a constant function. According to this model, the BRDF is written $f_r = \rho/\pi$, and ρ is referred to as the *albedo*. Lambertian reflectance greatly simplifies the vision problem, since according to this model, the emitted radiance is independent of the viewing direction. This leads to the ubiquitous brightness-constancy constraint exploited by many stereo and structure-from-motion systems. The Lambertian model is also used for simple image-based modeling in which the surface reflectance is represented by a diffuse texture map (see, e.g., [67].)

Other, more complex, low-dimensional representations fall into two categories: parametric models and non-parametric representations.

Parametric Models

A number of models exist in which the BRDF is represented using a small number of parameters (typically 3-12 per colour channel.) The benefit of these models is that they offer a very compact representation. In addition, some of these models are simple enough to be used analytically for vision tasks such as photometric stereo [36].

In some cases, parametric models are useful for interpolating and representing measured BRDF data, since the model parameters can typically be estimated from a small number of noisy samples. At the same time, because these models are non-linear functions of their parameters, the fitting process can be difficult. A more serious drawback to using parametric models is their lack of flexibility. Each model is designed to represent a small class of surfaces, so in general, reasonable approximation of measured data requires that the correct model be manually selected *a priori*. (Of course, it is possible that a measured BRDF is not accurately fit by any known parametric model.) In addition, this lack of flexibility means that a parametric model cannot represent a general BRDF with arbitrary accuracy.

Parametric models can be divided into two categories. The first category consists of phenomenological models, i.e., empirical models that are not physically-derived. Examples include the Phong model [86] and its close relative the Blinn-Phong model [11]. Both of these models represent the BRDF as a weighted sum of a Lambertian component and a single-parameter specular lobe whose shape is described by an exponentiated cosine. These representations are isotropic, and they can violate the conservation of energy property. They require the specification of three parameters per colour channel. Other phenomenological models include the Ward model [111], which is capable of describing anisotropic reflectance, and Minnaert's model [77] for lunar reflectance.

The most prominent phenomenological model is perhaps the Lafortune model. This model is based on a primitive function referred to as the *generalized cosine lobe*,

$$g(\hat{\mathbf{i}}, \hat{\mathbf{e}}) = [\hat{\mathbf{i}}^T \mathbf{C} \hat{\mathbf{e}}]^n,$$

where $\mathbf{C} = \text{diag}(C_x, C_y, C_z)$. This non-linear function is capable of approximating a range of typical BRDF behaviours, such as anisotropy, off-specular reflection, retro-reflection, grazing-angle specularities (to a certain degree), and non-Lambertian diffuse reflection like that described by the Minnaert model. Lafortune et al. [57] introduced a BRDF representation based on the linear combination of generalized cosine lobes, and in the literature

this representation (the ‘Lafortune model’) is often described as a sum of non-linear basis functions capable of representing any BRDF with arbitrary accuracy. The problem with this characterization is that, since the representation is non-linear in its parameters, the fitting process requires iterative optimization. The functional that is minimized contains many local minima, especially when many lobes are used, and selecting appropriate initial conditions is non-trivial. As a result, when using this model to represent measured BRDF data, we are practically limited to a representation based on only two or three lobes [74]. Due to this limitation, the Lafortune model behaves much like other empirical models and can be considered as such.

The second category of parametric models are those derived from physical principles. The most common is the Cook-Torrance model [19], which is based on the Torrance-Sparrow microfacet model of surface microstructure [107]. According to this model, the surface is represented as a distribution of long, v-shaped grooves, each of which is a perfect mirror. The Cook-Torrance model describes the reflectance of many plastics and metals. Oren and Nayar [84] used the same surface model, but with Lambertian facets, to create a BRDF model that describes non-Lambertian, ‘rough-diffuse’ reflectance. Another is the Poulin-Fournier model [88] for anisotropic reflectance, which is derived from a surface microstructure consisting of parallel cylinders. All of these parametric models, like their empirical counterparts, are completely specified by a small number of parameters.

Non-parametric representations

While parametric models can yield useful approximations in some cases, they are too constrained to represent the fine detail of arbitrary, complex BRDFs. At the same time, without low-dimensional models, the need to sample the BRDF at a high angular resolution yields an unwieldy representation. There are a number of representations that attempt to fill this gap. These non-parametric representations have the flexibility to represent general BRDFs with arbitrary accuracy, but at the same time, they can be truncated to provide compact approximations.

One approach is to represent the BRDF as a linear expansions in terms of pre-chosen basis functions, such as spherical harmonics [99, 114, 90] or Zernike polynomials [54]. While spherical harmonics have proved useful in analyzing low-frequency (e.g., Lambertian) reflectance [1, 89] even moderately complex BRDFs require a large number of non-zero coefficients to be accurately represented using these bases.

Spherical wavelets have also been used to represent measured BRDF data [96, 58, 72]. Since wavelet basis functions have finite spatial support, they can be used to take advantage of the fact that typical BRDFs vary slowly over much of their domain. For example, Matusik et al. [72] densely sampled the BRDFs of 100 different isotropic materials and determined that only 69 000 non-zero wavelet coefficients (equivalent to about 5% of the data) were required to accurately represent all of the BRDFs. In Chapter 6 we present an alternative basis that also provides significant compression.

In addition to their reduced wavelet basis, Matusik et al. [72] propose the direct use of their BRDF database as a set of basis functions. Arbitrary measured BRDFs can either be written as a linear combination of the 100 BRDFs in their database, or as a linear combination of a reduced set of principle components [71]. These are rather cumbersome representations, however, since they require the storage of the basis BRDFs, each of which contains 1.5 million elements per colour channel.

Other BRDF representations, designed for efficient rendering in addition to compact representation, are the factored representations of Kautz and McCool [50] and McCool et al. [75]. They express the BRDF as a product of lower-dimensional (2D) functions,

$$f_r = \prod_i g_i \left(\pi_i(\hat{\mathbf{i}}, \hat{\mathbf{e}}) \right),$$

where the projections, $\pi_i : \Omega \times \Omega \longrightarrow \mathbb{R}^2$, are fixed *a priori* and the 2D functions g_i are fit to sampled BRDF data using SVD, normalized decomposition [50] or purely-positive factorization [75]. Since the recovered 2D functions can be stored as textures in graphics hardware, these representations are very suitable for real-time rendering applications.

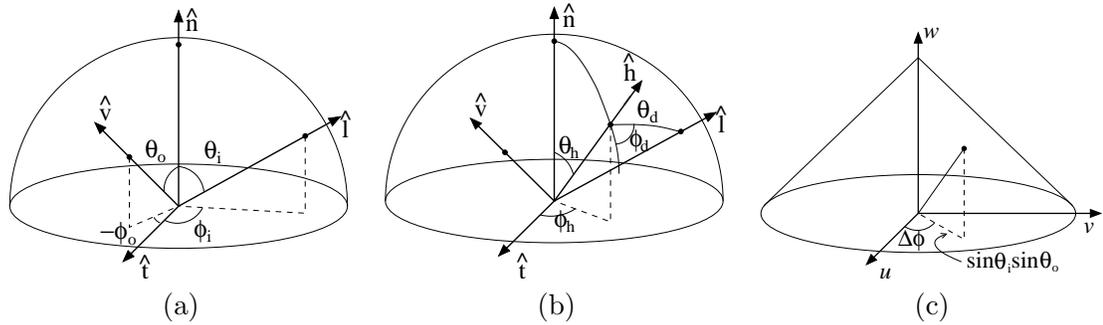


Figure 2.4: (a) Conventional in/out BRDF parameterization; (b) The halfway/difference parameterization of Rusinkiewicz; and (c) Marschner’s isotropic, bilateral-symmetric parameterization.

Parameterization

As we have described, the BRDF domain, $\Omega \times \Omega$, consists of the double hemisphere of all pairs of incident and reflected directions. Thus far, we have assumed that this 4D domain is parameterized by the spherical coordinates of these directions, as shown in Fig. 2.4(a). We are free to choose any parameterization, of course, and there are significant advantages to doing so. We discuss two additional parameterizations here, and in Chapter 6 we introduce an alternative parameterization that combines some of their benefits.

A very useful parameterization is Rusinkiewicz’s halfway/difference parameterization [93] shown in Fig. 2.4. According to this parameterization, a point in the BRDF domain is written $(\phi_h, \theta_h, \phi_d, \theta_d) \in [0, 2\pi) \times [0, \frac{\pi}{2}) \times [0, \pi) \times [0, \frac{\pi}{2})$. This parameterization has been shown to reduce sampling requirements and increase compression rates since common BRDF phenomena such as specular and retro-reflective peaks are aligned with the coordinate axes. For example, the number of required non-zero wavelet coefficients can be drastically reduced using the coordinates $(\phi_h, \theta_h, \phi_d, \theta_d)$ instead of $(\phi_i, \theta_i; \phi_o, \theta_o)$ [93, 72]. In addition, isotropy and Helmholtz reciprocity are very conveniently described using the halfway/difference parameterization. Helmholtz reciprocity implies $\phi_d \longrightarrow \phi_d + \pi$, so that ϕ_d can be restricted to $[0, \pi)$; and isotropy implies that the BRDF is a constant function of ϕ_h , meaning that this dimension can be ignored in the isotropic case.

Due to the singularities at $\theta = 0$ and the periodicity of ϕ , both the conventional in/out

parameterization and the halfway/difference parameterization are unsuitable for the application of general interpolation techniques in \mathbb{R}^3 or \mathbb{R}^4 . To overcome this, Marschner [67] presented an alternative parameterization for bilaterally-symmetric BRDFs based on a non-linear mapping of the in/out parameters:

$$(u, v, w) = (\sin \theta_i \sin \theta_o \cos \Delta\phi, \sin \theta_i \sin \theta_o \sin \Delta\phi, \cos \theta_i \cos \theta_o) \subset [0, 1]^3 \quad (2.3)$$

where $\Delta\phi = \phi_o - \phi_i$. As in the halfway/difference parameterization, this parameterization ensures reciprocity. The great advantage of this parameterization is that it is well defined everywhere in its domain, making it suitable for the application of general scattered-data interpolation techniques such as local polynomial regression [67] or radial basis functions (Chapter 6.)

2.2 Defining a suitable scale

The previous section discusses many representations of reflectance in terms of the BRDF defined at a point on a surface. In order for any of these representations to provide an accurate description of an object's appearance, the assumptions underlying the definition of the BRDF must be satisfied. If we use the BRDF to represent the appearance of a surface at the pixel level, for example, this means that the surface scattering properties must be uniform and isotropic over each region containing the area observed by each pixel. (See Fig. 2.3.)

General surfaces that are considered in image-based modeling do not satisfy this requirement, however, because they are not composed of planar facets with uniform and isotropic scattering properties. Indeed, as shown in Fig. 2.1, surfaces are generally composed of structure at a continuum of scales. The question of scale is considered briefly by Nicodemus et al. [81], where it is noted that for the BRDF, the requirement for isotropic scattering properties could be dropped, and the requirement for uniformity could be relaxed to, "a requirement for only statistical uniformity, with variations only over small enough distances to that they are not significant." ([81] p30) (The idea of statistical uniformity is also in-

investigated by Snyder [102, 103].) In general, if we wish to use the BRDF representation of reflectance for an arbitrary surface, we must be careful to choose an appropriate scale when making the division between BRDF and shape; *the scale must be chosen so that the requirement for statistical uniformity is satisfied.*

In cases where surface structure cannot be cleanly divided into shape and BRDF, we must seek alternative representations for appearance. The most common example is the bidirectional texture function (BTF) [23] in which appearance is represented (often statistically) using gross shape and an *apparent BRDF* [55] at each point.

Chapter 3

Helmholtz Stereopsis: Decoupling Shape and Reflectance

As described in Chapter 1, the radiance measurements in an image depend on the shape and reflectance of the surfaces in a scene as well as the illumination of that scene. Images typically provide only indirect information about the surface shape, and the problem considered in this chapter is the recovery of shape from these indirect measurements. We are given one or more images,

$$\mathbf{E} = g(\textit{shape}, \textit{reflectance}, \textit{illumination}),$$

and our goal is to invert the image formation process, finding $g^{-1}(\mathbf{E})$, to recover shape. We will assume that the illumination is known (as it typically is in IBM systems), in which case the coupling of shape and reflectance in the image set \mathbf{E} is the greatest obstacle to accurate shape recovery.

Most existing methods (e.g., stereo and photometric stereo techniques) approach the problem by making assumptions, either explicitly or implicitly, about surface reflectance. In contrast, this chapter presents Helmholtz stereopsis as a reconstruction process that enables the decoupling of shape and reflectance in images, so that restrictive assumptions are not required. This decoupling is achieved by capturing images to exploit the symmetry of the BRDF (Helmholtz reciprocity), and it thereby enables the accurate reconstruction of surfaces that have arbitrary, complex and unknown surface reflectance.

3.1 Related Work

Estimating 3D shape from images is a fundamental vision problem, and there is a vast literature in this area. In this section, we summarize some of the most common approaches, and since we are interested in image-based modeling, we focus on those that provide dense shape estimates. (This excludes feature-based stereo and structure-from-motion methods that typically provide shape estimates at only a sparse set of points.) Existing dense reconstruction techniques can be divided into three categories: geometric methods (fixed illumination, multiple views), photometric methods (single view, known or multiple illuminations), and photogeometric methods (multiple views and multiple illuminations.) We will discuss some of these methods here, detailing the reflectance assumptions on which they are based.

3.1.1 Geometric Methods

This category includes dense stereo and structure-from-motion techniques in which a static scene with fixed illumination is viewed from multiple, known camera positions. The goal is to recover the shape of the surfaces in the scene, a process that can be viewed as a binary classification problem: given a 3D point \mathbf{p} that is visible in two or more cameras, decide whether that point lies on a surface.

To help make this decision, one projects the point into each of the cameras and compares the radiance measurements, $\{e_1(\pi_1(\mathbf{p})), \dots, e_N(\pi_N(\mathbf{p}))\}$. (Here, $e : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is the radiance measured at an image point, and $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ projects a point \mathbf{p} onto the image plane.) Typically, the condition that these measurements are equal, i.e.,

$$e_1(\pi_1(\mathbf{p})) = e_2(\pi_2(\mathbf{p})) = \dots = e_N(\pi_N(\mathbf{p})), \quad (3.1)$$

is used as a necessary condition for \mathbf{p} to be a surface point. (There are also many stereo methods that use filtered intensities as opposed to the image intensities themselves.) While most geometric methods use additional surface information (e.g., surfaces are generally smooth) in one way or another, the constraint in Eq. 3.1, termed the brightness-constancy

constraint, is at the core of most conventional dense stereo methods. It is based on the assumption that scene radiance is independent of viewing direction, i.e., that the surface reflectance is Lambertian. The majority of surfaces, however, are not Lambertian and therefore violate this assumption. For these surfaces, noticeable variations in scene radiance occur as specularities shift with viewpoint, and smaller variations occur everywhere on the surface. In addition, if the BRDF is spatially-varying, these variations behave differently at every point on the surface. Under these conditions, recovering shape from a set of stereo images is difficult, if at all possible. (Most sparse, or feature-based, stereo methods also rely (albeit less heavily) on the Lambertian assumption; if the BRDF is arbitrary, the detected feature points may be viewpoint or lighting dependent.)

An example of a stereo method that does not assume strictly Lambertian reflectance is that proposed by Bhat and Nayar [10]. They assume predominantly Lambertian reflectance with a narrow specular lobe, and by using a trinocular rig, the specular measurements are treated as outliers and ignored.

3.1.2 Photometric Methods

In contrast to geometric methods such as stereo and structure-from-motion techniques, photometric methods recover surface shape from a single viewpoint using photometric information. A significant difference between the two classes of techniques is that for the most part, photometric techniques use shading information to directly estimate the Gauss map of a surface, whereas geometric techniques estimate the surface itself (e.g., the depth at a set of surface points). The direct estimation of surface normals is a very desirable property since accurate normals are critical for rendering and reflectance measurement (considered in Chapter 6.) In addition, if one assumes that the surface is continuous, the estimated surface normals can be integrated to recover the surface.

Photometric stereopsis is the process of using two or more illuminations of a static scene with a fixed viewpoint to estimate surface shape. Similar to conventional stereo techniques, many photometric stereo methods assume that the BRDF is Lambertian [116, 40, 59], or

that the BRDF is predominantly Lambertian with sharp specularities that can be treated as outliers [18, 27]. The methods that do not make these assumptions either assume that the BRDF is completely known *a priori* (usually through a set of reflectance maps [43]), or that it can be approximated by a simple parametric model [47, 80, 105, 36]. When we are dealing with arbitrary BRDFs, or when the form of the BRDF is spatially-varying, there is insufficient information to reconstruct both the geometry and the BRDF from a set of photometric stereo images.

Silver [100] presents a photometric stereo method that recovers the surface normals of homogeneous surfaces with an arbitrary BRDF assuming that a reference object with the same BRDF and known shape is available. Hertzmann and Seitz [42] elegantly extend this method to enable the reconstruction of surfaces whose (possibly spatially-varying) reflectance can be represented as a linear combination of the BRDFs of multiple homogeneous reference objects.

Another set of photometric methods consider the shape-from-shading problem, where the goal is to recover the surface shape using the photometric information in only a single image. Shape-from-shading methods are even more constrained than photometric stereo techniques, and they generally assume surfaces are homogeneous and that the reflectance is either Lambertian or known [46].

Magda et al. [65] presented a photometric method that can handle surfaces with arbitrary, spatially-varying BRDFs, and that is quite different from these other techniques. Their method uses many illuminations from point-sources located on two spheres of different radii surrounding the surface, and it provides depth estimates as opposed to directly estimating a field of surface normals.

3.1.3 Photogeometric Methods

A growing number of methods have been proposed that simultaneously exploit shading information *and* variation in viewpoint to recover shape. The term *photogeometric* was first used by Lu and Little [63] to describe their technique, but it seems like a suitable term for

the entire class of methods.

One set of photogeometric techniques results from incorporating shading information into stereo (fixed illumination, multiple view) systems [60, 21, 34, 49]. Because they incorporate shading cues, these methods are able to estimate shape in regions with little or no radiance variation. They assume Lambertian reflectance, and they often assume homogeneous surfaces.

Another set of photogeometric methods are those that consider photometric motion [85], in which an object moves under fixed illumination and viewpoint. In this case, both the viewpoint and illumination change (relative to the object) from frame to frame. These methods either assume Lambertian reflectance [66, 101, 119] or that optical flow is known *a priori* [85].

A more complex technique that makes less restrictive reflectance assumptions was proposed by Lu and Little [63]. They considered homogeneous surfaces with BRDFs that are isotropic (but otherwise arbitrary), and developed a method for using multiple viewpoints with collinear illumination to recover surface shape.

3.2 Reciprocal Image Pairs

While reconstruction techniques that rely on assumed reflectance models (such as those discussed in the previous section) may provide acceptable results for a restricted class of surfaces and for certain applications, many applications such as visual metrology and reflectometry require very accurate shape for surfaces in which there is no reflectance information *a priori*. In these cases, the assumptions made by these techniques can be violated, causing the recovered shape to be biased. This section shows that we can eliminate the need for restrictive reflectance assumptions by exploiting Helmholtz reciprocity. This enables the decoupling of shape and reflectance and the accurate recovery of surfaces with arbitrary BRDFs.

To see how Helmholtz reciprocity can be used to decouple shape and reflectance, consider obtaining a pair of images as shown in Fig. 3.1. The first image is captured while the object

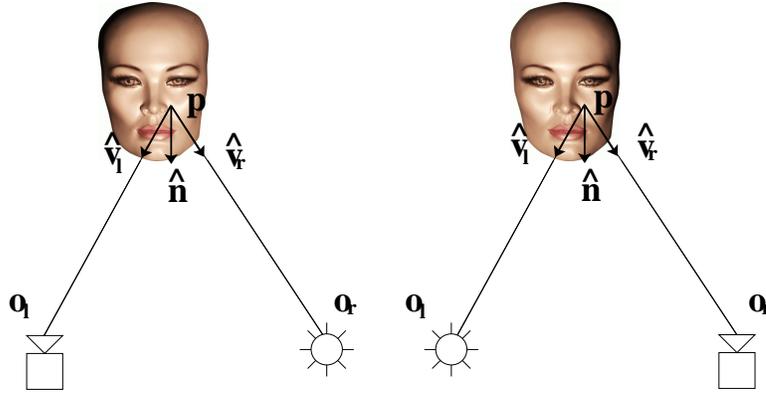


Figure 3.1: The setup for acquiring a *reciprocal pair of images* that exploits Helmholtz reciprocity. First an image is acquired with the scene illuminated by a single point source as shown on the left. Then, as shown on the right, a second image is acquired after the positions of the camera and light source are exchanged.

is illuminated by a single point light source, and the second image is captured once the camera and light source positions have been swapped. That is, the camera's center of projection is moved to the former location of the light source, and vice versa.

Figure 3.2 shows a reciprocal pair of images there were captured in this way. Although there are significant specular highlights in these images, the symmetry of the BRDF means that these specularities do not shift with viewpoint. As a result, instead of complicating the reconstruction problem, they remain fixed to the surface and becoming reliable features. A second important property of a reciprocal pair is that if a surface point is in shadow in one image, it is occluded (not visible) in the other. This is additional information that simplifies reconstruction, since shadows become reliable indicators of half-occlusion.

In addition to these qualitative properties, there is an important quantitative relationship between the radiance measurements in a reciprocal pair of images. This relationship is derived from the fact that, for any visible scene point, the ratio of the emitted radiance (in the direction of the camera) to the incident irradiance (from the direction of the light source) is the same for both images. Let \mathbf{o}_l and \mathbf{o}_r denote the positions of the camera and light source as shown in Fig. 3.1. Denote by \mathbf{p} and $\hat{\mathbf{n}}$ a point on the surface and its associated unit normal vector. Let the vectors $\hat{\mathbf{v}}_l = \frac{1}{|\mathbf{o}_l - \mathbf{p}|}(\mathbf{o}_l - \mathbf{p})$, and $\hat{\mathbf{v}}_r = \frac{1}{|\mathbf{o}_r - \mathbf{p}|}(\mathbf{o}_r - \mathbf{p})$ denote

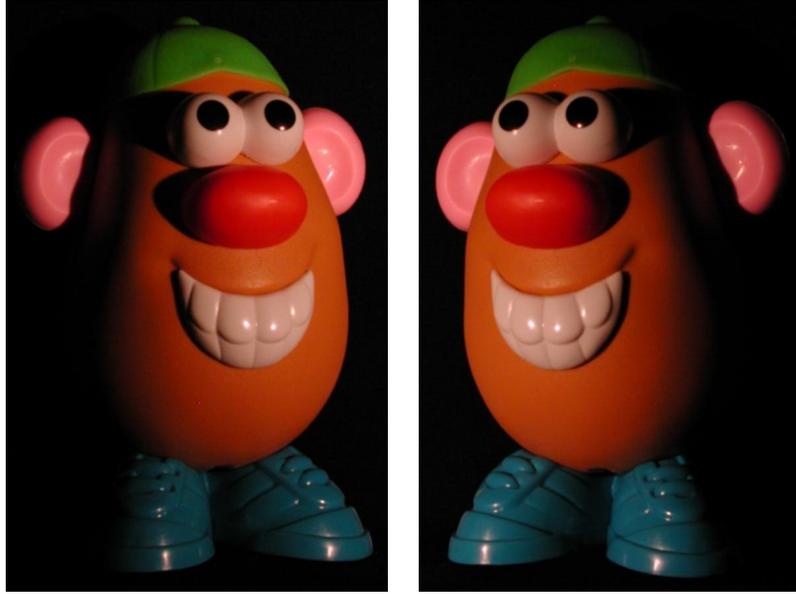


Figure 3.2: An example of a reciprocal pair of images. In contrast to a typical (fixed-illumination) stereo pair, specularities appear fixed to the surface. In addition, half-occluded and shadowed regions are in correspondence, i.e., if a point is shadowed in one image, it is not visible in the other.

the directions from \mathbf{p} to the camera and light source, respectively. Given this system, the radiance emitted from \mathbf{p} in the direction of the left camera is

$$e_l = f_r(\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_l) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \quad (3.2)$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_r$ gives the cosine of the angle between the direction to the light source and the surface normal, $\frac{1}{|\mathbf{o}_r - \mathbf{p}|^2}$ is the $1/r^2$ fall-off from a unit-strength, isotropic point light source, and f_r is the BRDF.

Now, consider the reciprocal case in which the light source is positioned at \mathbf{o}_l , and the camera observes \mathbf{p} from \mathbf{o}_r . In this case, the observed radiance is

$$e_r = f_r(\hat{\mathbf{v}}_l, \hat{\mathbf{v}}_r) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2}. \quad (3.3)$$

Because of Helmholtz reciprocity, $f_r(\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_l) = f_r(\hat{\mathbf{v}}_l, \hat{\mathbf{v}}_r)$, and we can eliminate the BRDF

term in the above two equations to obtain the *reciprocity constraint*¹

$$\left(e_l \frac{\hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2} - e_r \frac{\hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \right) \cdot \hat{\mathbf{n}} = 0. \quad (3.4)$$

In this equation, the notation e_l represents the radiance measured by the left camera at the projection of point \mathbf{p} , and can be interpreted as an abbreviation for $e_l(\pi_l(\mathbf{p}))$. Also, for calibrated cameras, the values of \mathbf{o}_l and \mathbf{o}_r are known, and the values of $\hat{\mathbf{v}}_l$, and $\hat{\mathbf{v}}_r$ can be computed for a given point \mathbf{p} . Thus, everything in the parentheses of this equation is determined by \mathbf{p} , and to reflect this, we make the definition

$$\mathbf{m}(\mathbf{p}) = \left(e_l \frac{\hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2} - e_r \frac{\hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \right).$$

Using this definition, Eq. 3.4 becomes

$$\mathbf{m}(\mathbf{p}) \cdot \hat{\mathbf{n}} = 0. \quad (3.5)$$

Note that the vector $\mathbf{m}(\mathbf{p})$ lies in the plane defined by \mathbf{p} , \mathbf{o}_r and \mathbf{o}_l (the epipolar plane).

Equation 3.4 relates surface geometry to the radiance measured in two images, and it can be used for reconstruction since it can be viewed as a constraint that must be satisfied by a surface. Unlike the brightness-constancy constraint used by conventional stereo, the reciprocity constraint is independent of the BRDF; it depends solely on the shape of the object (the point \mathbf{p} and surface normal $\hat{\mathbf{n}}$.)

There are a number of practical calibration issues that must be addressed when using Eq. 3.4 for reconstruction. First, the cameras are generally assumed to be calibrated so that the projections π and the camera/source positions \mathbf{o} are known. Second, since the reciprocity constraint involves scene radiance values (as opposed to pixel intensities), we require that the radiometric camera response functions of the two cameras be either known, or be equal and linear. If it exists, spatial variation in camera sensitivity due to optical fall-off and vignetting must also be calibrated. Finally, in developing the reciprocity constraint, we assume uniform and isotropic point light sources with equal intensity. If this is not the

¹In 1941, Minnaert [77] derived a special case of this constraint. It was used along with isotropy to increase the number of lunar reflectance measurements that could be made from Earth.

case, the sources must also be calibrated (i.e., the relative radiance of each light source as a function of output direction must be recovered.) An elegant procedure for calibrating the joint effects of source anisotropy and spatial variation in camera sensitivity for a Helmholtz stereo rig was recently developed by Janko et al. [48].

3.3 Helmholtz Stereopsis

Helmholtz stereopsis is the process of recovering shape from reciprocal image pairs. This could be achieved in a multitude of ways, in the same way that there are many different methods for reconstruction from stereo and photometric stereo images. Indeed, many Helmholtz stereo methods could be adapted from these existing techniques. In particular, the surface-evolution stereo methods of Faugeras and Keriven [29] and Yezzi and Soatto [117] seem well-suited for Helmholtz stereopsis, since they enable the use of simultaneous constraints on a surface and its Gauss map. Similarly, if the surface is the graph of a function, finite-element and parametric-surface approaches to shape-from-shading [45, 44, 104, 112] provide inspiration for the development of Helmholtz stereo methods.

A Helmholtz stereo method that is based on conventional, area-based stereo is developed in Sect. 3.5. Next, we present a comparison of Helmholtz stereopsis with typical stereo and photometric stereo methods.

3.4 Comparison to Existing Methods

Helmholtz stereopsis is a photogeometric process that allows us to combine the advantages of typical stereo and photometric stereo methods. Table 3.1 provides a summary of comparisons between typical Helmholtz stereo methods and these other two classes of methods, and this section discusses this comparison in four sections. It is important to note that this is a comparison between ‘typical’ methods, so not all of the observations made in this section will apply to all possible techniques.²

²For example, the Helmholtz stereo methods presented in this dissertation assume surface continuity, even though reciprocal pairs contain rich information about depth discontinuities that could be exploited by more general techniques.

	Assumed Reflectance	Shape: Textured Regions	Shape: Constant Regions	Shape: Depth Discont.	Handles Cast Shadows	Handles Half- Occlusion	Active/ Passive
Photometric Stereopsis	Lambertian/ Known	Normals	Normals	No	No	N/A	Active
Stereopsis	Lambertian	Depth	Nothing	Sometimes	Yes	Sometimes	Passive
Helmholtz Stereopsis	Arbitrary	Depth & Normals	Normals	Yes	Yes	Yes	Active

Table 3.1: A comparison of typical Helmholtz stereo methods and typical methods for conventional multinocular and photometric stereo. This is a comparison between ‘typical’ methods, and not all of the observations made in this table will necessarily hold for all techniques.

Assumed BRDF

As detailed in Sect. 3.1, both photometric stereo and conventional dense stereo techniques are predicated on assumptions about surface reflectance, usually assuming that the BRDF is Lambertian or of some other known parametric form. Many natural surfaces (e.g., human skin, the skin of a fruit, glossy paint) do not satisfy these assumptions, however, and cannot be accurately reconstructed by conventional techniques.

Recovered Surface Information

In stereo techniques, depth (and not a field of surface normals) is readily computed. Typically, the output of the system is a depth map—a discrete set of depth values at pixel or sub-pixel intervals. In most cases, unless a regularization process is used to smooth these depth estimates, the normal field found by differentiating the recovered depth map will be very noisy. Instead of direct differentiation of the depth map, regularized estimates of the normal field can be obtained, for example, based on an assumption of local planarity [25], or through the use of an energy functional [7]. In contrast to these methods, photometric stereo techniques generally provide a direct estimate of the field of surface normals which is then integrated (assuming continuity) to obtain a surface. Helmholtz stereo techniques are similar to photometric stereo methods (and different from typical stereo methods) in that

the Gauss map is directly estimated at each point based on photometric information.

In this way, typical Helmholtz stereo methods combine the advantages of stereo and photometric stereo methods by providing both a direct estimate of the surface depth *and* the field of surface normals. This is an important property, since a good estimate of the surface normals is critical for accurate reflectometry (as considered in Chapter 6) and rendering.

Constant Intensity Regions

Dense stereo methods work best when surfaces are highly textured; when they are not textured, the brightness-constancy constraint (Eq. 3.1) has little discriminatory power, and regularization is needed to infer the surface. (This is achieved, for example, using a statistical prior [87, 35, 69, 7], by computing a minimal surface [29], or by computing a maximally photo-consistent shape [56].) Sparse, feature-based stereo methods also have difficulty in these regions; these methods only reconstruct the geometry of corresponding feature points, so by their nature, they cannot directly reconstruct smoothly curving surfaces whose reflectance properties are constant. In contrast, photometric stereo and Helmholtz stereo methods are unaffected by lack of texture, since they can effectively estimate the field of surface normals which can be integrated to recover the surface. See Fig. 3.3 for a summary.

Depth Discontinuities

Depth discontinuities present difficulties for both stereo and photometric stereo techniques. When there is a depth discontinuity, it does not make sense to integrate the normal field that is output by photometric methods. Likewise, typical stereo algorithms have trouble locating depth discontinuities. This difficulty arises for two reasons. First, if a background object has regions of constant intensity and the discontinuity in depth occurs within one of these regions, it is quite difficult to reliably locate the boundary of the foreground object. Second, depth discontinuities induce half-occlusion in adjacent regions of the image, and these regions, which are not visible in at least one of the images, often confuse the reconstruction process.

The reciprocal image pairs used by Helmholtz stereo methods simplify the task of detect-

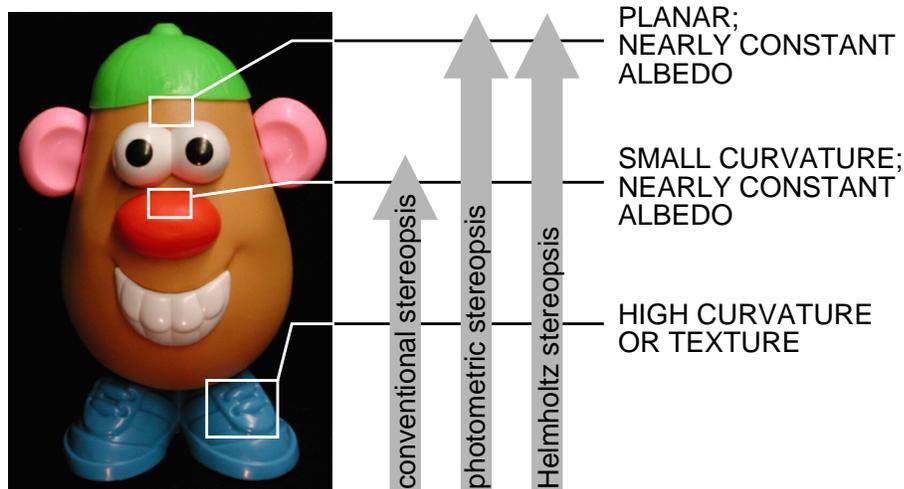


Figure 3.3: A summary of the surface properties required for Lambertian surface reconstruction by typical stereo and Helmholtz stereo techniques. Even when the BRDF is Lambertian, stereo methods are only capable of recovering surface geometry in regions of texture (i.e., varying albedo) or high curvature (e.g., edges). Neither photometric stereo nor Helmholtz stereo methods suffer from this limitation.

ing depth discontinuities since the lighting setup is such that the shadowed and half-occluded regions are in correspondence. The shadowed regions in a reciprocal image pair indicate depth discontinuities, so if one uses a method that exploits the presence of half-occluded regions for determining depth discontinuities (as done in some stereo algorithms [8, 20, 35, 6]), these shadowed regions can significantly enhance the quality of the reconstruction.

Active vs. Passive Imaging

Like photometric stereopsis and unlike conventional stereopsis, Helmholtz stereopsis is an active process. The requirement for reciprocal pairs means that the scene must be illuminated in a controlled manner, and images must be acquired as lights are turned on and off. As shown in Fig. 3.4, a suitable optical system can be constructed so that the camera and light source are not literally moved, but rather a virtual camera center and light source are co-located. Alternatively, as will be shown in the next section, a simple system can be developed that captures multiple reciprocal image pairs with a single camera and a single light source.

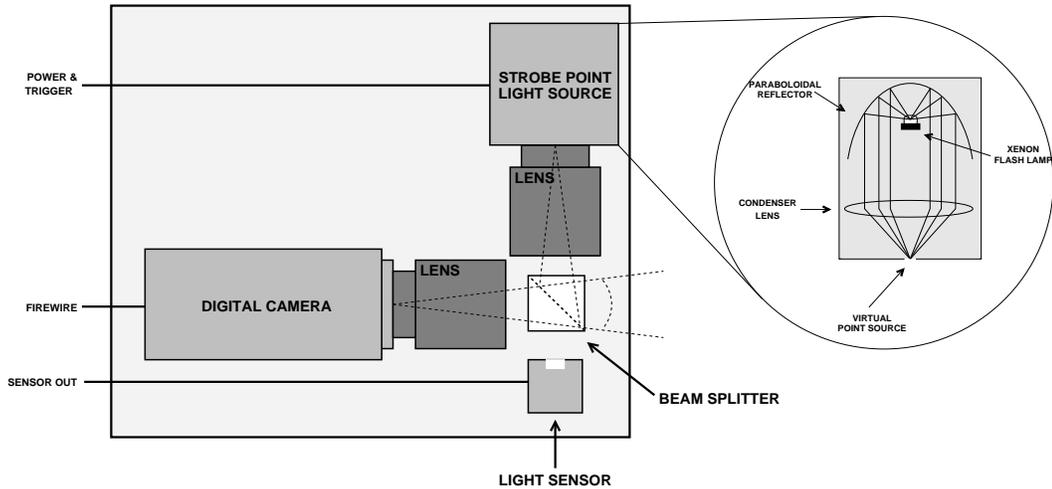


Figure 3.4: Proposed module for rapid acquisition of reciprocal image pairs. Using a beam-splitter, the camera and light source are effectively co-located. With multiple modules like this one, reciprocal image pairs can be acquired without moving the cameras and sources.

3.5 A Multinocular Helmholtz Stereo Method

In this section, we describe a Helmholtz stereo method related to traditional area-based stereo, and show how this method enables: 1) the reconstruction of surfaces with arbitrary, spatially varying BRDFs (surfaces that are neither Lambertian nor approximately Lambertian); 2) direct estimation of both surface depth and the field of surface normals; and 3) the reconstruction of surfaces in regions of constant radiance.

3.5.1 A Multinocular Reciprocity Constraint

In traditional area-based stereo methods, the surface is assumed to be the graph of a function, and an intensity constraint—usually the brightness-constancy constraint—is used to estimate a disparity function (at pixel resolution) which is linked to the surface depth. The reciprocity constraint in Eq. 3.4 is not well suited for this type of direct depth (or disparity) estimation, since it also depends on the surface normal. Using multiple reciprocal pairs, however, we can construct a multinocular reciprocity constraint that is independent of the surface normal and can be used directly as a constraint for depth estimation.

Suppose we capture N_P reciprocal image pairs as described in Sect. 3.2, and suppose

that each of these image pairs is captured from a unique pair of positions $(\mathbf{o}_{lj}, \mathbf{o}_{rj})$, $j = 1, \dots, N_P$. We can form N_P linear constraints like that in Eq. 3.4. Define $\mathbf{M}(\mathbf{p}) \in \mathbb{R}^{N_P \times 3}$ to be the matrix in which the j^{th} row is given by

$$\mathbf{m}_j^T(\mathbf{p}) = e_{lj} \frac{\hat{\mathbf{v}}_{lj}^T}{|\mathbf{o}_{lj} - \mathbf{p}|^2} - e_{rj} \frac{\hat{\mathbf{v}}_{rj}^T}{|\mathbf{o}_{rj} - \mathbf{p}|^2}.$$

Then the set of constraints from the N_P reciprocal pairs yields the *multinocular reciprocity constraint*,

$$\mathbf{M}(\mathbf{p}) \hat{\mathbf{n}} = 0. \quad (3.6)$$

For a surface point \mathbf{p} , the surface normal lies in the null space of $\mathbf{M}(\mathbf{p})$, and it can be estimated from a noisy matrix using singular value decomposition. In addition, the constraint

$$\text{rank } \mathbf{M}(\mathbf{p}) < 3, \quad (3.7)$$

provides a necessary condition that can be used to recover surface depth independent of surface orientation. At least three reciprocal pairs are required to exploit this constraint. In addition, there is a condition on $\{\mathbf{o}_i\}$, the set of camera/source positions: for each point \mathbf{p} , there must exist a triple of camera/source positions $\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3$ such that the set of points $\{\mathbf{p}, \mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3\}$ are not coplanar.

3.5.2 Capturing reciprocal images

A system that enables the acquisition of multiple reciprocal image pairs with a single camera and a single light source, can be constructed by mounting a camera and light source on a wheel as shown schematically in Fig. 3.5(a). Using this wheel, a reciprocal pair is acquired by capturing two images separated by a 180° rotation. We can capture any number of reciprocal pairs by rotating the wheel through 360° , stopping to capture images at reciprocal positions along the way.

A prototype of such a system is shown in Fig. 3.5(b). The camera is a Nikon Coolpix 990, and the light source consists of a standard 100W frosted incandescent bulb fitted with a small aperture. The camera is both geometrically and radiometrically calibrated. The

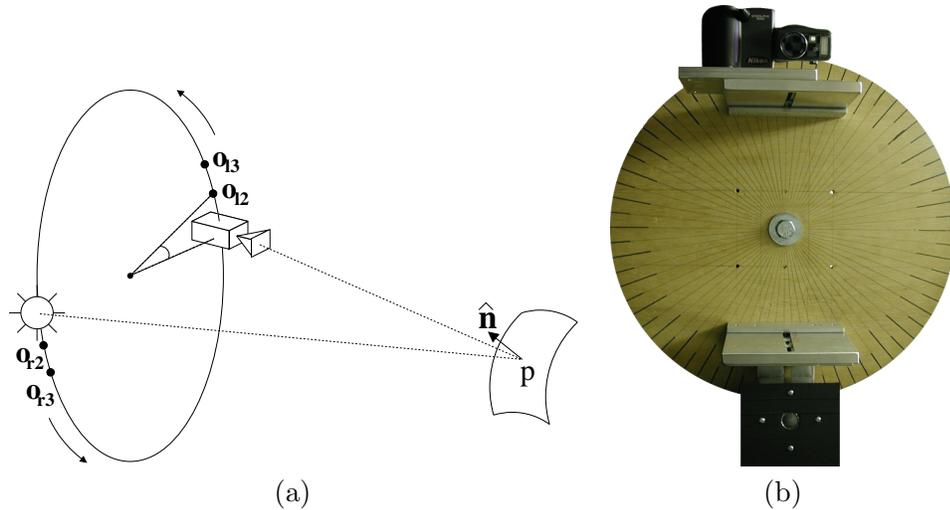


Figure 3.5: (a) A wheel is used to capture multiple reciprocal image pairs employing a single camera and a single light source. By rotating the wheel through 360° , any number of fixed-baseline pairs can be captured. (For example, images captured at \mathbf{o}_{12} and \mathbf{o}_{r2} will form a reciprocal pair.) (b) An example of the wheel design shown in (a). The light source consists of a standard 100W frosted incandescent bulb fitted with a small aperture.

former means that the intrinsic parameters and the extrinsic parameters of each camera position are known, while the latter means that we know the mapping from scene radiance values to pixel intensities (including optical fall-off, vignetting, and the radiometric camera response function.) Since the lamp is not an ideal isotropic point source, it also requires a radiometric calibration procedure in which we determine its radiance as a function of output direction. (As mentioned earlier, a procedure for performing this calibration has been developed by Janko et al. [48].)

An example of a set of images captured using this system is shown in Fig. 3.6. For all results shown in this paper the diameter of the wheel was 38cm and the distance from the center of the wheel to the scene was approximately 60cm.

3.5.3 Reconstruction

Suppose we capture n images (or $n/2$ reciprocal pairs) using the wheel in Fig. 3.5, and let the centers of projection of these n views be located at \mathbf{o}_c , $c = 1, \dots, n$. We assume that the surface is the graph of the function $z(x, y)$, $(x, y) \in U$, whose domain U is known. (The

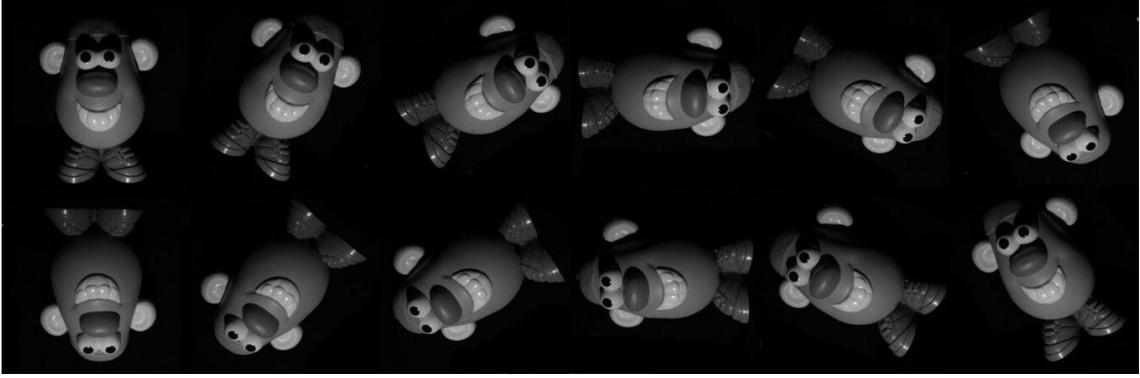


Figure 3.6: An example of 6 reciprocal images pairs captured using the rig described in Fig. 3.5. Reciprocal image pairs are arranged vertically.

function $z(x, y)$ is traditionally referred to as a depth map and is inversely related to the disparity map of binocular stereo.) We also assume that the unknown surface is completely contained within the common field of view of all n cameras, an assumption that gives upper and lower bounds on the depth map z . The depth is discretized so that $z \in \{z_{min}, \dots, z_{max}\}$, which is analogous to the discretized disparity function used in conventional binocular stereo.

We choose the (x, y) -plane to be parallel to the face of the wheel in Fig. 3.5, with the axis of rotation passing roughly through the center of U . For each point $\mathbf{x} \in U$, and for each depth value z we can construct a matrix $\mathbf{M}_{\mathbf{x}}(z)$ using Eq. 3.6. If the depth z corresponds to a surface point, this matrix will be rank 2, whereas it will be rank 3 in general. The correct depth at \mathbf{x} will be that which yields the matrix $\mathbf{M}_{\mathbf{x}}(z)$ that is ‘closest’ to being rank 2. Once the optimal depth is determined, the surface normal is uniquely determined as the unit vector that spans the 1D null space of the corresponding matrix.

While many possible measures of rank exist, since $\text{rank } \mathbf{M} \geq 2$ (assuming the camera/source condition in Sect. 3.5.1 is satisfied), a suitable measure is the ratio of the second to third singular values of \mathbf{M} . Given a matrix $\mathbf{M}_{\mathbf{x}}(z)$, we compute the singular value decomposition $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$, $\sigma_1 \geq \sigma_2 \geq \sigma_3$. Then, our measure of

rank used to select the correct depth is the ratio

$$r_{\mathbf{x}}(z) = \frac{\sigma_2}{\sigma_3}. \quad (3.8)$$

Note that at correct depth values, the ratio $r_{\mathbf{x}}(z)$ will be large.

The multinocular reciprocity constraint in Eq. 3.6 is a necessary condition for correct surface depth, but it is not sufficient. Similar to area-based stereo, we can increase the discriminative power of the constraint by assuming that the surface depth is locally constant and summing Eq. 3.8 over a window of points in U . (This is done, for example, when using SSD matching in binocular stereo.)

To estimate the depth at point \mathbf{x}_o , we consider the ratio r at this point as well as at points in a small rectangular window W around \mathbf{x}_o . Then, the estimated depth at \mathbf{x}_o is

$$z_{\mathbf{x}_o} = \arg \max_z \sum_{\mathbf{x} \in W} r_{\mathbf{x}}(z). \quad (3.9)$$

Once we have estimated the depth z_o , the least-squares estimate of the normal is

$$\hat{\mathbf{n}}_{\mathbf{x}_o} = \arg \min_{\hat{\mathbf{n}}} \|\mathbf{M}_{\mathbf{x}_o}(z_o)\hat{\mathbf{n}}\|^2, \quad \|\hat{\mathbf{n}}\| = 1, \quad (3.10)$$

which is simply given by the right singular vector corresponding to the smallest singular value of $\mathbf{M}_{\mathbf{x}_o}(z_o)$. The depth map that is recovered using Eq. 3.9 will be low in resolution due to the assumption of local depth constancy. This initial estimate of the depth can be refined using the high frequency information provided by the field of surface normals, however, and an example of this will be shown in the next section.

The Helmholtz stereo method makes no attempt at detecting half-occluded regions (even though this information is available through the visible shadows), and the reconstruction process is completely local. This method was chosen simply to demonstrate that reciprocity can be exploited for reconstruction, and we expect that improved results could be achieved using a more sophisticated Helmholtz stereo techniques. As shown in the next section, however, despite the simplicity of the method, the results are of reasonable quality.

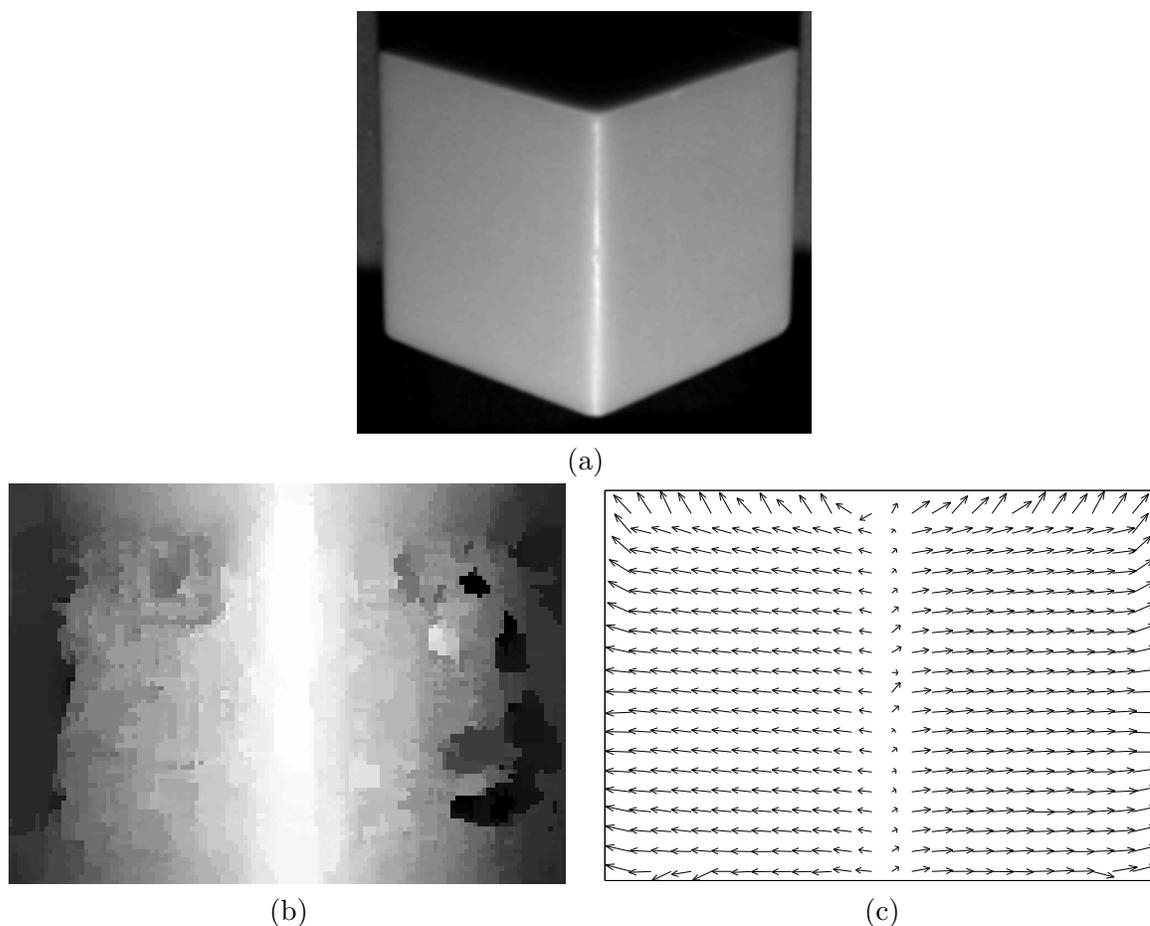


Figure 3.7: (a) one of 36 input images (18 reciprocal pairs), (b) the recovered depth map, and (c) a quiver plot of the recovered field of surface normals. As expected, even though we obtain a poor estimate of the depth due to lack of texture, the surface normals are accurately recovered.

3.5.4 Results

Figures 3.7-3.10 show the results of applying this procedure to four different objects. Each figure consists of: (a) one of the input images of the object, (b) the depth recovered using Eq. 3.9, and (c) the recovered field of surface normals.

Figure 3.7 is a demonstration of a surface reconstruction in the case of nearly constant image brightness. This surface (a wax cube) is a member of the class of surfaces described at the top of Fig. 3.3, and it is an example of a case in which conventional stereo has difficulty. Notice that our simple Helmholtz stereo method accurately estimates the normal field, even

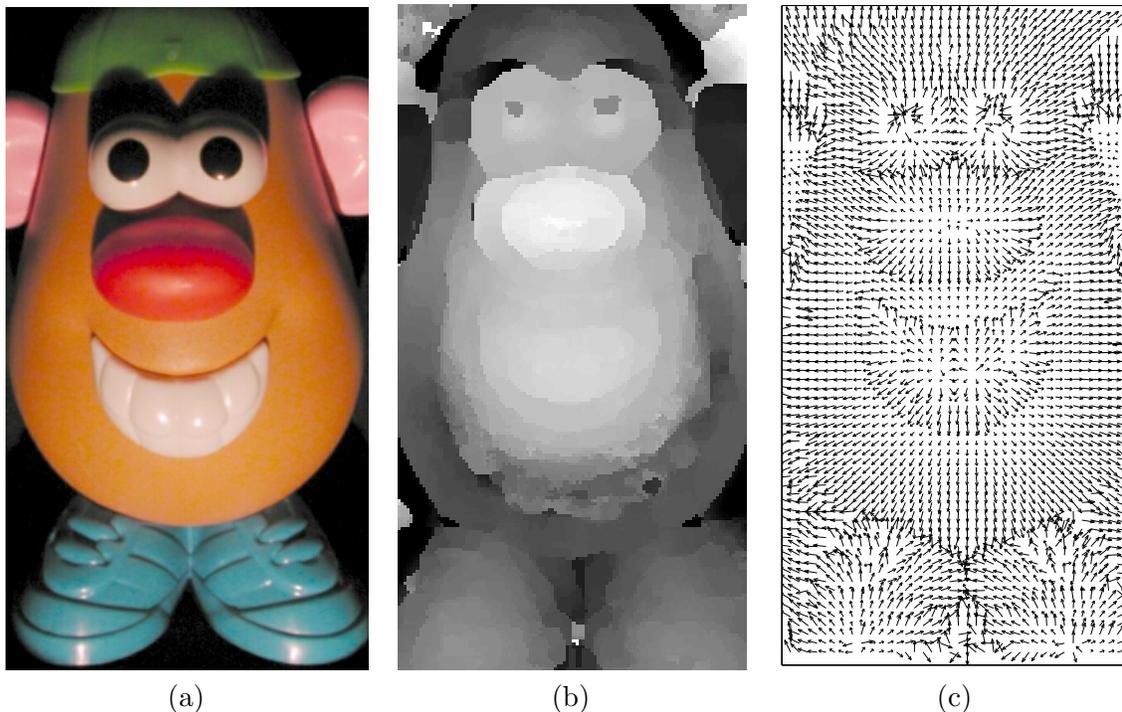


Figure 3.8: (a) one of 34 input images (17 reciprocal pairs), (b) the recovered depth map, and (c) a quiver plot of the recovered field of surface normals. As evidenced by the specularities in (a), the surface is non-Lambertian. Regions of very small albedo (e.g., the iris of the eyes, the background) are sensitive to noise and erroneous results are expected there. Elsewhere, the depth and orientation are accurately recovered. A 9×9 window was used in the depth search.

though the depth estimates are poor. The poor depth estimates are expected since at an image point \mathbf{x} , the ratio $r_{\mathbf{x}}(d)$ will be nearly constant for a small depth interval about the true surface depth. The normals are accurate, however, since each corresponding matrix $\mathbf{M}_{\mathbf{x}}(d)$ will have nearly the same null space.

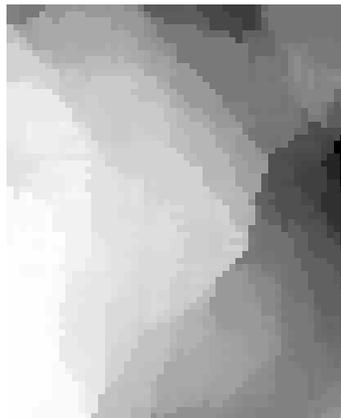
Figure 3.8 shows the results for a surface that is clearly non-Lambertian. The specularities on the nose, teeth and feet attest to this fact. Note that the reconstruction method is not expected to succeed in regions of very low albedo (e.g., the background as well as the iris of the eyes) since these regions are very sensitive to noise.

Figures 3.9 and 3.10 show two more examples of surface reconstructions. Again, note that the recovered surface normals are accurate despite the low resolution of the depth estimates, even in regions of nearly constant image brightness.

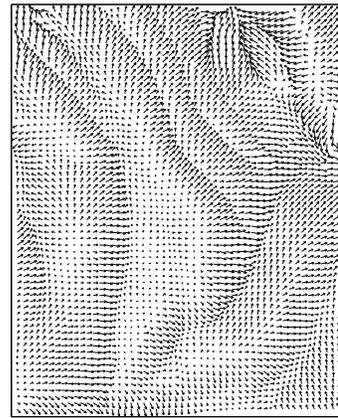
As mentioned at the end of the last section, it is possible to obtain a more precise surface reconstruction by integrating the estimated normal field. The examples above demonstrate that this field is accurately estimated, even in regions where the depth is not. To illustrate how surfaces can be reconstructed in this way, we enforced integrability (using the method of Frankot and Chellapa [33] with a Fourier basis) and integrated the vector fields shown in Figs. 3.7(c) and 3.10(c). The results are shown in Figs. 3.11 and 3.12. As seen in these figures, the high resolution information provided by the surface normals enables the recovery of precise surface shape — more precise than what we would expect from most conventional stereo methods. Note that it might be possible to obtain similar reconstructions using a photometric stereo method, but this would require an accurate model for the reflectance at each point on the surface.



(a)



(b)



(c)

Figure 3.9: A reconstruction for the marked interior region of a ceramic figurine shown in (a). Figures (b), and (c) are the depth map, and normal field. The low resolution of the depth map is caused by the 11×11 window used in the depth search, but this does not affect the accuracy of the estimated surface normals. Eighteen reciprocal image pairs were used.

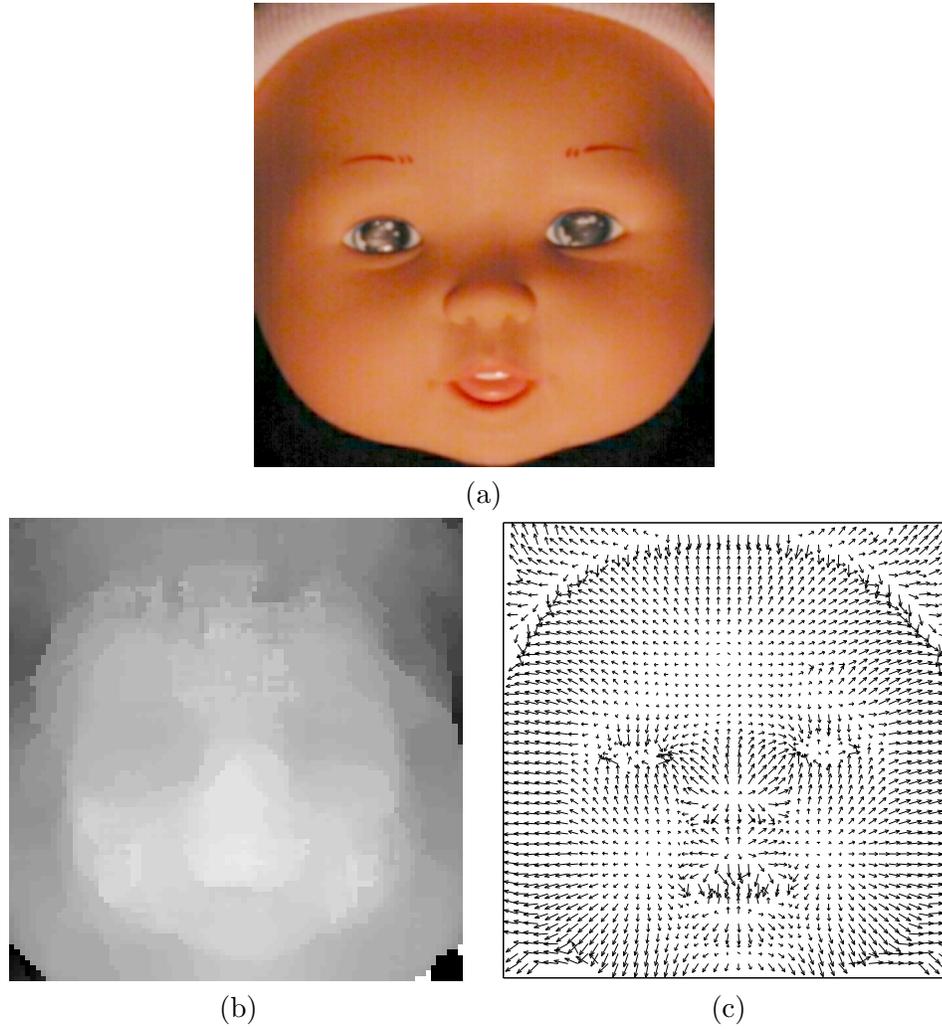


Figure 3.10: A reconstruction for the face of a plastic doll shown in (a). Figures (b) and (c) are the estimated depth map and normal field. Eighteen reciprocal image pairs and a 9×9 window were used.

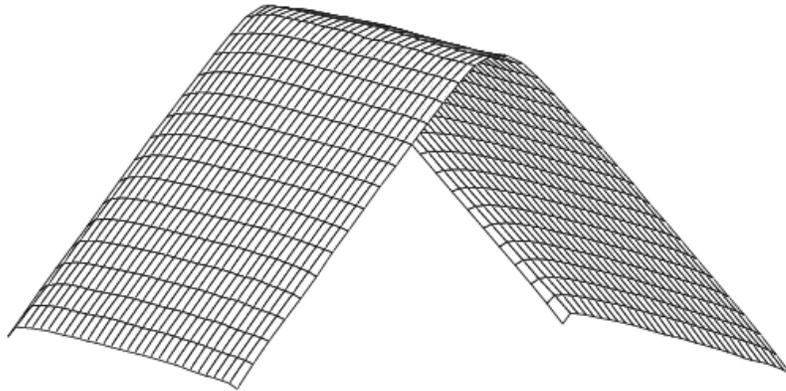


Figure 3.11: The surface that results from integrating the normal field shown in Fig. 3.7(c). Every third surface point is shown, and the surface is rotated for clarity.

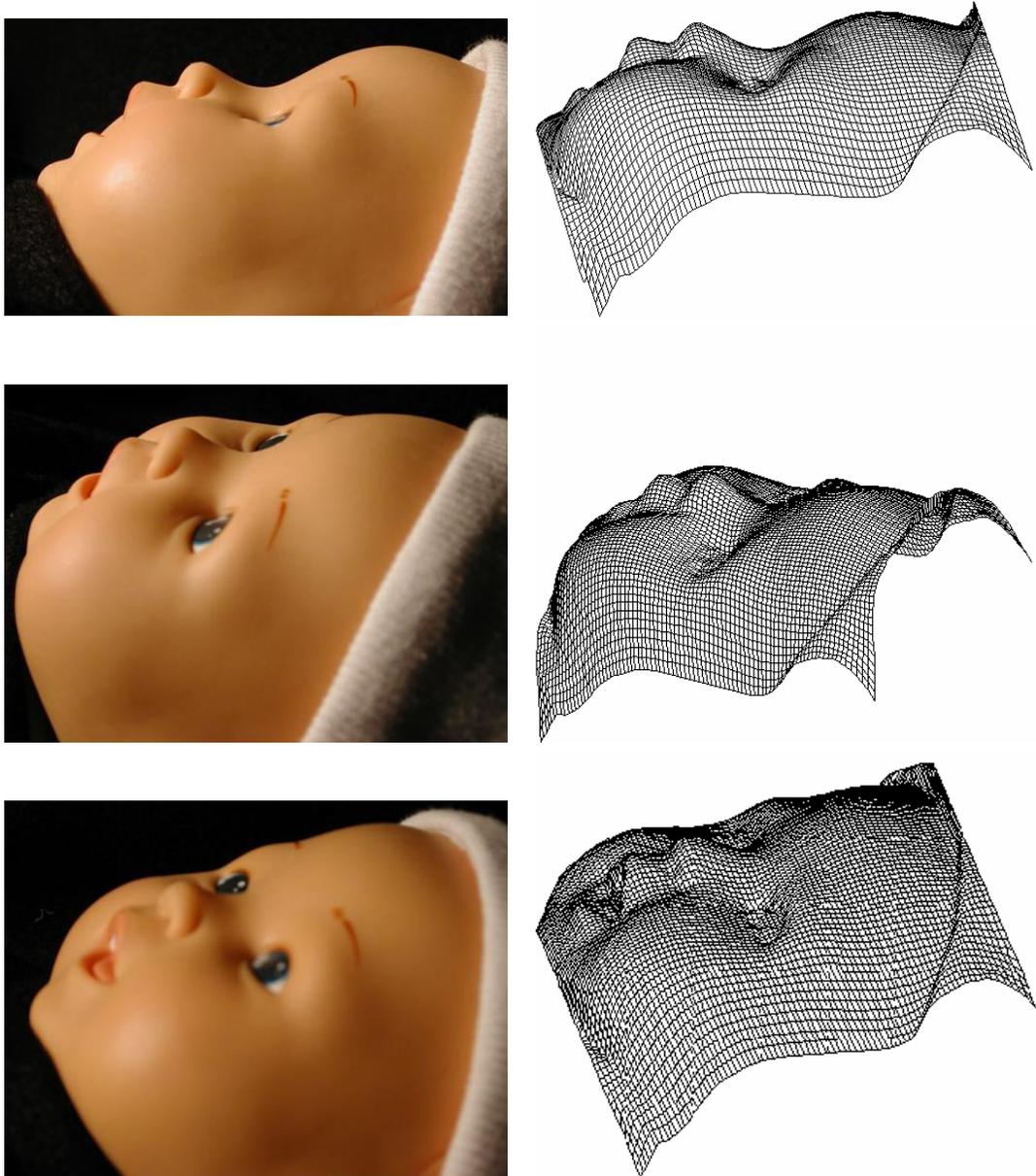


Figure 3.12: Three views of the surface that results from integrating the normal field shown in Fig. 3.10(c). To demonstrate the accuracy of the reconstruction, a real image taken from each corresponding viewpoint is displayed. The specularities on the doll's face clearly show that the surface is non-Lambertian.

3.6 Summary

This chapter introduces the notion of a reciprocal image pair as a means of decoupling shape and reflectance in image data. Due to Helmholtz reciprocity, reciprocal pairs have a unique property: the relationship between radiance measurements at corresponding image points is independent of the BRDF. An expression for this relationship is derived, and it is shown how it can be used as a constraint for shape recovery.

Helmholtz stereopsis is introduced as the process of recovering shape from reciprocal images, and since it enables the direct estimation of both a surface and its Gauss map, it allows us to combine the advantages of both conventional and photometric stereo methods. In contrast to these traditional methods, however, Helmholtz stereo methods can recover this geometric information for surface with arbitrary, complex and spatially-varying BRDFs.

Finally, we introduce a design that is capable of gathering reciprocal images in a controlled manner with a single camera and a simple approximation to a point light source, and we develop a simple Helmholtz stereo method (based on traditional area-based stereo) that uses multiple reciprocal pairs to recover accurate shape.

Chapter 4

Binocular Helmholtz Stereopsis

The last chapter introduced the reciprocity constraint which relates the image measurements at corresponding points in a reciprocal pair of images. It was shown that the combined constraints from at least three reciprocal pairs of images provide a simple means of directly estimating both a surface and its Gauss map.

In this chapter, we re-examine the reciprocity constraint, and investigate conditions in which shape can be recovered from a single reciprocal pair, such as that shown in Fig. 4.1. By writing the reciprocity constraint as a PDE, we develop a binocular Helmholtz stereo method that can reconstruct a surface when the depth along an initial curve is given and the surface is the graph of a C^1 function. We also show how regularization can be used to recover the surface when the initial conditions are not available.

4.1 The Reciprocity Constraint as a PDE

In the last chapter we derived the reciprocity constraint that relates the measured radiance at corresponding image points in a reciprocal pair of images. Given a reciprocal pair (captured as shown in Fig. 3.1), the constraint is

$$\left(e_l \frac{\hat{\mathbf{v}}_l(\mathbf{p})}{|\mathbf{o}_l - \mathbf{p}|^2} - e_r \frac{\hat{\mathbf{v}}_r(\mathbf{p})}{|\mathbf{o}_r - \mathbf{p}|^2} \right) \cdot \hat{\mathbf{n}} = 0,$$

where e_l is an abbreviation for $e_l(\pi_l(\mathbf{p}))$, i.e., the radiance measured by the left image at the projection of point \mathbf{p} . For clarity, we have written the unit vectors $\hat{\mathbf{v}}_l$, $\hat{\mathbf{v}}_r$ as explicit functions of \mathbf{p} .

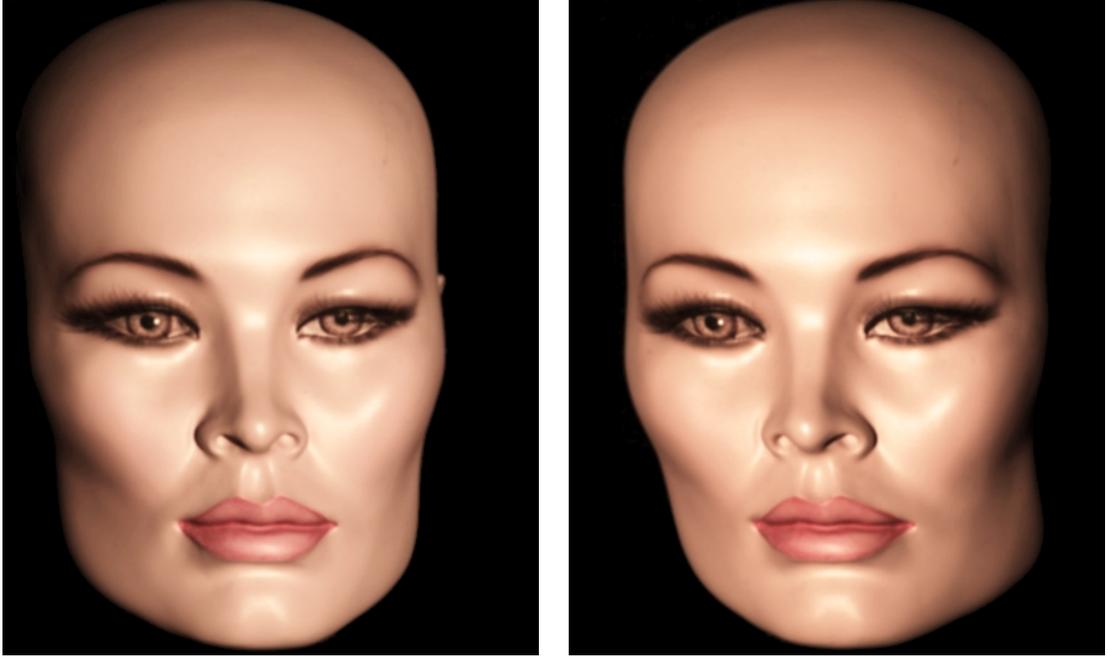


Figure 4.1: Two rectified images of a painted, plastic mannequin head acquired as a reciprocal pair. Note the prominent specularities.

Consider an imaging situation in which the distances from the light source and camera to the scene are large relative to the relief of the scene, and assume the camera fields of view to be narrow. Under these conditions, the cameras can be modeled by scaled orthographic projection, and the vectors $\hat{\mathbf{v}}_l(\mathbf{p})$ and $\hat{\mathbf{v}}_r(\mathbf{p})$ can be taken as constant over the scene. As well, the denominators $|\mathbf{o}_l - \mathbf{p}|^2$ and $|\mathbf{o}_r - \mathbf{p}|^2$ can each be taken as constant over the scene. The ratio $|\mathbf{o}_l - \mathbf{p}|/|\mathbf{o}_r - \mathbf{p}|$ can be easily determined when calibrating a Helmholtz stereo rig, and here we take this ratio to be one. Under these assumptions, the reciprocity constraint reduces to the *distant-source reciprocity constraint*,

$$(e_l \hat{\mathbf{v}}_l - e_r \hat{\mathbf{v}}_r) \cdot \hat{\mathbf{n}} = 0, \quad (4.1)$$

where $\hat{\mathbf{v}}_l$ and $\hat{\mathbf{v}}_r$ are constants determined during calibration.

The distant-source reciprocity constraint is a first order, nonlinear PDE relating the surface coordinates \mathbf{p} and their derivatives expressed through the normal $\hat{\mathbf{n}}$. This equation is very similar to the fixed-viewpoint, Lambertian case considered in a very different context

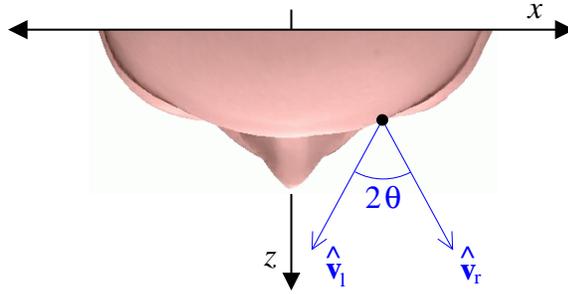


Figure 4.2: The cyclopean coordinate system used to simplify the distant-source reciprocity constraint. The x -axis is parallel to the epipolar planes, the z -axis bisects the two camera/source directions, and the surface is the graph of a C^1 function $z(x, y)$.

by Belhumeur and Jacobs [5] and Chen et al. [15]. As was done there, we can obtain a solution to Eq. 4.1 using the method of characteristic curves. That is, by performing a change of variables, we obtain a PDE in one variable, and under suitable conditions, the depth of the surface along a curve on the object is a solution to this PDE. In our case, each characteristic curve lies in a unique epipolar plane. (These planes are parallel since we assume scaled orthographic projection.)

As shown in Fig. 4.2, we establish a cyclopean coordinate system (as in [7]) by defining the z -axis to be the bisector of directions \hat{v}_r and \hat{v}_l , and the x -axis to be in direction $\hat{v}_r - \hat{v}_l$. In this coordinate system, epipolar planes are planes of constant y , and we have $\hat{v}_l = (-\sin \theta, 0, \cos \theta)$ and $\hat{v}_r = (\sin \theta, 0, \cos \theta)$ where 2θ is the angle between the two camera/source directions. Let the coordinates of points in the world be expressed in this system as (x, y, z) , and assume the surface is a graph of a C^1 function $z(x, y)$. Furthermore, assume the pair of images is rectified, so that corresponding scanlines in the left and right images lie in an epipolar plane. In this system, the point (x, y, z) will project to $(x \cos \theta + z \sin \theta, y)$ in the left image and to $(x \cos \theta - z \sin \theta, y)$ in the right image. The disparity is given by $2z \sin \theta$.

Expressing the surface depth as $z(x, y)$ and noting that the unnormalized surface normal is $(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1)$, we can write the distant-source reciprocity constraint as

$$-\frac{\partial z}{\partial x} \sin \theta (e_l(x_l, y) + e_r(x_r, y)) - \cos \theta (e_l(x_l, y) - e_r(x_r, y)) = 0,$$

where $x_l = x \cos \theta + z \sin \theta$ and $x_r = x \cos \theta - z \sin \theta$. This holds for all y . Rewriting this, we have

$$\frac{\partial z}{\partial x} = \cot \theta \frac{e_r(x_r, y) - e_l(x_l, y)}{e_r(x_r, y) + e_l(x_l, y)}, \quad (4.2)$$

which can be numerically integrated as

$$z(x, y) = \int_{x_0}^x \frac{\partial z}{\partial x} dx + z(x_0, y). \quad (4.3)$$

In other words, for each epipolar plane (indexed by y), this integral can be independently evaluated to estimate a profile of the surface corresponding to its intersection with the epipolar plane. There is no search for correspondence over disparity space, as correspondence is determined as a byproduct of integration. In order to evaluate Eq. 4.3, however, we require the initial conditions $z(x_0, y)$. There are two ways to look at this issue. On one hand, knowing $z(x_0, y)$ for some (x_0, y) in an epipolar plane amounts to having the means to estimate the depth for all points in that plane. Alternatively, one can view Eq. 4.3 as defining a one-parameter family of reconstructed curves in each epipolar plane, with each element of the family indexed by a different depth values at (x_0, y) . In Sec. 4.2, we take the latter view and use regularization to algorithmically choose a member of this family for each epipolar line, thereby obtaining a complete reconstruction of the surface.

In order to evaluate the practicality of Eq. 4.3, we gathered reciprocal pairs of images of three cylinders made of an approximately Lambertian material, a rough-diffuse non-Lambertian material [84], and a specular plastic material. Images were acquired with a Kodak DCS 760 (12-bit) digital camera, and the scene was illuminated with a 150W halogen bulb. The cameras were geometrically calibrated, and the distance from the cameras to the object was about two meters, which satisfies the assumptions made in deriving Eq. 4.1. Figure 4.4 shows, for each cylinder, a pair of rectified images and a plot of the image intensities in a single epipolar plane. These intensity curves are characteristic of these three material types.

For each epipolar plane, a family of reconstructed profiles was obtained by repeatedly using Runge-Kutta integration in Matlab with a discrete set of initial depths. Figure 4.3

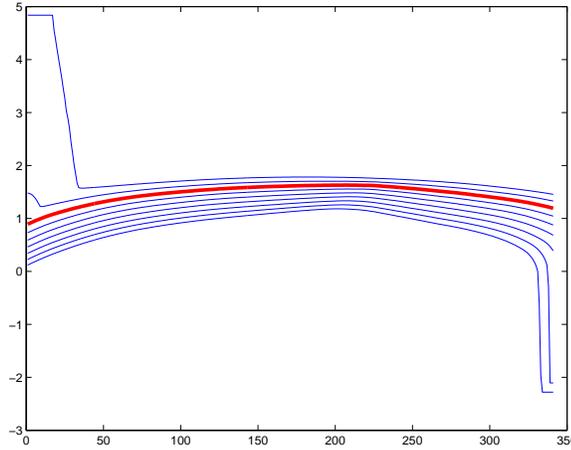


Figure 4.3: A family of reconstructed curves in one epipolar plane for the specular cylinder images in the bottom row of Fig. 4.4. The family arises from different initial conditions $z(x_0, y)$ when integrating Eq. 4.3. The thick (red) curve is the member of this family with the correct geometry, and is redrawn with a different scaling in the lower right of Fig. 4.4.

shows such a family for the specular cylinder. The optimal reconstruction can be selected from this family from knowledge of a single correspondence or by some other means (e.g., smoothness, a shape prior, etc.) The last column of Fig. 4.4 shows the reconstructed profile across one epipolar line overlaid on a circular cross section. (The initial conditions were chosen manually.) The RMS errors between the reconstructed curve and overlaid circle as a percentage of radius are 0.11%, 1.7%, and 0.94%, respectively, for the Lambertian, generalized Lambertian, and specular cylinders. The reconstructed curve for the Lambertian cylinder is indistinguishable from the ground truth circle whereas there is a slight deviation for the specular cylinder.

4.2 Surface Reconstruction

As discussed in the previous section, we can recover a surface by integrating a characteristic curve in each epipolar plane, but this requires knowledge of an initial depth in each plane. This section presents a simple and efficient technique that uses regularization to algorithmically determine these initial conditions.

Finding an initial depth in an epipolar plane is equivalent to finding a single binoc-

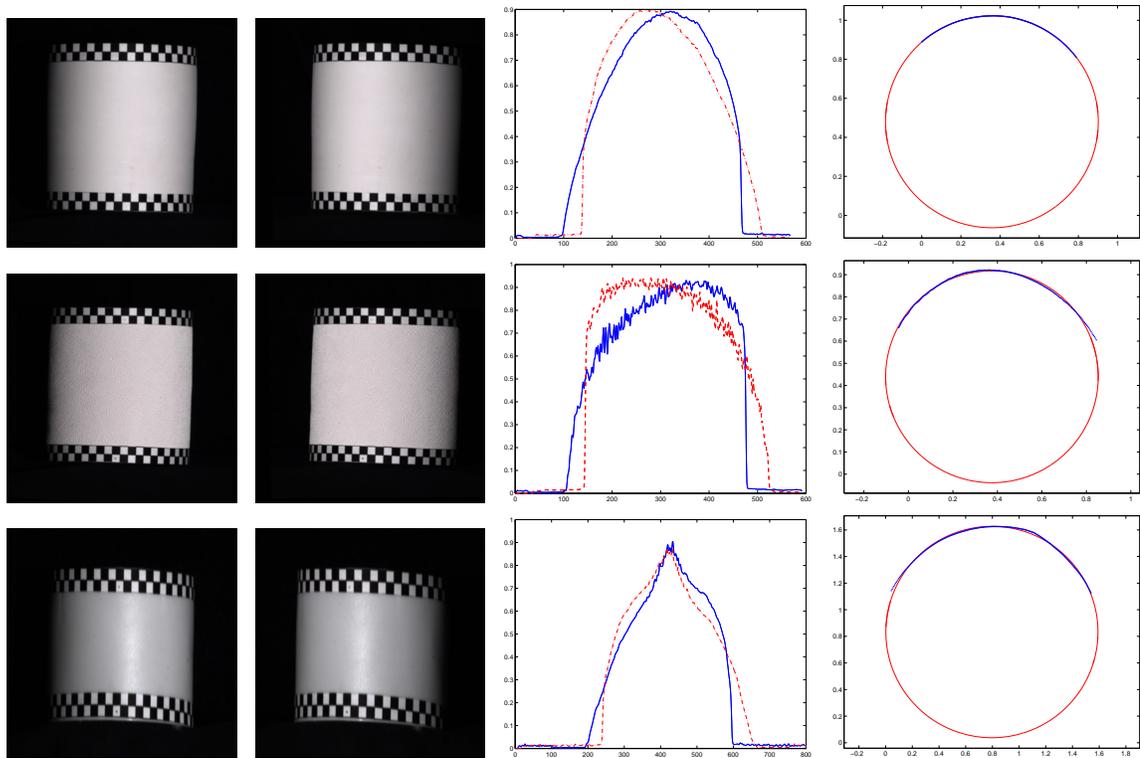


Figure 4.4: Reconstruction of three real cylinders of three material types: The cylinder in row 1 is approximately Lambertian, the cylinder in row 2 has rough-diffuse non-Lambertian reflectance [84], and the plastic cylinder in row 3 is highly specular. The first two columns show a rectified pair of images of the cylinder. The third column shows a plot of the image intensities across one epipolar line in the left (blue, solid) and right (red, dashed) images. The fourth column shows the reconstructed profile (thick, blue) in the epipolar plane superimposed on a circular cross section.

ular correspondence. (This is a much simpler problem than establishing correspondence at *all* points as required by conventional binocular stereo.) To solve this problem, one could develop an algorithm that used salient features (such as edges) to establish the initial correspondence, since surfaces generically produce rapid variations in image intensity corresponding to rapid variations in the surface normal and variations in the BRDF across the surface. Additionally, the fact that shadowed surface regions correspond to half-occluded regions in a reciprocal pair of images provides a means of establishing initial image correspondence since these events generally cause discontinuities in image intensity. (Tu and Mendonça [108] used shadow/half-occlusion boundaries as surface control points in their

binocular Helmholtz stereo method that was developed independently from our work.)

While coupling a feature-based stereo algorithm with the solution of Eq. 4.1 might work, it would rely on the identification of feature points, and would break down ungracefully when the initial correspondences were incorrectly assigned. An alternative approach is to eliminate the requirement for initial conditions altogether, exploiting instead the fact that the surface is continuous. In this approach, one recovers the surface $z(x, y)$ that minimizes a functional

$$E(z) = E_{data}(z) + \lambda E_{surface}(z) \quad (4.4)$$

where

$$E_{data} = \iint \left(\frac{\partial z}{\partial x} - r(z) \right)^2 dx dy \quad (4.5)$$

with

$$r(z) = \cot \theta \frac{e_r(x_r, y) - e_l(x_l, y)}{e_r(x_r, y) + e_l(x_l, y)}. \quad (4.6)$$

The second term in Eq. 4.4 reflects prior knowledge about the surface, and it can take a variety of forms. As typically done in variational shape-from-shading [45], we can exploit the fact that the surface is continuous (and therefore integrable) using

$$E_{surface}(z) = \iint \left(\frac{\partial^2 z}{\partial x \partial y} - \frac{\partial^2 z}{\partial y \partial x} \right)^2 dx dy.$$

Alternatively, we can bias our solution toward those that are smooth in some sense. Commonly used smoothness terms (stabilizers) include

$$E_{surface}(z) = \iint \left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2 dx dy$$

and

$$E_{surface}(z) = \iint \left(\frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z}{\partial y^2} \right)^2 dx dy.$$

Of course, we can also use a combination of an integrability term and a smoothness term.

This section describes a fast technique that *approximates* the minimization of Eq. 4.4, using the smoothness term

$$E_{surface}(z) = \iint \left(\frac{\partial z}{\partial y} \right)^2 dx dy. \quad (4.7)$$

This functional is chosen for two reasons. First, since the characteristic curves are each continuous by construction, it is unnecessary to enforce smoothness in the x direction. Second, using Eq. 4.7, we can find an approximation to the global minimum of Eq. 4.4 very efficiently.

Our approximate minimization of the functional in Eq. 4.4 happens in two passes. First, each epipolar plane is considered independently, and instead of estimating a single curve in each plane, we recover a family of possible curves. In the second pass, we use the functional in Eq. 4.7 to choose one curve from each epipolar family, thereby obtaining a complete surface. Both passes are implemented very efficiently using dynamic programming.

Pass 1: Within Epipolar Planes

Consider a single epipolar plane (i.e., a plane of constant y) and denote the profile of the surface in this plane by $z(x)$. In order to estimate $z(x)$, we minimize the functional

$$E_{data}(z) = \int \left(\frac{\partial z}{\partial x} - r(z) \right)^2 dx + \alpha \int |\nabla e_l - \nabla e_r|^2 dx \quad (4.8)$$

where $e_l = e_l(x_l, y)$, $e_r = e_r(x_r, y)$, α is a weighting term and $r(z)$ is given by Eq. 4.6 with the known value of y . This functional represents a robust version of Eq. 4.5, one that combines our reciprocity-based reconstruction with conventional feature-based stereo. In regions with little texture, the image gradients are small, and the first term dominates. Therefore, in these regions we effectively recover the surface by estimating its slope from the photometric variation in the reciprocal pair. In textured regions, however, the finite resolution in the images means that the ratio $r(z)$ cannot be reliably estimated, and we instead estimate the surface depth by aligning ‘features’ (i.e., the image gradient) in the two images.

For each epipolar plane, we use dynamic programming to find a discrete approximation to the curve $z(x)$ that minimizes Eq. 4.8. For n discrete values of x , we consider m possible depth values, and the computational cost of finding the global minimum of Eq. 4.8 is $O(nm^2)$; see [9] for details.

Note that by finding the global minimum of Eq. 4.8, we have eliminated the need for an initial condition. When an epipolar plane includes an image feature such as an albedo edge, the gradient matching term locks onto the edge, effectively providing the initial depth needed for the integration. Thus, we expect good results for epipolar planes that include features such as the eyes, eyebrows, and lips in Fig. 4.1. We are not guaranteed, however, that each epipolar plane will have a significant feature that enables the minimization to find the correct solution. (See, for example, the forehead and the bridge of the nose in Fig. 4.1.) To overcome this, we use regularization *across* epipolar planes to recover the surface.

Pass 2: Across Epipolar Planes

Regularization across epipolar planes can be implemented very efficiently using the following basic idea. First, we compute as the output from Pass 1 a family of n solutions minimizing Eq. 4.8 with the additional constraint that the endpoints of the solutions vary over a range of possible z values. (The endpoints are used to index the family of solutions since this family is obtained as a by-product of dynamic programming. Note that this family differs from the one arising in Sec. 4.1 and shown in Fig. 4.3.) If the family of solutions is big enough (i.e., if our sampling of the depth z is fine enough), then the correct solution should be well represented by one member from the family. This is true whether or not the epipolar plane has a salient image feature. To choose the optimal curve for each epipolar plane, we simply choose the collection of curves (one per plane) that minimizes our smoothness functional.

More precisely, let $x = x_e$ denote the end of the epipolar line. Let $z(x_e, y) = z_e$ denote the surface depth at the endpoint $x = x_e$ for epipolar plane y . For each y and for each ending point z_e in the range of possible z values, we compute a solution

$$\hat{z}(x, y | z(x_e, y) = z_e) = \arg \min_{(z(x, y) | z(x_e, y) = z_e)} E_{data}(z(x, y)). \quad (4.9)$$

In other words, $\hat{z}(x, y | z(x_e, y) = z_e)$ is the solution along epipolar line y that minimizes Eq. 4.8 subject to the constraint $z(x_e) = z_e$. For each epipolar plane y , we have a family of curves, and this family is indexed by z_e . Within this family there is a value of z_e , and a

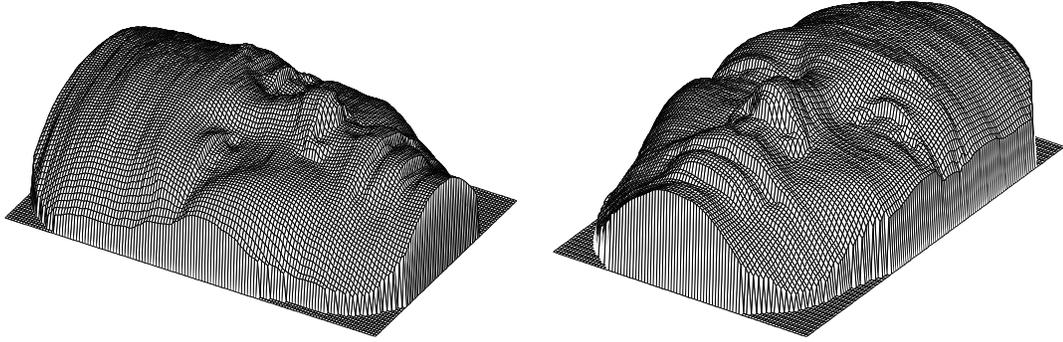


Figure 4.5: Two views of a reconstructed plastic mannequin created using a binocular Helmholtz stereo technique.

corresponding solution $\hat{z}(x, y | z(x_e, y) = z_e)$, that is close to the correct solution. We denote the family of solutions over all epipolar planes by $\hat{Z}(z_e; y) = \{\hat{z}(x, y | \hat{z}(x_e, y) = z_e)\}$.

In order to recover the surface, we must select one curve from the family of solutions in each epipolar plane. We do this by minimizing the functional in Eq. 4.7, and we take as our solution to Pass 2,

$$\hat{z}(x, y) = \arg \min_{\hat{z}} E_{surface}(z(x, y)). \quad (4.10)$$

As in Pass 1, we use dynamic programming to find $\hat{z}(x, y)$. The computational cost of this dynamic programming step is $O(m^2l)$ where m is the number of endpoints (depth values), and l is the number of epipolar planes (values of y).

Note that this formulation has an inherent asymmetry, as the second pass considers a range of ending points and not a range of starting points. We correct this by re-running this two stage process in reverse. Specifically, we run Pass 1 and Pass 2 across the data to find a collection of optimal ending points $\hat{z}_e = \hat{z}(x_e, y)$ for each epipolar line. We then re-run Pass 1 in reverse (i.e., from right to left), fixing the endpoint such that $z(x_e, y) = \hat{z}_e$ for each y . At this stage, for each y we now have a family of solutions indexed by the value of the beginning point z_b . The overall solution is then chosen by re-running Pass 2 to select the optimal curve from each family.

This algorithm has only one parameter: the weighting α of the image gradient term in Eq. 4.8. The second pass is parameter free. The method does not smooth the solution

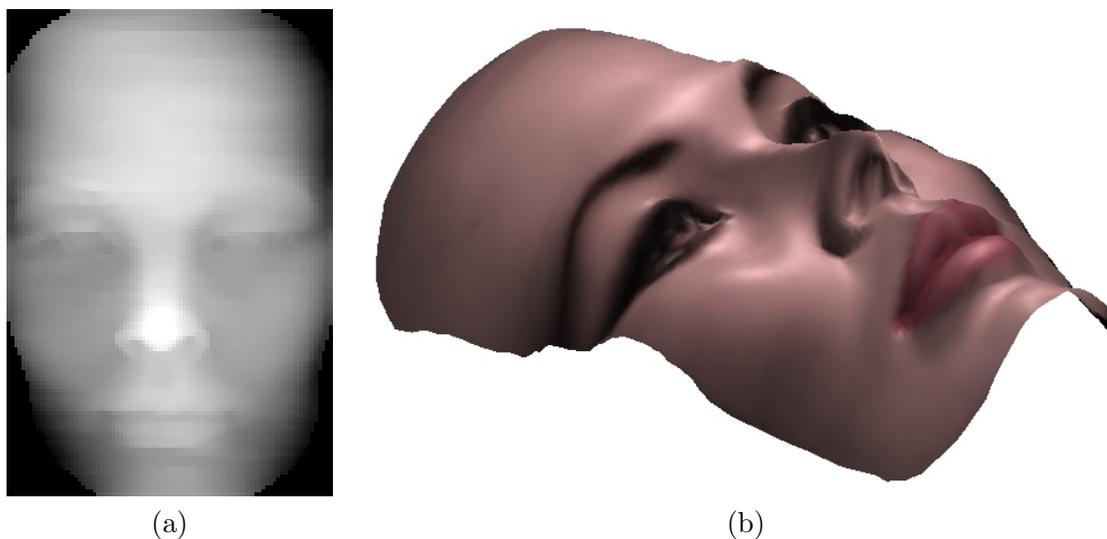


Figure 4.6: (a) Recovered depth map from the cyclopean viewpoint of the mannequin face in which light corresponds to near and dark to far. (b) Reconstructed surface, texture-mapped with one input image.

within epipolar planes, rather it chooses the solutions which together form the surface that is smoothest *across* epipolar planes.

Figures 4.5 display two views of a surface mesh reconstructed using the two-pass dynamic programming method. (The value $\alpha = 0.1$ was used.) Figure 4.6 shows a depth map in which light corresponds to near and dark to far, and the recovered surface texture-mapped with the left input image. Notice that the method is unhampered by the specularities and is able to both lock onto the features such the eyes, eyebrows, and lips, but also provide good reconstructions in textureless regions such as the forehead.

4.3 Summary

In this chapter, we look at the shape information that is available in a single reciprocal pair of images. We show that the reciprocity constraint is a PDE that, under suitable conditions, can be integrated along characteristic curves provided that initial conditions are known. We also show how regularization can be used to determine the initial conditions algorithmically, enabling a complete reconstruction of the surface from a single pair of images.

The results in this chapter demonstrate that, like its multinocular counterpart, binocular Helmholtz stereopsis allows the reconstruction of surfaces with arbitrary BRDFs, and unlike typical stereo techniques, it provides accurate surface shape in regions of constant brightness.

Binocular Helmholtz stereopsis requires minimal data. Only two images are required as compared to the 36 images used in the last chapter. By using only two images, binocular Helmholtz stereo methods will generally be faster, simpler, and cheaper to implement. As a result, they can be applied to a much broader range of applications.

Chapter 5

Stratified Helmholtz Stereopsis

In Chapter 3 we derived the reciprocity constraint (Eq. 3.4) that relates the radiance measurements at corresponding points in a reciprocal pair of images, and we demonstrated how this constraint can be used to estimate the shape of surfaces with arbitrary reflectance. Since the constraint depends on the acquisition geometry (i.e., the intrinsic and extrinsic camera parameters), direct application of this constraint requires that the acquisition geometry is accurately known.

In this chapter we examine the surface information that can be obtained from reciprocal pairs when we do not have knowledge of the acquisition geometry. This discussion is analogous to studies of numerous other surface reconstruction techniques (e.g., structure from motion, stereo and photometric stereo methods) that have been adapted to handle uncalibrated or weakly calibrated image and illumination equipment. For these conventional techniques, the vision community has established a precise understanding of what 3D information can be obtained under stratified levels of prior knowledge about the acquisition system. A stratified approach is useful because it tells us what can be obtained under a given calibration condition (e.g., projective, affine or metric reconstruction) and what assumptions are required in order to obtain more. This approach has been applied to the problems of binocular and multinocular stereo [30, 31], structure from motion [53, 98] and photometric stereo [40, 92]. This chapter presents a similar stratification of Helmholtz stereopsis. It is assumed that the reader is familiar with multiple-view geometry; the required

background can be found in a number of texts (e.g., [39]).

In the first section, we introduce an ‘uncalibrated’ reciprocity constraint that is derived from the multinocular reciprocity constraint of Eq. 3.6, but is different in that it requires little knowledge about the cameras and sources. This new constraint can be used, for example, to obtain a projective reconstruction in the uncalibrated case. Following this, we explore the distant-source case (as considered in Chapter 4) in which the cameras and sources are far from the object so that the cameras can be modeled as affine cameras. In this case, we show that it is possible to obtain a reconstruction of the surface and its field of surface normals up to an unknown affine transformation. Finally, we demonstrate how knowledge about the acquisition system can be used to upgrade this affine reconstruction to a metric one; both geometric and photometric pieces of information are considered.

5.1 An Uncalibrated Reciprocity Constraint

We begin this section with a brief review of the multinocular reciprocity constraint derived in Sect. 3.5.1. We then construct a new constraint that does not require knowledge of the acquisition system.

In Chapter 3 a wheel configuration was used to capture multiple reciprocal pairs using a single camera and a single light source. Here, we consider a different acquisition system consisting of M isotropic point light sources co-located at the camera centers of M pinhole cameras. (This can be accomplished in numerous ways, such as by swapping cameras and sources, or by using half-mirrors or beam-splitters as shown in Fig. 3.4. Co-location can also be approximated by placing each light source near a camera.) Images are acquired in the following fashion. Light source i is turned on while the other sources are turned off, and $M - 1$ images are acquired from all cameras but camera i . This process is repeated M times, each with a different source turned on, until $M(M - 1)$ images are acquired. These images are comprised of $M(M - 1)/2$ reciprocal image pairs, and Fig. 5.1 shows a set of these images for $M = 8$. In this figure, the vertical direction corresponds to camera position, the horizontal direction to source position; reciprocal pairs are in symmetric positions.

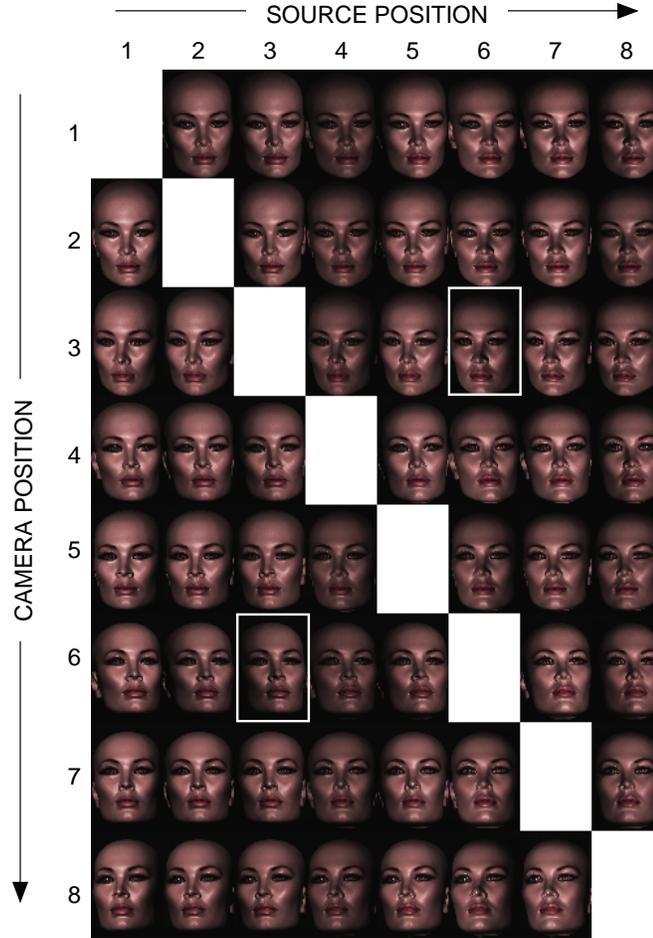


Figure 5.1: Reciprocal images from eight camera/source positions. Columns contain fixed illumination (stereo) images and rows contain fixed viewpoint (photometric stereo) images. One reciprocal pair is shown highlighted.

Given $M(M - 1)/2$ reciprocal pairs taken from camera/source positions $\mathbf{o}_1 \dots \mathbf{o}_M$ we have $M(M - 1)/2$ reciprocity constraints on the radiance measurements at corresponding image points. As in Eq. 3.4, we can write these constraints as

$$\left(e_{ij} \frac{s_i \hat{\mathbf{v}}_i^\top}{|\mathbf{o}_i - \mathbf{p}|^2} - e_{ji} \frac{s_j \hat{\mathbf{v}}_j^\top}{|\mathbf{o}_j - \mathbf{p}|^2} \right) \cdot \hat{\mathbf{n}} = 0, \quad (5.1)$$

where e_{ij} is the image irradiance at the projection of surface point \mathbf{p} (with surface normal $\hat{\mathbf{n}}$) in camera \mathbf{o}_i when illuminated by a point source at \mathbf{o}_j . The unit vector $\hat{\mathbf{v}}_i$ points in the direction from \mathbf{p} to \mathbf{o}_i . We have also included the relative source strengths s_i . (In previous chapters, all sources were assumed to be equal or of known strength—a non-trivial

assumption.)

As noted in Sect. 3.5.1, when the camera and source positions (and the source strengths) are known, everything inside the parentheses is determined by the point \mathbf{p} . By stacking the reciprocity constraints into rows of a matrix \mathbf{M} , we obtain the multinocular reciprocity constraint,

$$\mathbf{M}(\mathbf{p})\hat{\mathbf{n}} = \mathbf{0} \quad (5.2)$$

where $\mathbf{M} \in \mathbb{R}^{M(M-1)/2 \times 3}$, and \mathbf{M} is written as a function of \mathbf{p} to stress the fact that it is completely determined by a single 3D point.

This constraint is used in Chapter 3 to determine which points \mathbf{p} in a scene are surface points. In addition, for each established surface point, Eq. 5.2 provides an estimate of the surface normal, since for a surface point \mathbf{p} , the normal $\hat{\mathbf{n}}$ spans the 1D nullspace of $\mathbf{M}(\mathbf{p})$.

Reconstruction using Eq. 5.2 requires knowledge of the camera and source positions and the source strengths. It also requires that the radiometric camera responses are known or are linear and equal (since Eq. 5.1 involves scene radiance values and not pixel intensities), and that the M point sources are isotropic and uniform. By a simple rearrangement of Eq. 5.2, however, we can eliminate the need for most of these assumptions.

Making the definition

$$\tilde{w}_i = \frac{s_i \hat{\mathbf{v}}_i \cdot \hat{\mathbf{n}}}{|\mathbf{o}_i - \mathbf{p}|^2} \quad (5.3)$$

we can write Eq. 5.1 as $e_{ij}\tilde{w}_i - e_{ji}\tilde{w}_j = 0$, and Eq. 5.2 becomes

$$\mathbf{E}\tilde{\mathbf{w}}^\top = \mathbf{0} \quad (5.4)$$

where $\mathbf{E} \in \mathbb{R}^{(M(M-1)/2) \times M}$ and $\tilde{\mathbf{w}} \in \mathbb{R}^M$. For example, if $M = 4$, we have

$$\begin{bmatrix} e_{12} & -e_{21} & 0 & 0 \\ e_{13} & 0 & -e_{31} & 0 \\ e_{14} & 0 & 0 & -e_{41} \\ 0 & e_{23} & -e_{32} & 0 \\ 0 & e_{24} & 0 & -e_{42} \\ 0 & 0 & e_{34} & -e_{43} \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \tilde{w}_3 \\ \tilde{w}_4 \end{bmatrix} = \mathbf{0}.$$

Equation 5.4 is satisfied for visible points on a surface in the scene. Therefore, similar

to the use of Eq. 5.2, the *uncalibrated reciprocity constraint*,

$$\text{rank } \mathbf{E} < M, \tag{5.5}$$

can be used to recover shape information. When the epipolar geometry of the camera system is known (i.e., we have weak calibration), then \mathbf{E} will be a function of disparity, and a multinocular stereo search process can be performed using this constraint. When it is unknown, this constraint would have to be used within the context of a robust structure from motion algorithm (e.g., [2].) In either case, one obtains a set of corresponding points, and in turn, one can reconstruct the scene up to a projective transformation [30].

The key advantage of the uncalibrated reciprocity constraint (Eq. 5.5) is that it depends only on measured radiance values, so we do not need to know the positions of the cameras and sources, nor do we need to know the source strengths. All that we require is: 1) that the radiometric responses of the cameras are linear and equal (or are known); and 2) that the light sources are isotropic and uniform. Also, note that we do not require radiance measurements from all $M(M - 1)/2$ reciprocal pairs in order to use Eq. 5.4. We only require measurements from P pairs, where $P \geq M$.

In the remainder of this chapter, we will discuss a stratified reconstruction technique based on this constraint, even though it is a weaker correspondence constraint than that based on Eq. 5.2 (see Appendix A.) While this may seem disconcerting, note that the original reciprocity constraint is itself only a necessary condition for correspondence, and that the advantage of allowing an uncalibrated system is an important one.

5.2 Distant Sources

In the previous section, we derived a correspondence constraint that does not require knowledge of the cameras and sources. This suggests that it is possible to establish a projective reconstruction of scene points from uncalibrated reciprocal pairs. This is similar to conventional uncalibrated stereo [30], except that since we have carefully varied the lighting between views, we are able to use a constraint that is independent of reflectance (as opposed

to relying on brightness-constancy.)

In this section, we show that when the cameras and sources are far from the scene, we can quite easily obtain more: the reciprocal image pairs provide accessible information about both the surface normals and the light sources.

As shown in the binocular case of Chapter 4, when the cameras and sources are far from the scene, we can write Eq. 5.1 as

$$(e_{ij}\mathbf{s}_i^\top - e_{ji}\mathbf{s}_j^\top) \hat{\mathbf{n}} = 0, \quad (5.6)$$

where \mathbf{s}_i is a product of the effective source strength s_i and direction $\hat{\mathbf{s}}_i$, both of which are the same for all points \mathbf{p} in the scene. Accordingly, the vector $\tilde{\mathbf{w}}$ in Eq. 5.4 simplifies to

$$\mathbf{w} = [\mathbf{s}_1^\top \hat{\mathbf{n}} \quad \mathbf{s}_2^\top \hat{\mathbf{n}} \quad \cdots \quad \mathbf{s}_M^\top \hat{\mathbf{n}}]^\top. \quad (5.7)$$

Now, suppose that we have established correspondence for N points. That is, we have corresponding observations of N unknown scene points $\mathbf{X}_1 \dots \mathbf{X}_N \in \mathbb{R}^3$ in each of M viewpoints. (This could be achieved using the constraint $\text{rank } \mathbf{E} < M$, for example.) For a given point \mathbf{X}_k , we have $M(M-1)$ radiance measurements, one for each source/camera pair, and we can form a matrix \mathbf{E}_k for that point as in Eq. 5.4. Since these radiance measurements correspond to a single surface point, this matrix in general has rank $(M-1)$, and its 1D null space can be expressed as $c_k \mathbf{w}_k = c_k [\mathbf{s}_1^\top \hat{\mathbf{n}}_k \quad \mathbf{s}_2^\top \hat{\mathbf{n}}_k \quad \cdots \quad \mathbf{s}_M^\top \hat{\mathbf{n}}_k]^\top$, $c_k \in \mathbb{R}$. (It may be possible for the rank of \mathbf{E} to drop below $(M-1)$, but do not consider these cases here.) Letting \mathbf{W} denote the $N \times M$ matrix (recall that N is the number of points and M is the number of sources) whose rows are the transposed null vectors $c_1 \mathbf{w}_1^\top \dots c_N \mathbf{w}_N^\top$, we have

$$\mathbf{W} = \begin{bmatrix} c_1 \mathbf{s}_1^\top \hat{\mathbf{n}}_1 & c_1 \mathbf{s}_2^\top \hat{\mathbf{n}}_1 & \cdots & c_1 \mathbf{s}_M^\top \hat{\mathbf{n}}_1 \\ c_2 \mathbf{s}_1^\top \hat{\mathbf{n}}_2 & c_2 \mathbf{s}_2^\top \hat{\mathbf{n}}_2 & & \vdots \\ \vdots & & \ddots & \\ c_N \mathbf{s}_1^\top \hat{\mathbf{n}}_N & \cdots & & c_N \mathbf{s}_M^\top \hat{\mathbf{n}}_N \end{bmatrix}. \quad (5.8)$$

The matrix \mathbf{W} is of a familiar form; it is precisely this matrix that one considers in the bilinear calibration-estimation problem of uncalibrated photometric stereo [27, 40, 52]. The difference here is that the ‘albedo values’ are not due to surface reflectance properties, but

are instead artifacts of the homogeneity of Eq. 5.4. Indeed, the reflectance of the surface has been effectively removed through the use of reciprocal images. Each column of \mathbf{W} corresponds to a Lambertian image of the N points on the surface under a fixed source direction, and Fig. 5.2 shows what these images look like for dense points on a real surface.

In order to extract the surface normal and source information embedded in \mathbf{W} , we must find the decomposition $\mathbf{W} = \mathbf{B}\mathbf{S}$, where \mathbf{B} is the $N \times 3$ matrix of surface normals (each scaled by a constant c_k), and \mathbf{S} is the $3 \times M$ matrix of source vectors (source directions scaled by source strength.) A common strategy is to find a preliminary rank 3 factorization of \mathbf{W} using SVD, and then to correct that decomposition using additional constraints [40, 52]. That is, one computes $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and defines $\tilde{\mathbf{B}} = \mathbf{U}\mathbf{\Sigma}^{1/2}$ and $\tilde{\mathbf{S}} = \mathbf{\Sigma}^{1/2}\mathbf{V}^T$, keeping only the rows and columns of \mathbf{U} , $\mathbf{\Sigma}$, and \mathbf{V} corresponding to the first three singular values. (Here, it is assumed that $N \geq 3$, $M \geq 3$ and that not all normals or sources lie in a plane.) This decomposition is not unique (since $\tilde{\mathbf{B}}\mathbf{Q}^{-1}\mathbf{Q}\tilde{\mathbf{S}} = \tilde{\mathbf{B}}\tilde{\mathbf{S}}$ for all $\mathbf{Q} \in GL(3)$), and the true decomposition $\mathbf{W} = \mathbf{B}\mathbf{S}$ can be obtained by finding the matrix \mathbf{Q} that satisfies

$$\begin{aligned} \mathbf{B} &= \tilde{\mathbf{B}}\mathbf{Q}^{-1} \\ \mathbf{S} &= \mathbf{Q}\tilde{\mathbf{S}}. \end{aligned} \tag{5.9}$$

Note that \mathbf{Q} can only be defined up to scale, which is an expression of the fact that we can apply a global scaling to the source strengths (and the inverse scaling to the c_k 's) without affecting \mathbf{W} . Thus, \mathbf{Q} has eight degrees of freedom. We examine relevant methods to determine \mathbf{Q} in Sect. 5.3.

In the previous section, we derived a constraint that can be used to establish correspondence in the uncalibrated case. In the present section, we showed that in the case of distant sources, we can go further; we can use the available photometric information to estimate the surface normals at the points of observation as well as the strength and direction of the light sources. In making this statement, we are ignoring the problem of solving for \mathbf{Q} in Eq. 5.9, but notice that we have not yet used the available geometric information. Since the source directions are equivalent to the viewing directions, they can alternatively

be computed using established techniques of structure from motion. The interplay between the geometric and photometric constraints is discussed in the next two sections, leading to a number of ways in which we can establish a dense metric reconstruction in the uncalibrated case of distant-source Helmholtz stereopsis.

5.3 Affine Reconstruction

In this section we demonstrate that geometric information can be used to resolve the ambiguity in the calibration-estimation problem. As a result, for $N \geq 4$ observed points and $M \geq 4$ camera/source positions we can obtain a dense affine reconstruction of the scene (including surface normals and source strengths) without making any assumptions beyond those of the previous section.

When the cameras are far from the scene, they can be accurately described using the parallel projection model. Here, we use the most general such model—that of an affine camera [78]. To represent an affine camera, we use the tuple $[\mathbf{P}, \mathbf{t}]$ where $\mathbf{P} \in \mathbb{R}^{2 \times 3}$ and $\mathbf{t} \in \mathbb{R}^2$. In this notation, the image point $\mathbf{x}_k^i \in \mathbb{R}^2$ that results from projecting scene point $\mathbf{X}_k \in \mathbb{R}^3$ into image i is given by

$$\mathbf{x}_k^i = \mathbf{P}_i \mathbf{X}_k + \mathbf{t}_i. \quad (5.10)$$

The matrix \mathbf{P} can be decomposed as

$$\mathbf{P} = \begin{bmatrix} \alpha_x & s \\ 0 & \alpha_y \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \end{bmatrix} \quad (5.11)$$

where \mathbf{r}_1^\top and \mathbf{r}_2^\top are the first two rows of a rotation matrix that describes the camera orientation in the world coordinate system, s is the pixel skew, and α_x and α_y are the horizontal and vertical scale factors. (The aspect ratio is given by α_x/α_y .) Also, given a matrix \mathbf{P} , the viewing direction (and in the present case, the source direction) in the world coordinate system is given by the unit vector in the negative direction of the cross product of the two row vectors \mathbf{p}_1^\top and \mathbf{p}_2^\top . That is,

$$\hat{\mathbf{s}} = -\frac{\mathbf{p}_1^\top \times \mathbf{p}_2^\top}{|\mathbf{p}_1^\top \times \mathbf{p}_2^\top|}. \quad (5.12)$$

It is well known that by observing $N \geq 4$ non-coplanar rigid¹ points over two or more unknown affine views, one can establish an affine coordinate frame, and thereby obtain the scene points and the cameras up to an unknown affine transformation [53, 98]. In the present case, this reconstruction includes the (affine) source directions, since they are equivalent to the viewing directions. We can show that given a sufficient number of camera/source positions, knowledge of these directions enables a unique solution to the calibration-estimation problem of Sect. 5.2.

Each known source direction $\hat{\mathbf{s}}_i$ gives two linear constraints on the matrix \mathbf{Q} in Eq. 5.9, since for each we have $\hat{\mathbf{s}}_i = \alpha_i \mathbf{Q} \tilde{\mathbf{s}}_i$ for some $\alpha_i > 0$. As noted in Sect. 5.2, \mathbf{Q} is only defined up to scale, and thus has eight degrees of freedom. It follows that in general, \mathbf{Q} can be uniquely recovered given $M \geq 4$ camera/source positions by solving the corresponding constrained linear system of equations.

To summarize, given $N \geq 4$ observed points over $M \geq 4$ cameras/sources in general position, we can obtain the cameras (and source directions), the scene points, the source strengths, and the surface normals at the observed points. All of this information is in an affine coordinate frame. (See Fig. 5.3 for an example.) In Sect. 5.4, we discuss ways in which we can upgrade to a metric reconstruction.

Note that if we have fewer views (if $M = 3$) we can still establish correspondence using Eq. 5.4, and we can still establish an affine reconstruction of the cameras and the observed points. We cannot, however, determine the source strengths or the surface normals without further information.

5.4 Metric Reconstruction

The reconstruction obtained in the previous section differs from a metric reconstruction (i.e., up to scale) by an unknown affine transformation $\mathbf{A} \in GL(3)$. The problem of “upgrading” the reconstruction to a metric one is thus the problem of estimating the nine parameters of

¹The general affine camera model allows smooth non-rigid transformations of the scene points between views (see [53].) Here we assume fixed source positions, however, and in order to make use of the photometric constraints in our system, we require that the scene be rigid.

this transformation, commonly termed the autocalibration problem. In order to solve this problem, we require more information, either geometric or photometric, about the system.

Autocalibration is well studied, and numerous techniques exist for obtaining a metric reconstruction using constraints on the intrinsic parameters of the cameras (see [39].) For example, we can obtain a metric reconstruction if we know the aspect ratio of the cameras (e.g., the pixels are square) and there is no pixel skew (valid for CCD cameras.) With this knowledge in hand, metric reconstruction follows directly from the methods in Sect. 5.3 with no extra work, since in this case, structure from motion yields a metric reconstruction of the points and cameras [53]. Then, the Euclidean source directions can be used to resolve \mathbf{Q} , yielding the Euclidean surface normals and the relative source strengths. This is perhaps the most practical method, and the one we use in Sect. 5.5.

In addition to knowledge about that camera parameters, we can also make use of photometric information about the system. For example, knowledge of the relative strength of the light sources was used by Hayakawa [40] to partially resolve the ambiguity in the calibration-estimation problem for uncalibrated Lambertian photometric stereo. In that paper, it was shown that knowledge of the relative strength of six sources was enough to resolve the surface normals up to a unitary transformation [40, 92]. Similar analysis could be applied here. As another example, if the BRDF of the surface is known to be highly peaked in the specular direction, we can use detected specularities to constrain \mathbf{A} . If a specularity is detected at the projection of the scene point \mathbf{X} in the reciprocal images corresponding to camera/source directions $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{s}}_j$, it follows that the normal at that point must bisect these two directions, or $\hat{\mathbf{n}} = (\hat{\mathbf{s}}_i + \hat{\mathbf{s}}_j)/2$. Detected specularities have already been used to reduce ambiguities in photometric stereo and affine binocular stereo [27].

Any of these geometric or photometric constraints can be combined to determine the unknown affine transformation and thereby upgrade to a metric reconstruction. In the next section, we provide a demonstration of one method of resolving \mathbf{A} and obtaining a metric reconstruction.

5.5 Implementation and Results

For the results in this section, we acquired reciprocal pairs of images using a single 12-bit Kodak DCS760 camera (whose radiometric response function is linear) and multiple Lowel Pro-light sources with 250-watt halogen bulbs. M separate sources were placed at each of M camera/source positions which were far (~ 2.5 meters) from the scene, and the collection process was as follows. First, the camera replaced the source at position 1, and $M-1$ images were captured, each with illumination by one of the other sources. Then, the original source was returned to position 1, and the camera replaced the source at position 2. The process was repeated, resulting in a total of $M(M-1)$ images, or $M(M-1)/2$ reciprocal pairs. An example of a typical dataset for $M = 8$ is shown in Fig. 5.1.

5.5.1 Affine

The first step of the reconstruction is the establishment of an affine coordinate system. Currently, this is done by clicking a small number of points, although it could be achieved using a robust structure from motion technique (e.g., [2].) Given $N \geq 4$ correspondences over all M views, we compute the points $\mathbf{X}_1 \dots \mathbf{X}_N$ and cameras $[\mathbf{P}_1, \mathbf{t}_1], \dots, [\mathbf{P}_M, \mathbf{t}_M]$ using the Tomasi-Kanade factorization algorithm [106] generalized to affine cameras as in [98, 113]. The corresponding source directions $\hat{\mathbf{s}}_i$ are computed using Eq. 5.12.

Having established an affine coordinate system, we use the uncalibrated reciprocity constraint (Eq. 5.4) to establish dense correspondence in the multinocular system. This is accomplished by an exhaustive search as follows. The affine coordinate system is transformed so that the xy -plane is parallel to the image plane of the second camera in Fig. 5.1, meaning that $\hat{\mathbf{s}}_2 = [0 \ 0 \ -1]^\top$. We assume that the surface is the graph of a function $z(x, y)$, $(x, y) \in U$ where the domain U is known. (The domain was manually selected to eliminate the background from the reconstruction.) As in Sect. 3.5.3, the depth z is discretized so that $z \in \{z_{min}, \dots, z_{max}\}$. For each point $\mathbf{x} \in U$, and for a given depth z , we sample all $M(M-1)$ images at the projected points and build $\mathbf{E}_{\mathbf{x}}(z)$ using Eq. 5.4. As a measure of the likelihood that $\text{rank } \mathbf{E}_{\mathbf{x}}(z) = M-1$ we use the ratio of its two smallest

singular values,

$$r_{\mathbf{x}}(z) = \frac{\sigma_{M-1}}{\sigma_M}. \quad (5.13)$$

Since the constraint in Eq. 5.4 is necessary for a valid depth value but not sufficient, as we did in Sect. 3.5.3, we use simple windowing to make the depth search more robust. For a given point $\mathbf{x}_o \in U$ and a depth value z , we compute the ratio $r_{\mathbf{x}}(z)$ at this point as well as at points in a small window W_r around \mathbf{x}_o . Then, the estimated depth at \mathbf{x}_o is given by

$$z_{\mathbf{x}_o} = \arg \max_z \sum_{\mathbf{x} \in W_r} r_{\mathbf{x}}(z). \quad (5.14)$$

Once we have estimated the depth $z_{\mathbf{x}_o}$, the corresponding null vector $\mathbf{w}_{\mathbf{x}_o}$ can be obtained as the linear least-squares estimate,

$$\mathbf{w}_{\mathbf{x}_o} = \arg \min_{\mathbf{w}} \|\mathbf{E}_{\mathbf{x}_o}(z_{\mathbf{x}_o})\mathbf{w}\|^2, \quad \|\mathbf{w}\| = 1, \quad (5.15)$$

which is simply given by the right singular vector corresponding to the smallest singular value of $\mathbf{E}_{\mathbf{x}_o}(z_{\mathbf{x}_o})$. However, due to sampling noise, and noise in the estimation of the epipolar geometry and the depth estimate, the radiance measurements in $\mathbf{E}_{\mathbf{x}_o}(z_{\mathbf{x}_o})$ are in general noisy. This affects the accuracy of the estimated null vector $\mathbf{w}_{\mathbf{x}_o}$, especially when these measurements are made near discontinuities in scene radiance (e.g., at albedo edges.) In order to mitigate these effects, we use a second, weighted windowing scheme. Given the depth estimates at and near \mathbf{x}_o , we construct the \mathbf{E} matrices at \mathbf{x}_o and in a neighborhood $W_{\mathbf{E}}$ around it. The null vector at \mathbf{x}_o is estimated using

$$\mathbf{w}_{\mathbf{x}_o} = \arg \min_{\mathbf{w}} \sum_{\mathbf{x} \in W_{\mathbf{E}}} r_{\mathbf{x}} \|\mathbf{E}_{\mathbf{x}}\mathbf{w}\|^2, \quad \|\mathbf{w}\| = 1, \quad (5.16)$$

where $r_{\mathbf{x}}$ is given by Eq. 5.13. Again, this calculation can be done using SVD, since $\mathbf{w}_{\mathbf{x}_o}$ is simply the null vector of an augmented matrix created by stacking the weighted $\mathbf{E}_{\mathbf{x}}$ matrices vertically.

To summarize, we use a two-pass technique to establish depth and to estimate the null vector $\mathbf{w}_{\mathbf{x}}$ for each point $\mathbf{x} \in U$. In the first pass, the ratio in Eq. 5.13 is computed and stored for each \mathbf{x} and for each possible depth. By maximizing (with respect to depth) this



Figure 5.2: The columns of the \mathbf{W} matrix defined in Eq. 5.8 displayed as images. These correspond to images of the surface with the reflectance removed.

ratio summed over a window in U , we estimate the true depth associated with each point. In the second pass, we use these established depth estimates and the associated ratios to compute an estimate of the null vectors over a second weighted window. Note that these two steps need not be performed at the same resolution.

At this point, we have a null vector $\mathbf{w}_{\mathbf{x}}$ for every $\mathbf{x} \in U$, and we can construct the matrix \mathbf{W} as in Eq. 5.8. The columns of this matrix for the dataset in Fig. 5.1 are shown in Fig. 5.2. As mentioned previously, these columns of \mathbf{W} correspond to images of the surface with the reflectance removed. In order to recover the surface normals and the source strengths as described in Sect. 5.3, we compute a preliminary factorization of \mathbf{W} and then resolve the ambiguity in the factorization using the known affine source directions $\hat{\mathbf{s}}_i$. As a result of this procedure, we obtain the field of surface normals as shown in Fig. 5.3(a). If we treat the affine coordinate frame as though it were metric, and integrate this normal field (after enforcing integrability [33]), we obtain the result shown in the top row of Fig. 5.4.

For this result, we used all of the images shown in Fig. 5.1. Twenty correspondences were used to establish the affine coordinate system, the resolution of the reconstruction was

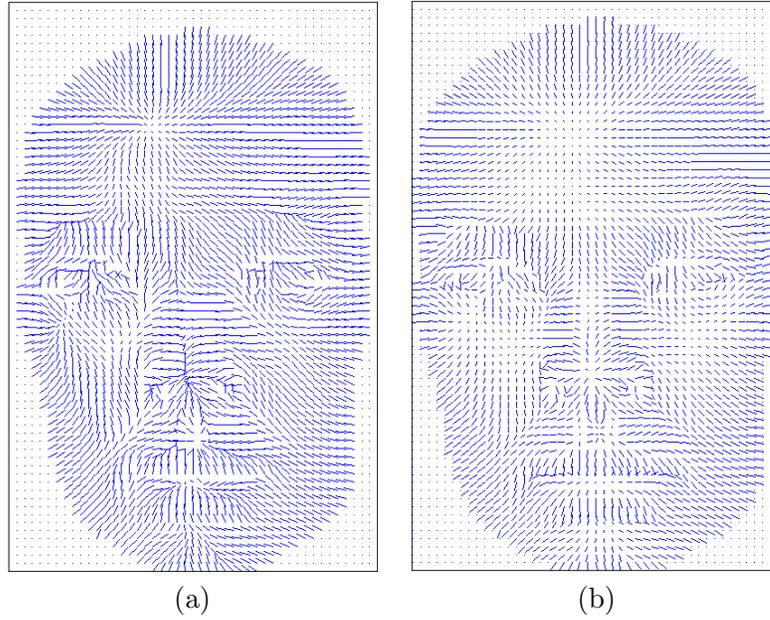


Figure 5.3: Surface normals seen from the viewpoint of the principle camera. (a) Normal field that results from the affine reconstruction as described in Sect. 5.3. (b) That which results from enforcing known aspect ratio and zero skew in all cameras.

153×105 , and we used square windows with $W_r = 9 \times 9$, and $W_E = 3 \times 3$.

5.5.2 Metric

In order to obtain a metric reconstruction, we make the assumption that the cameras have unit aspect ratio and zero skew, and the procedure of the previous section is repeated with only minor changes. We take the set of affine points \mathbf{X}_k and cameras $[\mathbf{P}_i, \mathbf{t}_i]$ and enforce known aspect ratio and zero skew by finding $\mathbf{G} \in GL(3)$ that satisfies the set of quadratic constraints [98, 113]

$$\begin{aligned} \mathbf{p}_{i1}^\top \mathbf{G} \mathbf{G}^\top \mathbf{p}_{i1} &= \mathbf{p}_{i2}^\top \mathbf{G} \mathbf{G}^\top \mathbf{p}_{i2} \\ \mathbf{p}_{i1}^\top \mathbf{G} \mathbf{G}^\top \mathbf{p}_{i2} &= 0, \end{aligned}$$

where \mathbf{p}_{i1}^\top and \mathbf{p}_{i2}^\top are the rows of the i^{th} camera matrix \mathbf{P}_i . By applying the transformation \mathbf{G} to the system (the cameras transform as $\mathbf{P}\mathbf{G}$ and the points as $\mathbf{G}^{-1}\mathbf{X}$), we obtain the points and cameras in a metric coordinate frame, and we can compute the true Euclidean source directions $\hat{\mathbf{s}}_i$.

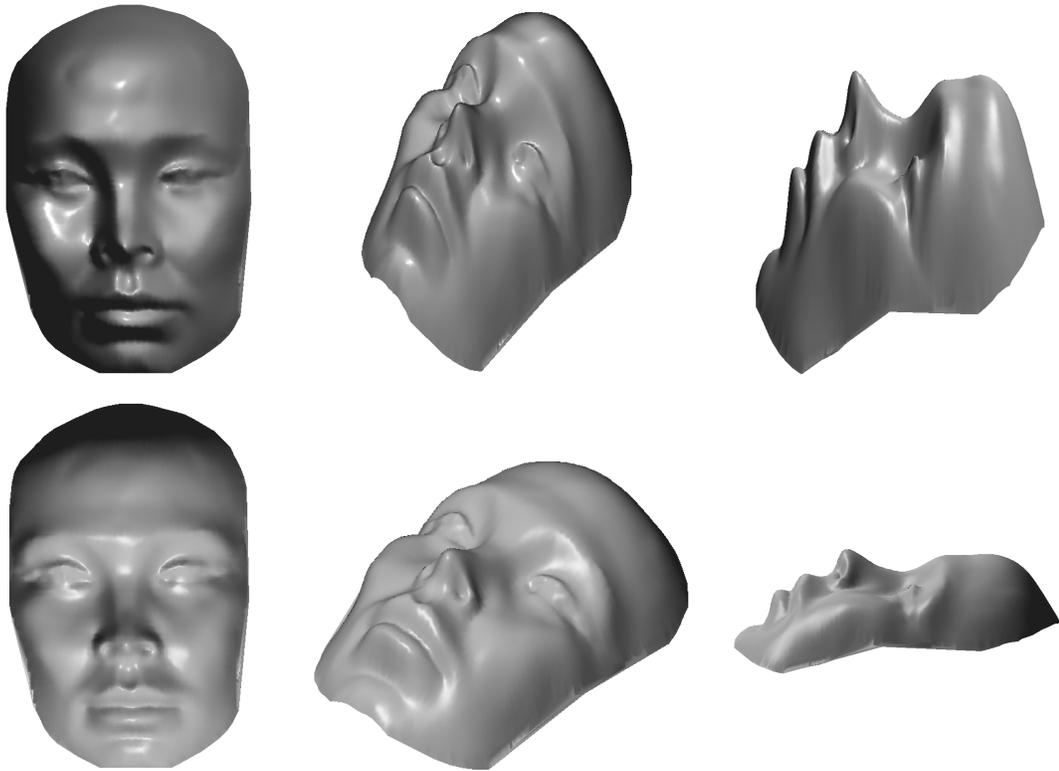


Figure 5.4: Three views of the two surfaces that result from integrating the normal fields in Fig. 5.3. The top row corresponds to the affine reconstruction (see Fig. 5.3(a)), and the bottom row the metric reconstruction (see Fig. 5.3(b)).



Figure 5.5: The same surface shown in the bottom row of Fig. 5.4, but interpolated and texture mapped with one input image.

Since the correspondences remain unchanged, the matrix \mathbf{W} need not be recomputed. We simply resolve the ambiguity in the factorization using the Euclidean source directions in place of the affine directions. The resulting normal field is shown in Fig. 5.3(b) and the integrated surface (after enforcing integrability [33]) is shown in the bottom row of Fig. 5.4, and in Fig. 5.5.

5.6 Summary

This chapter takes a stratified look at uncalibrated Helmholtz stereopsis (i.e., where we have little or no information about the cameras and light sources.) We present a new reciprocity constraint that can be used to establish correspondence without knowledge of the cameras and sources and thus obtain a projective reconstruction of the observed points. Like the original reciprocity constraint, this new uncalibrated constraint has the important property of not assuming a reflectance model for the surfaces in the scene being reconstructed.

We also show that in the distant-source case (i.e., when the distance from the scene to the cameras/sources is large), we can obtain an affine reconstruction that, in addition to the observed points, includes the surface normal at each observed point and the relative strength of the light sources. We discuss ways in which further information about the cameras and sources can be used to upgrade from this affine reconstruction to a metric reconstruction.

Chapter 6

Reflectance Sharing: Reflectance from a Sparse Set of Images

As discussed in the introduction to this dissertation, an image E is determined by the shape and reflectance properties of the surfaces in a scene as well as the scene illumination,

$$E = g(\text{shape}, \text{reflectance}, \text{illumination}).$$

The last three chapters addressed the problem of inverting the function g to recover the scene shape. The approach was to capture a set of images under controlled illumination such that shape and reflectance could be decoupled, and the shape could be recovered without making assumptions about reflectance.

In this chapter, we consider a different problem: assuming the shape is given (by Helmholtz stereopsis or by some other means) and the illumination is known, recover the reflectance of the surfaces in the scene. Recovering reflectance is difficult because of the high-dimensionality of the problem. At each point on the surface, the reflectance is described by the BRDF, and as discussed in Chapter 2, the BRDF is a four dimensional function of the view and lighting directions. It can vary sharply, especially when a surface is specular. In addition, the BRDF generally changes spatially over an object's surface. Without further assumptions, recovering the spatially-varying BRDF (or 6D SBRDF) requires an input set of images large enough to observe high-frequency radiometric events, such as sharp specular highlights, at each point on the surface. This set consists of a near

exhaustive sampling of images of the scene from all viewpoints and lighting directions, which can be tens-of-thousands of images or more.

In previous work, recovering spatial reflectance has been made tractable in one of three ways. Either: 1) the SBRDF is approximated using a parametric reflectance model, 2) only low-frequency effects are recovered, or 3) accurate non-parametric reflectance is recovered, but only for a subset of the reflectance function (e.g., for only one of view or lighting variation.) In this chapter we present an alternative approach. We exploit the fact that, although it is high-dimensional, the spatial reflectance function typically varies slowly over much of its spatial and angular domain. It varies slowly in the spatial dimensions since reflectance often varies smoothly across much of an object’s surface; and the fact that it varies slowly in certain angular dimensions is simply a re-statement of the observation that the reflectance at a single point—the BRDF—is highly compressible. By taking advantage of these properties, we develop a technique for recovering an accurate SBRDF from a sparse set of images without using a parametric model. In fact, as shown in Fig. 6.10, we can achieve a plausible result even for the extreme case of a single input image.

We approach SBRDF estimation as a scattered-data interpolation problem, with images providing dense 2D slices of data embedded in the higher dimensional SBRDF domain. To solve this interpolation problem we introduce a new representation of reflectance based on radial basis functions. This representation is compatible with many parameterizations of the angular domain (i.e., the BRDF domain), so we are free to choose one that exploits compressibility in these dimensions. In addition, by interpolating in both the spatial and angular dimensions, we exploit spatial coherence by *sharing* reflectance information across the surface.

6.1 Assumptions and Related Work

Although there has been substantial recent research in estimating spatial reflectance, this chapter represents an alternative approach to the problem, relying on a different set of assumptions. It is beneficial, therefore, to discuss these assumptions in the context of those

made by existing techniques.

Our method is directly applicable to acquisition systems that provide a set of images of an object with known geometry, viewpoint, and either point-source or directional illumination (see, e.g., [95, 24, 70]). Ignoring interreflections, each pixel in these images provides a sample of the high dimensional SBRDF, and our method works by interpolating and extrapolating from these samples to recover a continuous reflectance function. Although it is not explicitly considered here, in the case of directional illumination and orthographic projection, global effects such as sub-surface scattering and interreflection will be absorbed into our representation, so we expect our method to handle some of these effects. (This is similar to the *non-local reflectance field* defined by Debevec et al. [24].)

Our reflectance representation is based on radial basis functions, which provide a general scattered-data interpolation technique. This means that we do not assume the SBRDF can be represented by a parametric reflectance model. The only assumption we make about the SBRDF is that it is “smooth” in some sense. (This is discussed in more detail in Sect. 6.3.1 and in [28, 110].) It also means that, while we expect less accurate results far from the convex hull of the input samples (i.e., far from the set of input viewing and illumination directions), within this convex hull we expect accurate results for arbitrary reflectance functions. This is in sharp contrast to methods using parametric reflectance models, such as Lensch et al. [62] and McCallister et al. [73], who represent the SBRDF using the Lafortune model [57], Sato et al. [95] who use a simplified Torrance-Sparrow model, and Yu et al. [118] who use the Ward model. In many cases, parametric approaches can provide useful approximations from sparse input, even far from the convex hull of input samples, but they lack the flexibility to represent general reflectance functions with arbitrary accuracy.

The benefits of a data-driven, non-parametric approach have been demonstrated in previous work. For example, Wood et al. [115] use over 600 images of an object under fixed illumination to estimate the 2D view-dependent reflectance variation at each point, and Debevec et al. [24] use 2048 images to measure the 2D lighting-dependent variation

corresponding to fixed viewpoint.¹ These methods differ from ours in that they neither assume nor exploit spatial coherence, relying instead on an exhaustive sampling of (a subset of) the spatial reflectance function. They are not designed to work well for sparse input, and the large amount of data they require makes it difficult to extend them to include *both* light and view variation in a non-parametric way. For example, Matusik et al. [70] estimate a surface reflectance function (both view and lighting) on the visual hull of an object using more than 12,000 images. Even this large number of images provides only a sparse sampling of the BRDF at each point, and as a result, images can be synthesized only with low-frequency lighting.

Our method differs from previous non-parametric methods in that it exploits spatial coherence to reduce the number of required images. Our assumption that the SBRDF varies smoothly in the spatial dimensions is similar to assumptions made by previous parametric methods, that have assumed that specular parameters are the same across a surface [118] or can be estimated at a sparse set of points [95]. To properly exploit spatial coherence we require curved surfaces, since a planar surface with distant illumination and orthographic view provides no information about angular reflectance variation. (Note, however, that in these cases more angular reflectance information can be obtained using near-field illumination and perspective views.) Additionally, while we demonstrate how our method can handle rapid spatial variation in terms of a multiplicative albedo or diffuse texture, in cases where the shape of the BRDF itself changes rapidly, we currently assume these discontinuities to be given as input.

In addition to assuming spatial smoothness, our second main assumption is that spatial reflectance is highly compressible in the angular dimensions (i.e., in the BRDF domain.) This property has been previously exploited for 3D shape reconstruction [42], efficient BRDF acquisition [72], efficient rendering of BRDFs [75], and efficient evaluation of environment maps [13, 91]. In the present work, angular compressibility is exploited by assuming that

¹In [24], the method is extended to handle viewpoint variation through a special-purpose parametric model.

the BRDF typically varies rapidly only in certain dimensions (e.g., the half-angle.) Since these dimensions are sampled densely in a *single* image, only a small number of images are generally required for reflectance estimation. To accurately recover reflectance variation in other dimensions (e.g., that corresponding to Fresnel reflection,) we expect that additional images will be required. Since variation in these other dimensions is generally low in frequency, however, a sparse set of images usually suffices.

Other related work includes methods for compressing image-based representations, such as factoring them using eigen-textures [82, 16], but these generally require full, densely sampled reflectance information before compression can be applied. In addition, although they do not address the problem of spatially-varying reflectance, our work is directly related to, and inspired by, methods for image-based BRDF measurement [64, 68, 72]. Here, it has been observed that a single image of a curved, homogeneous surface represents a very dense sampling of a 2D slice of the global 3D or 4D BRDF.

6.2 Notation and BRDF Parameterization

At the core of our approach is the interpolation of scattered data in many (3-6) dimensions. The success of any interpolation technique depends heavily on how the SBRDF is parameterized. This section introduces some notation and presents one possible parameterization. Based on this parameterization, our interpolation technique is discussed in detail in Sects. 6.3 and 6.4.

The SBRDF is a function of six dimensions written $f(\mathbf{x}, \theta)$, where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ is the pair of spatial coordinates that parameterize the surface geometry (a surface point is written $\mathbf{s}(x, y)$), and $\theta \in \Omega \times \Omega$ are the angular coordinates that parameterize the double-hemisphere of view/illumination directions in a local coordinate frame defined on the tangent plane at a surface point (i.e., the BRDF domain.) As discussed in Sect. 2.1.3, a common parameterization of the BRDF domain is $\theta = (\theta_i, \phi_i, \theta_o, \phi_o)$ which represent the spherical coordinates of the light and view directions. When the BRDF is isotropic, the angular variation reduces to a function of three dimensions, commonly parameterized by $(\theta_i, \theta_o, \phi_o - \phi_i)$. In this work,

we restrict ourselves to this isotropic case and consider the SBRDF to be a function defined on a 5D domain. In the special case when the SBRDF is a constant function of the spatial dimensions (i.e., $f(\mathbf{x}, \theta) = f(\theta)$) we say that the surface is *homogeneous* and is described by a 3D function.

The angular dimensions (the BRDF domain) can be parameterized in a number of ways, and as discussed in Sect. 2.1.3 one good choice is Rusinkiewicz’s halfway/difference parameterization [93], shown in Fig. 6.1(b). Using this parameterization in the isotropic case, the BRDF is written as $\theta = (\theta_h, \phi_d, \theta_d) \subset [0, \frac{\pi}{2}) \times [0, \pi) \times [0, \frac{\pi}{2})$. (Note that ϕ_d is restricted to $[0, \pi)$ since $\phi_d \mapsto \phi_d + \pi$ by reciprocity.)

The existence of singularities at $\theta_h = 0$ and $\theta_d = 0$ and the required periodicity ($\phi_d \mapsto \phi_d + \pi$) make the standard halfway/difference parameterization unsuitable for most interpolation techniques. Instead, we define the mapping $(\theta_h, \phi_d, \theta_d) \mapsto (u, v, w)$, as

$$(u, v, w) = \left(\sin \theta_h \cos 2\phi_d, \sin \theta_h \sin 2\phi_d, \frac{2\theta_d}{\pi} \right). \quad (6.1)$$

This modified parameterization is shown in Fig. 6.1(c). The mapping eliminates the singularity that occurs at $\theta_h = 0$ and ensures that the BRDF $f(u, v, w)$ satisfies reciprocity. This mapping is undefined when $\theta_d = 0$ (i.e., when the light and view directions are equivalent), but this is not a severe limitation since that configuration is difficult to create in practice and can usually be avoided during synthesis.

6.2.1 Considerations for Image-based Acquisition

The halfway/difference parameterization of Rusinkiewicz has been shown to reduce sampling requirements and increase compression rates since common features such as specular and retro-reflective peaks are aligned with the coordinate axes [93, 72]. The modified parameterization of Eq. (6.1) maintains this property, since specular events cluster along the w -axis, and retro-reflective peaks occur in the plane $w = 0$.

These parameterizations are useful in image-based modeling for an additional reason: a single (orthographic, directional illumination) image of a homogeneous surface provides a dense sampling of the BRDF along a plane of constant θ_d (or equivalently, of constant

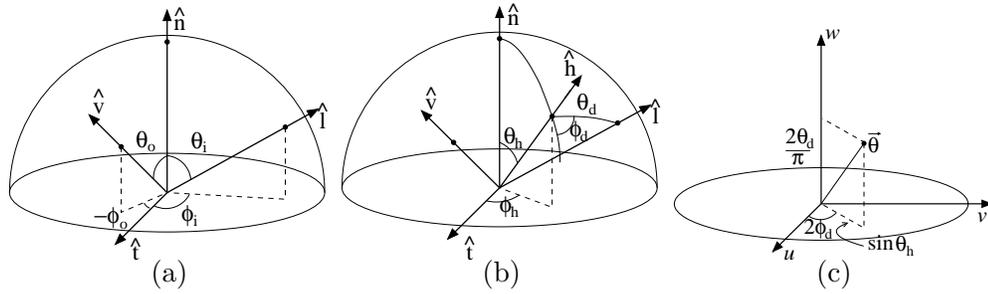


Figure 6.1: (a) The input/output parameterization of the BRDF domain defined in a local coordinate system defined by the surface normal and tangent vector. (b) The halfway/difference parameterization of Rusinkiewicz. In the isotropic case, the BRDF domain is parametrized by $(\theta_h, \phi_d, \theta_d)$. (c) The modified parameterization defined by Eq. (6.1) that is suitable for interpolation. The BRDF $f(u, v, w)$ is guaranteed to satisfy reciprocity; the parameterization is defined for all values of θ_h ; BRDF samples from a single image of a homogeneous curved surface lie on planes of constant w ; and specular events are clustered near the w -axis, enabling significant compression.

w .) Thus, these parameterizations effectively separate the dimensions of the BRDF that can be sampled densely using image based methods (θ_h and ϕ_d) from that which is sampled sparsely (θ_d). This is clear from Fig. 6.1(b). In a single image, θ_d is constant for all surface points, independent of the surface normal $\hat{\mathbf{n}}$. As a result, we have only as many samples of this dimension as we have images. In contrast, when a surface is curved, a single image provides a nearly continuous sampling of θ_h and ϕ_d .

Conveniently, the asymmetric sampling obtained from image-based data corresponds well with the behavior of general BRDFs, which vary slowly in the sparsely sampled dimension θ_d , especially when θ_d is small. By imaging curved surfaces, we ensure that the sampling of the half-angle θ_h is high enough to accurately recover the high-frequency variation that is generally observed in that dimension.

It is worth emphasizing that our interpolation method can be implemented using a number of parameterizations, with the halfway/difference parameterization being one of many possible choices. Another suitable parameterization that may prove useful, for example, is the Marschner parameterization described in Sect. 2.1.3.

6.3 Homogeneous Surfaces

With this notation in hand, we can proceed to introduce our reflectance representation and our method for sharing reflectance. Recall that our goal is to estimate a continuous SBRDF $f(\mathbf{x}, \theta)$ from a set of samples $\{f_i\}$ drawn from images of a surface with known geometry. In this section, we introduce the method using the special case of homogeneous surfaces, where we seek to estimate a single global BRDF that is not spatially-varying. This special case is considered for two reasons. First, the homogeneous case provides a convenient platform to evaluate the performance of our method. Second, although it is not the focus of this paper, the application of our method to the homogeneous case provides an alternative and potentially useful representation for image-based BRDF measurement [64, 68, 72], that is well-suited for sparse datasets. The inhomogeneous case is discussed in the next section.

In order to recover a continuous representation of the BRDF from scattered, image-based samples, the BRDF can be interpolated using local polynomial regression [67], or expressed as a linear expansion of pre-chosen basis functions such as spherical harmonics, wavelets, Zernike polynomials, or the data-driven basis functions of Matusik et al. [72]. Here, we choose an alternative representation, expressing the BRDF as a linear combination of radial basis functions (RBFs). Radial basis functions have been used successfully for scattered data interpolation problems in many different contexts, including 3D reconstruction [14, 26]; and there is a growing body of research investigating convergence rates and methods for efficient computation [12]. They are chosen here for three main reasons.

1. The coefficients of an RBF interpolant are estimated directly from scattered data with few restrictions on the sample locations (e.g., without requiring samples on a regular grid.) This is an important property that is not shared by methods such as spherical harmonics and spherical wavelets. Indeed, since the sampling rate available in image-based data is highly irregular, use of these other basis functions often requires a preprocessing step to resample the data at regular intervals.
2. Essential to this work is the fact that RBF interpolation can easily be extended to the

spatially-varying case. The cost of computing the coefficients of an RBF interpolant is dimension-independent, so the RBF interpolation methods we develop here for the homogeneous BRDF case (a 3D domain) can be easily adapted for the spatially-varying case (a 5D domain). In contrast, spherical harmonics and Zernike polynomials are defined only in the angular domain, and could not easily be used to model spatial variation.

3. As has been shown by Carr et al. [14] an RBF representation can be compact, providing substantial compression of the input data. This is true for BRDF (and SBRDF) data provided a suitable parameterization is chosen.

6.3.1 Radial Basis Functions

To briefly review RBFs, consider a general function $g(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ from which we have N samples $\{g_i\}$ at sample points $\{\mathbf{x}_i\}$. This function is approximated as a sum of a low-order polynomial and a set of scaled, radially symmetric basis functions centered at the sample points;

$$g(\mathbf{x}) \approx \tilde{g}(\mathbf{x}) = p(\mathbf{x}) + \sum_{i=1}^N \lambda_i \psi(\|\mathbf{x} - \mathbf{x}_i\|), \quad (6.2)$$

where $p(\mathbf{x})$ is a polynomial of degree at most m , $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a continuous function, and $\|\cdot\|$ is the Euclidean norm. The sample points \mathbf{x}_i are referred to as *centers*, and the RBF interpolant \tilde{g} satisfies the interpolation conditions $\tilde{g}(\mathbf{x}_i) = g(\mathbf{x}_i)$.

Given a choice of m , an RBF ψ , and a basis for the polynomials of order m or less, the coefficients of the interpolant are simply determined as the solution of the linear system

$$\begin{bmatrix} \Psi & P \\ P^\top & 0 \end{bmatrix} \begin{bmatrix} \vec{\lambda} \\ \vec{c} \end{bmatrix} = \begin{bmatrix} \vec{g} \\ 0 \end{bmatrix}, \quad (6.3)$$

where $\Psi_{ij} = \psi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $\vec{\lambda}_i = \lambda_i$, $\vec{g}_i = g_i$, $P_{ij} = p_j(\mathbf{x}_i)$ where $\{p_j\}$ are the polynomial basis functions, and $\vec{c}_i = c_i$ are the coefficients in this basis of the polynomial term in \tilde{g} . This system is invertible (and the RBF interpolant is uniquely determined) in arbitrary dimensions for many choices of ψ , with only mild conditions on m and the locations of the data points [28, 76]. For example, in two dimensions the familiar “thin-plate spline”

corresponds to the RBF $\psi(r) = r^2 \log r$, and the corresponding system is invertible whenever $m \geq 1$ and the sample points are not collinear.

In the case of homogeneous BRDF data, the function being interpolated is a function of three dimensions (i.e., $d = 3$). A good choice in this case is the linear (or biharmonic) RBF, $\psi(r) = r$, with $m = 1$. Using this RBF, the resulting interpolant \tilde{g} minimizes a generalization of the thin-plate energy and is the “smoothest” in some sense [28]. In Sect. 6.4 we show that the linear RBF performs well in the spatially-varying case (in which $d = 5$) as well.

In practical cases the samples $\{g_i\}$ are affected by noise, and it is desirable to allow the interpolant to deviate from the data points, balancing the smoothness of the interpolant with its fidelity to the data. This is accomplished by replacing Ψ in Eq. (6.3) with $\Psi - \rho \mathbf{N}\mathbf{I}$, where \mathbf{I} is the identity matrix and ρ is a stiffness parameter. Further details are presented in [110].

So far we have dealt exclusively with isotropic radial functions. In many cases we benefit from using *anisotropic* radial functions in which the basis functions are ‘stretched’ in certain directions. Anisotropic RBFs were used by Dinh et al. [26] to accurately represent corners and edges in 3D surfaces. Here we use them to manage the asymmetry in our sampling pattern. (Recall from Fig. 6.1 that the dimensions u and v are sampled almost continuously while we have only as many samples of the w dimension as we have images.) Following Dinh et al. [26], an anisotropic radial function is created by scaling the Euclidean distance function in Eq. (6.2) so that the basis functions become

$$\psi(\|M(\mathbf{x} - \mathbf{x}_i)\|), \quad (6.4)$$

where $M \in \mathbb{R}^{d \times d}$. In our case we choose $M = \text{diag}(1, 1, s_d)$ to reflect the varying sampling rates in the three dimensions. For $s_d < 1$, the basis functions are elongated in the w dimension, which is appropriate since our sampling rate is much lower in that dimension. The appropriate value of this parameter depends on the angular density of the input images, and empirically we have found that typical values for s_d are between 0.1 and 0.5.

As a final note, when the number of samples is large (i.e., $N > 10\,000$), solving Eq. (6.3) requires care and can be difficult (or impossible) using direct methods. This limitation has been addressed quite recently, and iterative fitting methods [3], and fast multipole methods (FMMs) for efficient evaluation [4] have been developed for many choices of ψ in many dimensions. In some cases, solutions for systems with over half a million centers have been reported [14]. The next section includes an investigation of the number of RBF centers required to accurately represent image-based BRDF data, and we find this number to be sufficiently small to allow the use of direct methods.

6.3.2 Radial Basis Function BRDFs

Using the linear radial basis function with $m = 1$, and using the modified halfway/difference parameterization from Eq. (6.1), the BRDF is expressed as

$$\tilde{f}(\theta) = c_1 + c_2u + c_3v + c_4w + \sum_{i=1}^N \lambda_i \|\theta - \theta_i\|, \quad (6.5)$$

where $\theta_i = (u_i, v_i, w_i)$ represent the BRDF sample points from the input images, and $\vec{\lambda}$ and \vec{c} are found by solving Eq. (6.3).

As a practical consideration, since each pixel in a set of images represents a sample point θ_i , even with modest resolution in the images, using all available samples as RBF centers in Eq. (6.5) is computationally prohibitive. Much of this data is redundant, however, and an accurate representation of the BRDF can be achieved using only a small fraction of these centers. Carr et al. [14] use a simple greedy algorithm that selects which of the input samples are used as centers for the RBF, and a slightly modified version of the same algorithm can be applied here. The procedure begins by selecting a small subset of the sample points θ_i and fitting an RBF interpolant to these. Next, this interpolant is evaluated at all sample points and used to compute the radiance residuals, $\varepsilon_i = (f_i - \tilde{f}(\theta_i)) \cos \theta_i$, where θ_i is the angle between the surface normal at the sample point and the illumination direction. Finally, points where ε_i is large are appended as additional RBF centers, and the process is repeated until the desired fitting accuracy is achieved.

6.3.3 Evaluation

Equation (6.5) is an alternative representation for the BRDF. In order to test this model, we performed numerical experiments simulating the data available from image-based BRDF measurement systems such as that of Matusik et al. [72]. For these simulations, we used both specular and diffuse reflectance, one drawn from measured data (the metallic-blue BRDF, courtesy of Matusik et al. [72]), and the other generated using a physically-based BRDF model (Oren-Nayar [84]).

Figure 6.2 provides an indication of the number of centers required to accurately represent a BRDF. The BRDF model in Eq. (6.5) was fit to ten images of a homogeneous sphere (with the measured metallic-blue BRDF), and the accuracy of this fit was measured by its ability to predict the appearance of the sphere under novel view and lighting conditions. The figure shows the accuracy of our recovered BRDF representation as the number of centers is increased using the greedy algorithm. Convergence with less than 400 centers suggests that only a small fraction of the available centers are required to accurately summarize the reflectance information available in the input images. For this experiment, the ten input images were uniformly spaced in θ_d over the range $[0, \frac{\pi}{2}]$, and the resolution of each image was 100×100 . The relative RMS radiance error was computed using images of the sphere for 21 view and light directions (also uniformly spaced in θ_d) that were not used as input. Images predicted by the model were compared to the true images, and the error was summed over all pixels corresponding to surface points.

The example in Fig. 6.2 demonstrates the efficiency of this representation. For inhomogeneous surfaces, we obviously need more centers to represent spatial variation, but even in that case, accurate results can be obtained with fewer than 2000 centers. (See Fig. 6.5.) This suggests that this representation may be useful for compression of large image-based datasets.

In a second set of experiments, we explored the accuracy of the BRDF model in Eq. (6.5) as the number of input images increases. The results are shown in Figs. 6.3 and 6.4 for predominantly specular and diffuse reflectance, respectively. These results demonstrate that:

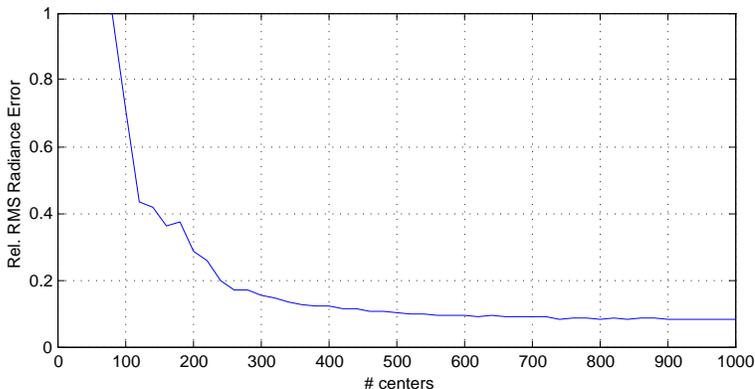


Figure 6.2: The accuracy of the RBF representation as the number of centers is increased using the greedy algorithm described in Sect. 6.3.2. The input is 10 images of a sphere synthesized using the metallic-blue BRDF measured by Matusik et al.

1) the RBF representation provides a reasonable fit from a small number of images, and
 2) the accuracy of this fit improves as the number of images is increased. The convergence of the fit is expected since, like other data-driven representations, the RBF representation is free to use the samples from additional images to flexibly represent the subtleties of the material. In contrast, while parametric models yield a more compact representation, these figures show that they can be too restricted to accurately fit arbitrary BRDF data. (This is true even as we increase the number of generalized cosine lobes in the Lafortune representation; additional lobes do not significantly improve accuracy, and furthermore, they cannot be reliably fit to a small number of images.) The bottom of each of these figures shows synthetic spheres rendered for novel view and light directions using the 1000-center RBF and parametric BRDF models recovered from 16 input images. The bottom of Fig. 6.3 demonstrates how the flexibility of the RBF representation enables the accurate representation of the BRDF at grazing angles, whereas behavior at even moderate grazing angles is not captured by the parametric models.

6.4 Inhomogeneous Surfaces

In the inhomogeneous case, the reflectance function varies in the spatial dimensions in addition to the angular dimensions, and our goal is to estimate the 5D SBRDF. An important

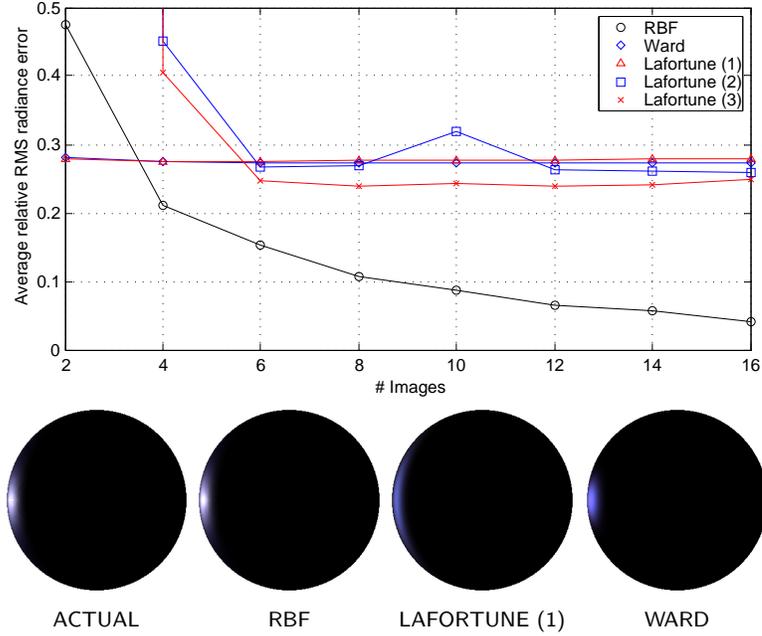


Figure 6.3: Top: Error in the estimated BRDF for an increasing number of input images of a sphere with the specular metallic-blue BRDF measured by Matusik et al. As the number of images increases, the RBF representation converges to the true BRDF. For comparison, the isotropic Ward model and the Lafortune model with one, two and three lobes (in addition to a Lambertian diffuse lobe) were fit to the same data, and these were found to be too restricted to provide an accurate fit. (Note: due to the difficulty in fitting multi-lobe Lafortune representations, for each set of images the Lafortune model was fit ten times using LM-iteration with the first and third lobes randomly initialized to $C_x = C_y = -1 \pm 0.25, C_z = 1 \pm 0.25, n = 20 \pm 5$ (forward scattering) and the second to $C_x = C_y = C_z = 1 \pm 0.25, n = 20 \pm 5$ (backward scattering). The graph shows the average error over the ten trials.) Bottom: Synthesized images using the three BRDF representations estimated from 16 input images. (The RBF representation uses 1000 centers.) The angle between the source and view directions is 140° . Only the RBF representation accurately captures the effects at this moderate grazing angle.

advantage of the RBF representation developed in the previous section is that it extends very naturally to the inhomogeneous case, with a unified representation of both spatial and angular variation. This representation allows us to share reflectance information spatially across the surface, and we demonstrate how this spatial sharing can reduce the number of required images by one or two orders of magnitude. We begin by assuming that the 5D SBRDF varies smoothly in the spatial dimensions. In the next section, we show how this can be generalized to handle rapid spatial variation in terms of a multiplicative albedo or

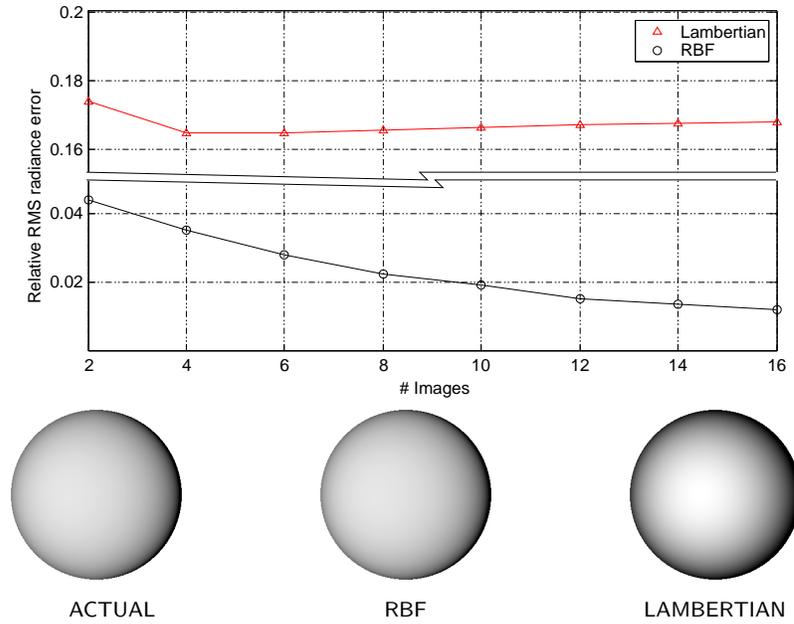


Figure 6.4: Top: Plot comparing the error in the estimated BRDF for an increasing number of input images of a sphere synthesized with diffuse Oren-Nayar reflectance. As the number of images increases, the RBF representation converges to the true BRDF. When the diffuse BRDF is modeled as Lambertian, of course this is not the case. Bottom: Synthesized images comparing the 1000-center RBF and Lambertian models estimated from 16 input images. The angle between the source and view directions is 10° .

diffuse texture.

6.4.1 Radial Basis Function SBRDFs

Letting $\mathbf{q} = (x, y, u, v, w)$ be a point in its domain, the SBRDF is represented using the linear RBF with $m = 1$, giving

$$\tilde{f}(\mathbf{q}) = p(\mathbf{q}) + \sum_{i=1}^N \lambda_i \|\mathbf{q} - \mathbf{q}_i\|, \quad (6.6)$$

where $p(\mathbf{q}) = c_1 + c_2x + c_3y + c_4u + c_5v + c_6w$. We can use any parameterization of the surface \mathbf{s} , and there has been a significant amount of recent work on determining good parameterizations for general surfaces [61, 38]. The ideal surface parameterization is one that preserves distances, meaning that $\|\mathbf{x}_1 - \mathbf{x}_2\|^{\frac{1}{2}}$ is equivalent to the geodesic distance between $\mathbf{s}(\mathbf{x}_1)$ and $\mathbf{s}(\mathbf{x}_2)$. For simplicity, here we treat the surface as the graph of a function, so that $\mathbf{s}(x, y) = (x, y, s(x, y))$, $(x, y) \subset [0, 1] \times [0, 1]$.

The procedure for recovering the parameters in Eq. (6.6) is almost exactly the same as in the homogeneous case. The coefficients of \tilde{f} are found by solving Eq. (6.3) using a subset of the SBRDF samples available in the input images, and this subset is chosen using the greedy algorithm. Anisotropic basis functions are realized using $M = (s_x, s_x, 1, 1, s_d)$, where s_x is chosen to reflect the expected relative rates of variation in the spatial and angular dimensions. The value of s_x depends on the choice of surface parameterization, and we have found typical values to be between 0.2 and 0.4.

6.4.2 Evaluation

The SBRDF representation of Eq. (6.6) was evaluated using experiments similar to those used in Sect. 6.3.3 for the homogeneous case. The SBRDF is fit to synthetic images of a hemisphere rendered using the physically-based Torrance-Sparrow model with linearly varying roughness parameter. Five images of the hemisphere are shown in the top of Fig. 6.6, and they demonstrate how the specular highlight sharpens as we move from left to right across the surface.

Figure 6.5 shows the accuracy of the recovered SBRDF as a function of the number of RBF centers when the SBRDF is fit to ten images of the hemisphere with fixed, orthographic viewpoint and ten uniformly distributed illumination directions. It is clear from this figure that fewer than 2000 centers are required to accurately represent the spatially-varying reflectance information available in the input images. This is a very compact representation, and it does not grow significantly as more input images are added. For this experiment, the RMS error was computed using images synthesized with 40 novel illumination directions, also uniformly distributed over the hemisphere.

Figure 6.6 demonstrates that reflectance sharing can drastically reduce the number of required input images. In this experiment, the SBRDF is fit to five images of the hemisphere (shown in the top of the figure.) Even with these sparse images, we can accurately predict the appearance of the hemisphere under novel illumination directions as shown in the middle of the figure. For comparison, the right column shows the result obtained without exploiting

spatial coherence, i.e., by interpolating only in the angular dimensions to estimate a separate BRDF at each point. (This is the technique used by Matusik et al. [70], and is similar in spirit to [115, 24], since these methods also estimate a unique reflectance function at each point.) Interpolating only in the angular dimensions means that at most five reflectance samples are available at each point. As a result, severe aliasing occurs when the surface is re-lit with a high frequency illumination environment like the directional illumination shown here. The bottom of Fig. 6.6 demonstrates that even 150 images are not enough to accurately estimate the SBRDF without exploiting spatial coherence. Thus, reflectance sharing effectively reduces the number of required input images by one or two orders of magnitude—from over 150 to only 5. (The accompanying video includes animated sequences of these results.)

Even though all of the input images are captured from the same viewpoint in this experiment, a full SBRDF is recovered, and as shown in Figs. 6.7(a)–(b), view-dependent effects can be accurately predicted. This is made possible by the spatial sharing of information (since each point on the surface is observed from a unique view in its local coordinate frame) and by reciprocity (since we effectively have observations in which the view and light directions have been exchanged.) As shown in Fig. 6.7(c), however, reflectance effects far from the convex hull of input samples cannot be accurately recovered, since these effects are not observed in the input images. (Figure 6.7(c) shows the actual and predicted appearance with $\theta_d = 65^\circ$; the five input images were captured with $\theta_d < 30^\circ$.)

6.5 Generalized Spatial Variation

The previous sections have introduced a simple, compact representation for reflectance and described how it can be used to accurately estimate an SBRDF from a sparse set of images when the SBRDF varies smoothly over the surface. In this section, we soften the requirement for spatial smoothness and demonstrate the applicability of our approach to image-based rendering of a human face.

Rapid spatial variation can be handled using a multiplicative albedo or texture by

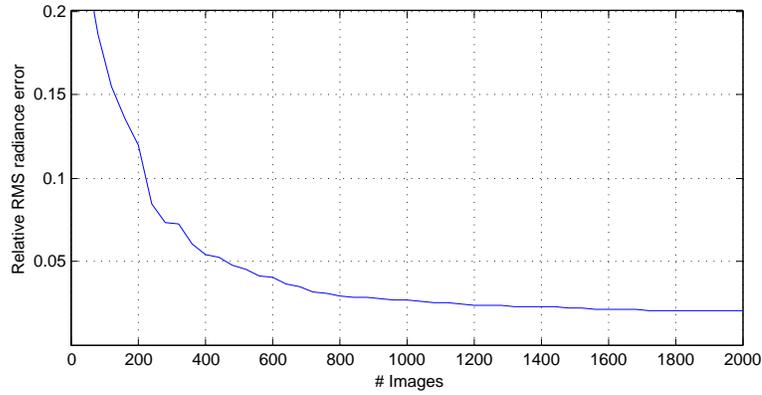


Figure 6.5: The accuracy of the RBF representation for spatially-varying reflectance as the number of centers is increased using the greedy algorithm described in Sect. 6.3.2. The input is 10 images of a hemisphere synthesized using a Torrance-Sparrow reflectance model with a linearly varying roughness parameter. Images of the sphere are shown at the top of Fig. 6.6.

writing the SBRDF as

$$f(\mathbf{x}, \theta) = a(\mathbf{x})d(\mathbf{x}, \theta),$$

where $a(\mathbf{x})$ is an albedo map for the surface and $d(\mathbf{x}, \theta)$ is a smooth function of five dimensions. As an example, consider the human face in Fig. 6.8(a). The function $a(\mathbf{x})$ accounts for rapid spatial variation caused by pigment changes in the skin, while $d(\mathbf{x}, \theta)$ models the smooth spatial variation that occurs as we transition from a region where skin hangs loosely (e.g., the cheek) to where it is taut (e.g., the nose.)

In some cases, it is advantageous to express the SBRDF as a linear combination of multiple 5D functions. For example, Sato et al. [95] and many others use the dichromatic model of reflectance [97] in which the BRDF at a point is written as the sum of an RGB diffuse component and a scalar specular component that multiplies the source colour. We employ a modified dichromatic model for the example in Fig. 6.9. Using this model, the SBRDF is given by

$$f_k(\mathbf{x}, \theta) = a_k(\mathbf{x})d_k(\mathbf{x}, \theta) + \hat{\mathbf{s}}_k g(\mathbf{x}, \theta), \quad k = R, G, B, \quad (6.7)$$

where $\hat{\mathbf{s}}$ is an RGB unit vector that describes the colour of the light source. In Eq. (6.7), a single function g is used to model the specular reflectance component, while each colour

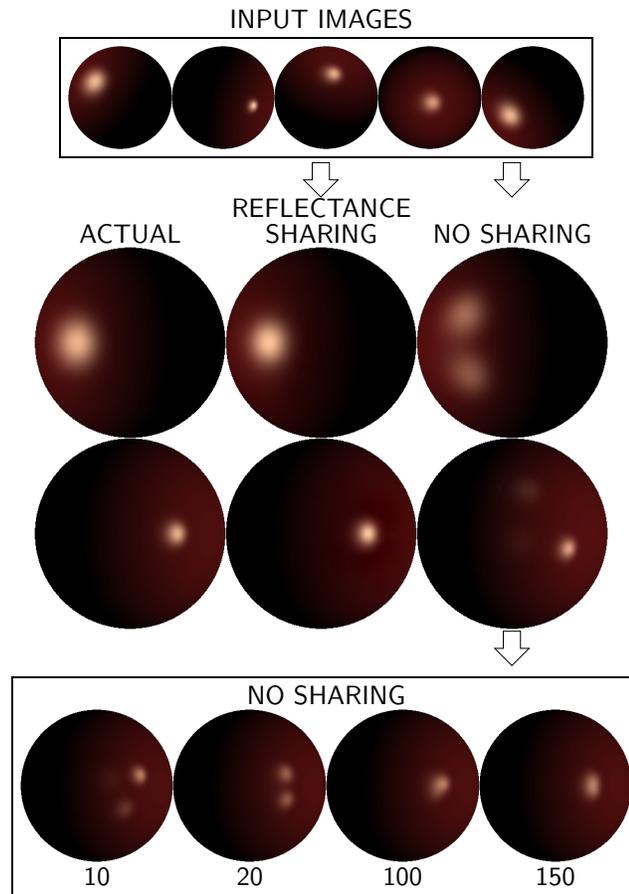


Figure 6.6: Estimating the spatially-varying reflectance function from a sparse set of images. Top: Five images of a hemisphere with spatially varying reflectance used as input. Middle: By sharing reflectance information spatially, we can recover the SBRDF from this sparse data. Severe aliasing occurs, however, if we do not share spatially, interpolating instead only in the angular dimensions and estimating a BRDF separately at each point. Bottom: Angular-only interpolation applied to an increasing number of input images. At least 150 images are required to obtain a result comparable to the reflectance sharing result with only five images.

channel of the diffuse component is modeled separately. This is significantly more general than the usual assumption of a Lambertian diffuse component, and it accounts for changes in diffuse colour as a function of θ , such as the desaturation of the diffuse component of skin at large grazing angles witnessed by Debevec et al. [24].

Finally, although not used in our examples, more general spatial variation can be modeled by dividing the surface into a finite number of regions, where each region has spatial

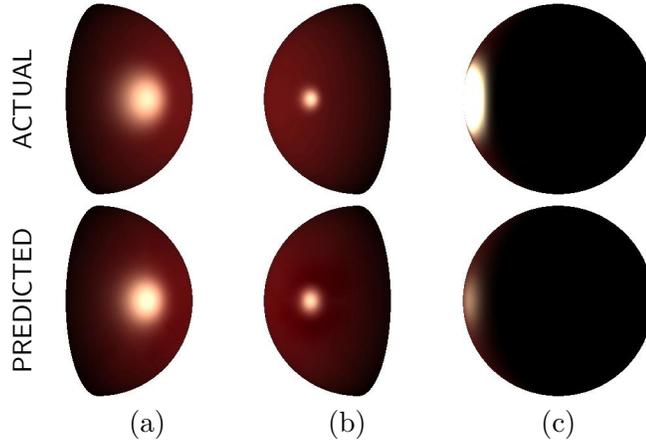


Figure 6.7: (a,b) Actual and predicted appearance of the hemisphere for two novel view-points. Even though the input (top of Fig. 6.6) is captured from a single view, a full SBRDF is recovered, including view-dependent reflectance effects. (c) Actual and predicted appearance far from the convex hull of input samples. The effects at grazing angles (here, $\theta_d = 65^\circ$) cannot be accurately predicted, since these effects are not observed in the input images.

reflectance as described above. Given a set of disjoint regions R_j , $j = 1, \dots, M$ that tile the spatial domain, the SBRDF is expressed as

$$f(\mathbf{x}, \theta) = \sum_{j=1}^M \delta_j(\mathbf{x}) f_j(\mathbf{x}, \theta), \quad (6.8)$$

where δ_j evaluates to 1 if $\mathbf{x} \in R_j$ and 0 otherwise, and f_j is the SBRDF in region R_j .

6.5.1 Data Acquisition and SBRDF Recovery

We require knowledge of the surface geometry and a set of images taken from known view-point and known directional illumination. In addition, in order to estimate the separate diffuse and specular reflectance components in Eq. (6.7), the input images must be similarly decomposed. While it can be accomplished in many ways (see, e.g., [94, 79]), we performed the specular/diffuse separation by placing linear polarizers on both the camera and light source and exploiting the fact that the specular component preserves the linear polarization of the incident radiance. Two exposures were captured for each view/lighting configuration, one with the polarizers aligned (to observe the sum of specular and diffuse components), and one with the source polarizer rotated by 90° (to observe only the diffuse component.) To

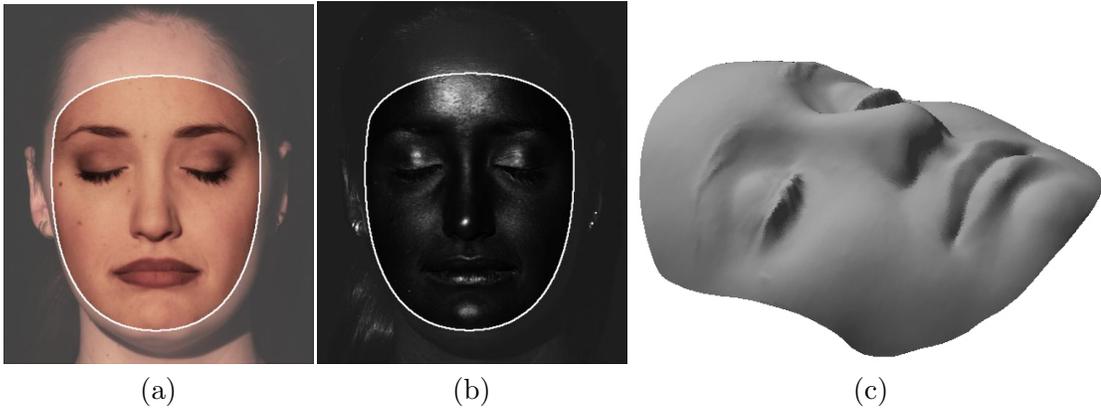


Figure 6.8: (a,b) An example of one input image separated into its specular and diffuse components. (c) The geometry of the face used for SBRDF recovery and rendering.

recover the geometry, we used a variant of photometric stereo, since it directly provides the precise surface normal estimates required for reflectometry. The geometry and an example of a decomposed input image are shown in Fig. 6.8.

Four view/illumination configurations, with fixed orthographic viewpoint and four known illumination directions, were used as input. (Two polarized exposures were captured in each configuration.) The average angular separation of the light directions is 21° (considerably more sparse than in previous work,) and these directions span a large area of frontal illumination.

The RGB albedo $a(\mathbf{x})$ in Eq. (6.7) is estimated as the median of the four diffuse samples at each surface point, and normalized diffuse reflectance samples are computed by dividing by $a(\mathbf{x})$ at each point. The resulting normalized diffuse samples are used to estimate the three functions, $d_k(\mathbf{x}, \theta)$, in Eq. (6.7) using the RBF techniques described in Sect. 6.3. The samples from the specular images were similarly used to compute g . 2000 RBF centers were used for each diffuse colour channel, and for the specular component, 5000 RBF centers were used.

Each diffuse RBF interpolant requires the storage of 2006 coefficients (the weights for 2000 centers and the six polynomial coefficients) and 2000 sample points $\{\mathbf{q}_i\}$, each of which has five-dimensions. (This could be substantially reduced, for example, by using the same

center locations for all three channels.) Similarly, the specular component requires 5006 coefficients and 5000 centers. The total required storage for the four interpolants is 66 024 single precision floating-point numbers, or 258kB. This very compact representation of both viewpoint- and lighting-dependent reflectance effects.

6.5.2 Rendering

Images under arbitrary view and illumination are synthesized using this representation as follows. The SBRDF coordinates \mathbf{q} at any surface point $\mathbf{s}(\mathbf{x})$ are determined by the spatial coordinates \mathbf{x} , the surface normal at that point, and the view and illumination directions in its local coordinate frame. The radiance emitted from that point toward the camera is

$$I_k(\mathbf{x}) = f_k(\mathbf{q})(\mathbf{l} \cdot \hat{\mathbf{n}}), \quad k = R, G, B, \quad (6.9)$$

where f_k is given by Eq. (6.7) and \mathbf{l} represents the strength and direction of the illumination. Because Eq. (6.7) involves sums over a large number (up to 5000) of RBF centers for each pixel, direct image synthesis can be quite slow. As Fig. 6.10 demonstrates, however, this representation can be used in conjunction with precomputed light transport methods for real-time rendering, making it a useful intermediate representation between acquisition and rendering.

6.5.3 Results

The recovered spatially-varying reflectance function is shown in Fig. 6.9, where a synthetic image of the surface is compared to a real image for illumination conditions that were not used as input.

The spatial variation of the recovered reflectance function is evident in these images, and is shown explicitly in Fig. 6.9(d). In this plot, the recovered specular lobe on the nose is shown to be substantially more peaked than that of the cheek. Note, however, that the shape of this lobe as seen in Fig. 6.9(b) is slightly more broad than that in the actual image. This is caused by small movements of the subject during acquisition and calibration errors (light source position and source anisotropy.) Both of these lead to slight misalignment of

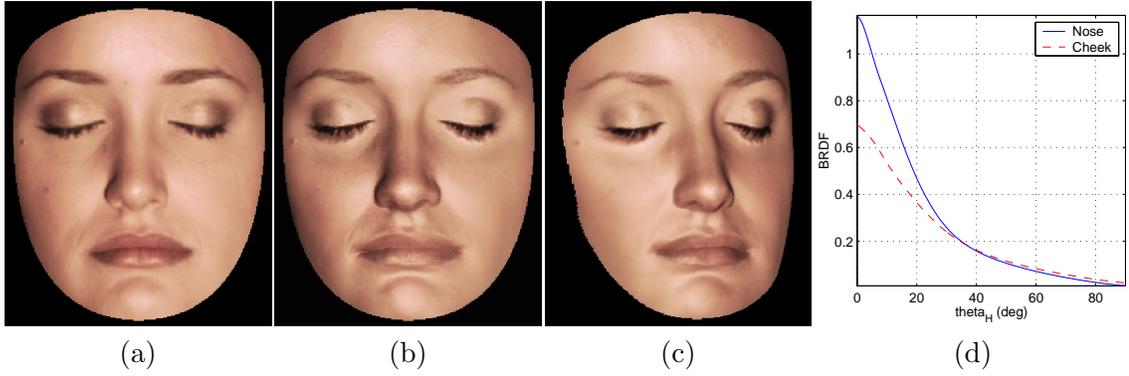


Figure 6.9: Comparison of an actual novel image (a) and synthesized image (b) that was rendered using the reflectance representation in Eq. (6.7). The spatially-varying reflectance was recovered by sharing the reflectance information from four input images. This representation can be used to synthesized viewpoint variation (c), and it accurately models the spatial variation of the reflectance function (d).

the images relative to the geometry and to each other and add noise to the SBRDF samples. (Accuracy could be improved, for example, using a high speed acquisition system such as that of Debevec et al. [24].) Figure 6.9(c) shows a synthetic image with changing view.

6.5.4 A Special Case: One Input Image

In the extreme case when only one input image is available, all reflectance samples lie on a hyperplane of constant w , reducing the dimension of the SBRDF by one. The dimension-independence of RBFs means this can be handled very naturally. Of course, by collapsing the θ_d (or w) dimension, we do not model Fresnel effects on the surface. (As an example, if we are given a single image captured with frontal illumination, we can say very little about the appearance at grazing angles.) Nevertheless, a single image *does* contain significant reflectance information, and as demonstrated in this section, our method obtains plausible results in this case.

For the single image case, we use a simplified SBRDF representation,

$$f_k(\mathbf{q}) = a_k(\mathbf{x}) + \hat{\mathbf{s}}_k g(\mathbf{q}), \quad k = R, G, B, \quad (6.10)$$

with

$$g(\mathbf{q}) = \sum_{i=1}^N \lambda_i \|\mathbf{q} - \mathbf{q}_i\|,$$

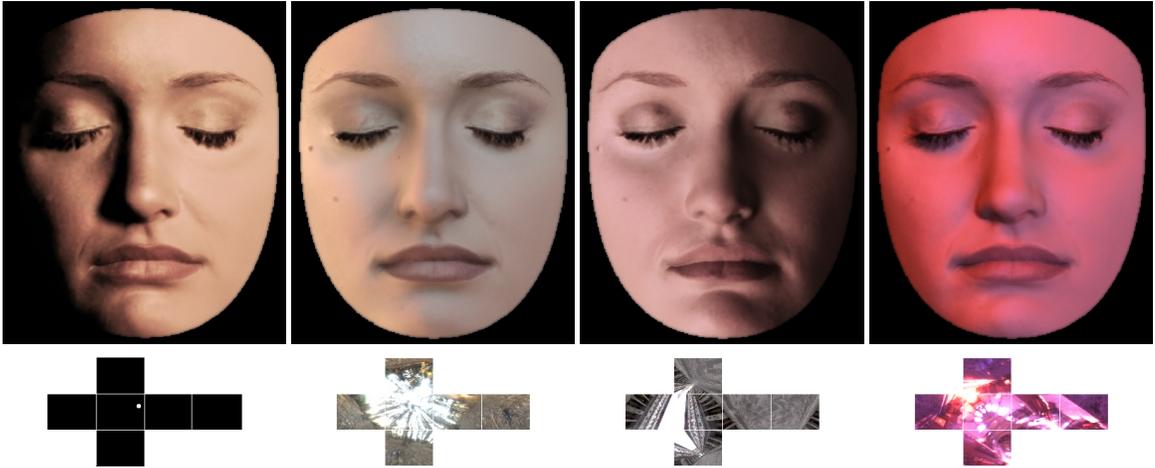


Figure 6.10: The face is rendered using a non-parametric, spatially varying reflectance function that is recovered from a single image (shown in Fig. 6.9(a)) of known geometry taken under directional illumination. These images were rendered in real-time using a technique based on precomputed information as described in [120].

where $\mathbf{q} = (x, y, u, v)$. The diffuse component of the reflectance function is modeled as Lambertian, and the albedo $a(\mathbf{x})$ is estimated directly from the reflectance samples in the diffuse component of the single input image. The specular component g is estimated from the specular reflectance samples using the fitting procedure of Sect. 6.3.2. 2000 RBF centers are used.

Figure 6.10 shows synthetic results under natural lighting from environment maps. These images were rendered in real-time, using the precomputed image technique discussed in [120].

6.6 Summary

In this chapter we show that, by careful parameterization and representation of reflectance, and by exploiting spatial coherence to share information spatially across the surface, we can estimate a non-parametric, spatially-varying reflectance function from a sparse set of images of known geometry. The method can be viewed as bridging parametric methods and dense, non-parametric methods that do not exploit spatial coherence. The spatial reflectance function can be approximated from a very small number of images, and at the

same time, the estimate converges to the true appearance as the number of images increases.

We formulate the task of estimating spatially-varying reflectance as a scattered-data interpolation problem, and to solve this interpolation problem we introduce a new representation of reflectance based on radial basis functions. This representation is compact and can be recovered directly from scattered, image-based data. It is shown to be a useful intermediate representation of spatially-varying reflectance, since it can be used in combination with existing real-time rendering techniques based on precomputed information.

Chapter 7

Conclusion

This dissertation addresses the problems of estimating both the shape and reflectance properties of an object from its images. For recovering accurate shape, it introduces the concept of a *reciprocal image pair* as a means of decoupling shape and reflectance in image data. These images exploit a symmetry property of reflectance known as Helmholtz reciprocity. *Helmholtz stereopsis* is defined as the process of recovering shape from reciprocal image pairs, and it is shown to enable accurate shape estimation for surfaces with arbitrary and complex reflectance. Since they operate independent of reflectance, Helmholtz stereo methods can be applied to a much broader class of surfaces than conventional techniques, and they perform well in cases where other techniques often fail. In addition, typical Helmholtz stereo techniques have the unique ability to provide direct estimates of both the surface and its Gauss map. This is a very desirable quality, since accurate surface normals are essential for reflectance measurement (as considered in the second part of this dissertation) and for rendering synthetic images (i.e., for predicting appearance under novel illumination and viewpoint.)

Chapters 3–5 discuss some of the significant properties of reciprocal images and the shape information they contain. These chapters show that more than one reciprocal pair is generally required to estimate a surface and its Gauss map, but that the shape of a continuous surface can often be recovered from a single reciprocal pair. Shape information can also be recovered in the uncalibrated case (when little is known about the imaging

system), and in some cases, the photometric information in a set of reciprocal pairs is sufficient for the recovery of both shape and the parameters of the acquisition system.

The second problem considered by this dissertation is that of estimating surface reflectance from images of known shape, a task made difficult by the vast amount of data it typically requires. *Reflectance sharing* is presented as a method that exploits the spatial coherence of reflectance to obtain an accurate estimate of an object's spatially-varying reflectance properties from a drastically reduced set of images. Unlike most existing techniques, this information is recovered without resorting to low-dimensional approximations (e.g., parametric models) of reflectance. Reflectance sharing shows that the task of estimating reflectance can be formulated as a scattered-data interpolation problem in five or six dimensions, and this dissertation presents a new representation of reflectance that can be used in this framework. The new representation has the added advantage of being very compact, making it useful for compression of large datasets.

The focus of this dissertation is the acquisition of appearance models (in terms of shape and reflectance) for surfaces with *arbitrary and complex surface reflectance*, meaning reflectance that is not necessarily well-represented using pre-chosen, low-dimensional reflectance models. By developing practical techniques that eliminate the restrictions imposed by assumed reflectance models, we move closer to the rapid acquisition of accurate appearance models for general objects and scenes.

Chapter 8

Future Work

8.1 Helmholtz Stereopsis

Chapters 3–5 demonstrate the ability of Helmholtz stereo methods to accurately recover the shape of general surfaces, but many properties of reciprocal images remain unexplored:

- Shadowed areas in reciprocal images provide shape information since they correspond to half-occluded regions and are indicators of depth discontinuities. This additional information was not used in the presented Helmholtz stereo methods, and we expect that it could increase the utility of the method. This is true in all three cases (multinocular, binocular and uncalibrated) that were considered.
- More direct methods of combining depth and surface orientation in the multinocular case must be explored. Model-based, surface-evolution approaches such as the stereo techniques developed by Faugeras and Keriven [29] and Yezzi and Soatto [117] are particularly attractive since they enable the use of simultaneous constraints on a surface and its Gauss map. Finite-element and parametric-surface approaches to shape-from-shading [45, 44, 104, 112] also provide inspiration for a second generation of Helmholtz stereo techniques.
- The distant-source (affine) uncalibrated case was studied in Chapter 5, but the perspective case was not discussed in detail. Since the reciprocity constraint is a Euclidean constraint, it is conceivable that reciprocal image pairs contain sufficient photometric

information to solve the autocalibration problem in the perspective case as well.

In addition, there are two fundamental limitations to Helmholtz stereopsis. First, we expect the accuracy of the results to be affected by the presence of significant interreflections. Second, the accuracy may be decreased for surfaces whose reflectance is not well represented by a BRDF, including surfaces with significant sub-surface scattering or mesostructure. The severity of these limitations remains to be studied in detail, but concerning the second limitation, it is important to note that most surfaces (including these supposed violators) can be described by a BRDF *at an appropriate scale*. The task becomes, then, one of automatically finding the scale at which these effects become negligible, which in turn determines the resolution of the recovered shape.

8.2 Reflectance Sharing

Chapter 6 showed that the spatial coherence of surface reflectance can be exploited to significantly reduce the number of images required to accurately estimate the spatially-varying reflectance of an object. To increase the practicality of the technique, the most immediate concern is computational efficiency. We have demonstrated that our compact RBF-based representation is a useful intermediate representation of spatially-varying reflectance since it can be used in combination with current rendering techniques based on precomputed information. In future work, it may be possible to develop real-time rendering techniques directly from the RBF-based representation. For example, fast multipole methods can be used to reduce the evaluation of Eq. (6.2) from $\mathcal{O}(N)$ to $\mathcal{O}(1)$ [4]. This may be a good alternative to using factored forms of BRDFs for homogeneous surfaces [75] and may provide a practical approach for real-time rendering of surfaces with spatially-varying, non-parametric BRDFs.

Another direction for future work is to consider anisotropic reflectance. The ideas of exploiting spatial coherence and using RBFs to interpolate scattered reflectance samples can still be applied to the anisotropic case. This requires a parameterization which is different from that presented in Sect. 6.2, however, and unless it is known *a priori*, it also requires

the simultaneous estimation of the field of tangent vectors on the surface.

Finally, we demonstrated reflectance sharing for surfaces with reflectance that varies smoothly across the surface or can be described by the product of a smooth function and a (possibly) discontinuous albedo or texture. An important direction of future work is to detect the regions of a surface where this is not an appropriate representation, and to develop and incorporate more suitable appearance representations for these regions.

Appendix A

Uncalibrated reciprocity constraint

This appendix compares the original multinocular and uncalibrated reciprocity constraints discussed in Sect. 5.1. Consider a reciprocal imaging system consisting of P reciprocal pairs captured from $M \leq P$ positions. In a calibrated system, the multinocular constraint (Eq. 5.2) can be used to establish correspondence, since for a valid correspondence, the matrix \mathbf{M} will satisfy $\text{rank } \mathbf{M} < 3$. In the uncalibrated case, we cannot compute \mathbf{M} , and instead we factor it as

$$\mathbf{M}_{P \times 3} = \mathbf{E}_{P \times M} \mathbf{U}_{M \times 3}, \quad (\text{A.1})$$

where $\mathbf{U}\hat{\mathbf{n}} = \tilde{\mathbf{w}}$, with $\tilde{\mathbf{w}}$ as defined in Eq. 5.3. Then, the uncalibrated constraint used to establish correspondence is $\text{rank } \mathbf{E} < M$. We can show that for $M \geq 4$ this is a weaker constraint than that based on Eq. 5.2.

There are two relevant rank inequalities for a general matrix product,

$$\begin{aligned} \text{rank}(\mathbf{AB}) &\leq \min(\text{rank } \mathbf{A}, \text{rank } \mathbf{B}) \\ \text{rank } \mathbf{A}_{n \times k} \mathbf{B}_{k \times m} &\geq \text{rank } \mathbf{A} + \text{rank } \mathbf{B} - k. \end{aligned}$$

In the present case (noting that $\text{rank } \mathbf{U} = 3$ for non-coplanar sources) these give

$$\text{rank } \mathbf{E} + 3 - M \leq \text{rank } \mathbf{M} \leq \min(\text{rank } \mathbf{E}, 3),$$

which tells us the following. First, if $M = 3$, $\text{rank } \mathbf{E} < M \iff \text{rank } \mathbf{M} < 3$, and the two constraints are equivalent. They are not equivalent, however, for $M \geq 4$. In this case, $\text{rank } \mathbf{M} < 3 \implies \text{rank } \mathbf{E} < M$, but the converse does not hold.

Bibliography

- [1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. In *Proc. Int. Conf. on Computer Vision*, volume 2, pages 383–390, 2001.
- [2] Paul A. Beardsley, Philip H. S. Torr, and Andrew Zisserman. 3D model acquisition from extended image sequences. In *Proc. European Conf. on Computer Vision*, pages 683–695, 1996.
- [3] G. Beatson, R.K. Goodsell and M.J.D. Powell. On multigrid techniques for thin plate spline interpolation in two dimensions. *Lectures in Applied Mathematics*, 32:77–97, 1995.
- [4] R.K. Beatson and G.N. Newsam. Fast evaluation of radial basis functions: I. *Comput. Math. Appl.*, 24:7–19, 1992.
- [5] P. Belhumeur and D.W. Jacobs. Comparing images under variable illumination. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1998.
- [6] P.N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. In *Proc. Int. Conf. on Computer Vision*, pages 431–438, 1993.
- [7] P.N. Belhumeur. A Bayesian approach to binocular stereopsis. *Int. Journal of Computer Vision*, 19, 1996.

- [8] P.N. Belhumeur and D. Mumford. A Bayesian treatment of the stereo correspondence problem using half-occluded regions. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 506–512, 1992.
- [9] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.
- [10] D.N. Bhat and S.K. Nayar. Stereo and specular reflection. *Int. Journal of Computer Vision*, 26(2):91–106, 1998.
- [11] J.F. Blinn. Models of light reflection for computer synthesized pictures. In *Computer Graphics (Proc. SIGGRAPH 1977)*, pages 192–198, 1977.
- [12] M.D. Buhmann. *Radial basis functions*. Cambridge University Press, 2003.
- [13] B. Cabral, M. Olano, and P. Nemeč. Reflection space image based rendering. In *SIGGRAPH 99*, pages 165–170, 1999.
- [14] J.C. Carr, R.K. Beatson, J.B. Cherrie, T.J. Mitchell, W.R. Fright, B.C. McCallum, and T.R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *Proc. SIGGRAPH*, pages 67–76, 2001.
- [15] H. Chen, P.N. Belhumeur, and D.W. Jacobs. In search of illumination invariants. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 254–261, June 2000.
- [16] W. Chen, J. Bouguet, M. Chu, and R. Grzeszczuk. Light field mapping: Efficient representation and hardware rendering of surface light fields. *ACM Transactions on Graphics (SIGGRAPH 2002 proceedings)*, 21(3):447–456, 2002.
- [17] F.J.J. Clarke and D.J. Parry. Helmholtz Reciprocity: its validity and application to reflectometry. *Lighting Research and Technology*, 17(1):1–11, 1985.
- [18] E.N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Vision, Graphics and Image Processing*, 18(4):309–328, April 1982.

- [19] R. Cook and K. Torrance. A reflectance model for computer graphics. In *Proc. SIGGRAPH*, pages 307–316, 1981.
- [20] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution. In D. Hogg and R. Boyle, editors, *British Machine Vision Conference*, pages 337–346. Springer-Verlag, 1992.
- [21] J.E. Cryer, P.-S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, 1995.
- [22] K.J. Dana and J. Wang. Device for convenient measurement of spatially varying bidirectional reflectance. *J. Optical Society of America A*, 21(1):1–12, January 2004.
- [23] Kristin J. Dana, Bram van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, 1999.
- [24] P. Debevec, T. Hawkins, C. Tchou, H.P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH*, pages 145–156, 2000.
- [25] F. Deverney and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 208–213, 1994.
- [26] H.Q. Dinh, G. Turk, and G. Slabaugh. Reconstructing surfaces using anisotropic basis functions. In *Proc. Int. Conf. on Computer Vision*, pages 606–613, 2001.
- [27] Ondřej Drbohlav and Radim Šára. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Proc. European Conf. on Computer Vision*, pages II: 46–60, 2002.
- [28] J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive theory of functions of several variables*, pages 85–100. Springer-Verlag, 1977.

- [29] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods, and the stereo problem. *IEEE Trans. Image Processing*, 7(3):336–344, 1998.
- [30] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. European Conf. on Computer Vision*, pages 563–578, 1992.
- [31] Olivier Faugeras. Stratification of 3-D vision: Projective, affine, and metric representations. *J. Optical Society of America A*, 12(7):465–484, 1995.
- [32] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2002.
- [33] Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.
- [34] P. Fua and Y.G. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *Int. Journal of Computer Vision*, 16(1):35–56, September 1995.
- [35] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions in binocular stereo. In *Proc. European Conf. on Computer Vision*, Santa Margherita Ligure, Italy, 1992.
- [36] A. Georghiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 816–823, 2003.
- [37] A. Glassner. *Principles of digital image synthesis*. Morgan Kaufmann, 1995.
- [38] Xianfeng Gu, Steven Gortler, and Hugues Hoppe. Geometry images. *ACM Transactions on Graphics*, 21(3):355–361, 2002.
- [39] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

- [40] K. Hayakawa. Photometric stereo under a light source with arbitrary motion. *J. Opt Soc. Am.*, 11(11), 1994.
- [41] H. von. Helmholtz. *Treatise on Physiological Optics*, volume 1. Dover (New York), 1925.
- [42] A. Hertzmann and S. Seitz. Shape and material by example: a photometric stereo approach. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003.
- [43] B.K.P. Horn. Understanding image intensity. *Artificial Intelligence*, 8(11), 1977.
- [44] B.K.P. Horn. Height and gradient from shading. *Int. Journal of Computer Vision*, 5(1):37–75, 1990.
- [45] B.K.P. Horn and M.J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 33(2):174–208, 1986.
- [46] B.K.P. Horn and M.J. Brooks, editors. *Shape from Shading*. MIT Press, Cambridge, Mass., 1989.
- [47] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3(6):661–669, 1981.
- [48] Z. Janko, O. Drbohlav, and R. Šára. Radiometric calibration of a Helmholtz stereo rig. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004. To Appear.
- [49] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004. To Appear.
- [50] J. Kautz and M. McCool. Interactive rendering with arbitrary BRDFs using separable approximations. In *EGRW 99*, pages 247–260, 1999.

- [51] M-J. Kim. Verification of the reciprocity theorem. *Applied Optics*, 27(13):2645–2646, 1988.
- [52] Jan J. Koenderink and Andrea J. van Doorn. The generic bilinear calibration-estimation problem. *Int. Journal of Computer Vision*, 23(3):217–234, 1997.
- [53] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. *J. Optical Society of America A*, 8(2):377–385, February 1991.
- [54] J.J. Koenderink and A.J. van Doorn. Bidirectional reflection distribution function expressed in terms of surface scattering modes. In *Proc. European Conf. on Computer Vision*, pages II:28–39, 1996.
- [55] M.L. Koudelka, S. Magda, P.N. Belhumeur, and D.J. Kriegman. Acquisition, compression and synthesis of bidirectional texture functions. In *3rd International Workshop on Texture Analysis and Synthesis (Texture 2003)*, pages 59–64, 2003.
- [56] K.N. Kutulakos and S.M. Seitz. A theory of shape by space carving. *Int. Journal of Computer Vision*, 38(3):199–218, 2000.
- [57] E. Lafortune, S. Foo, K. Torrance, and D. Greenberg. Non-linear approximation of reflectance functions. In *Proc. SIGGRAPH*, pages 117–126, 1997.
- [58] P. Lalonde and A. Fournier. A wavelet representation of reflectance functions. *IEEE Trans. Visualization and Computer Graphics*, 3(4):329–336, 1997.
- [59] M.S. Langer and S. W. Zucker. Shape-from-shading on a cloudy day. *J. Opt Soc. Am.*, pages 467–478, 1994.
- [60] Y.G. Leclerc and A.F. Bobick. The direct computation of height from shading. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 552–558, 1991.
- [61] Aaron Lee, W. Sweldens, P. Schroder, L. Cowsar, and D. Dobkin. MAPS: Multiresolution adaptive parameterization of surfaces. In *SIGGRAPH 98*, pages 95–104, 1998.

- [62] Hendrik Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatially varying materials. In *Proc. Eurographics Workshop on Rendering*, 2001.
- [63] J. Lu and J.J. Little. Reflectance and shape from images using a collinear light source. *Int. Journal of Computer Vision*, 32(3):1–28, 1999.
- [64] R. Lu, J.J. Koenderink, and A.M.L. Kappers. Optical properties (bidirectional reflection distribution functions) of velvet. *Applied Optics*, 37(25):5974–5984, 1998.
- [65] S. Magda, T.E. Kriegman, D.J. Zickler, and P.N. Belhumeur. Beyond Lambert: Reconstructing surfaces with arbitrary BRDFs. In *Proc. Int. Conf. on Computer Vision*, pages II: 391–398, 2001.
- [66] A. Maki, M. Watanabe, and C. Wiles. Geotensity: combining motion and lighting for 3D surface reconstruction. *Int. Journal of Computer Vision*, 48(2):75–90, 2002.
- [67] Stephen R. Marschner. *Inverse rendering for computer graphics*. PhD thesis, Cornell University, 1998.
- [68] Stephen R. Marschner, Stephen H. Westin, Eric P. F. LaFortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-based BRDF measurement including human skin. In *Proc. Eurographics Workshop on Rendering*, pages 139–152, 1999.
- [69] L. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *Int. Journal of Computer Vision*, 8(1):71–91, July 1992.
- [70] Wojciech Matusik, Hanspeter Pfister, Matthew Brand, and Leonard McMillan. Image-based 3D photography using opacity hulls. *ACM Transactions on Graphics*, 21(3):427–437, 2002.
- [71] Wojciech Matusik, Hanspeter Pfister, Matthew Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003.

- [72] Wojciech Matusik, Hanspeter Pfister, Matthew Brand, and Leonard McMillan. Efficient isotropic BRDF measurement. In *Proc. Eurographics Symposium on Rendering*, pages 241–248, 2003.
- [73] D. K. McAllister, A. Lastra, and W. Heidrich. Efficient rendering of spatial bi-directional reflectance distribution functions. In *Proc. Eurographics Workshop on Rendering*, 2002.
- [74] David Kirk McAllister. *A Generalized surface Appearance Representation for Computer Graphics*. PhD thesis, University of North Carolina at Chapel Hill, 2002.
- [75] M. McCool, J. Ang, and A. Ahmad. Homomorphic factorization of BRDFs for high-performance rendering. In *Proc. SIGGRAPH*, pages 171–178, 2001.
- [76] C. A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 1:11–22, 1986.
- [77] M. Minnaert. The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93:403–410, 1941.
- [78] J. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
- [79] S.K. Nayar, X. Fang, and T. Boult. Separation of reflection components using color and polarization. *Int. Journal of Computer Vision*, 21(3):163–186, 1997.
- [80] S.K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE J. of Robotics and Automation*, 6(4):418–431, 1990.
- [81] F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 160, National Bureau of Standards (US), 1977.

- [82] K. Nishino, Y. Sato, and K. Ikeuchi. Eigen-texture method: appearance compression and synthesis based on a 3D model. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1257–1265, Nov 2001.
- [83] R. Onn and A.M. Bruckstein. Integrability disambiguates surface recovery in two-image photometric stereo. *Int. Journal of Computer Vision*, 5(1):105–113, 1990.
- [84] M. Oren and S.K. Nayar. Generalization of the Lambertian model and implications for machine vision. *Int. Journal of Computer Vision*, 14:227–251, 1996.
- [85] A. Pentland. Photometric motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(9):879–890, 1991.
- [86] B. Phong. Illumination for computer-generated pictures. *Comm. of the ACM*, 18(6):311–317, 1975.
- [87] T. Poggio, V. Torre, and C. Koch. Computational vision and regularisation theory. *Nature*, 317:314–319, 1985.
- [88] P. Poulin and A. Fournier. A model for anisotropic reflection. In *Proc. SIGGRAPH*, pages 273–282, 1990.
- [89] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. *J. Optical Society of America A*, 18(10):2448–2458, October 2001.
- [90] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proc. SIGGRAPH*, pages 117–128, 2001.
- [91] R. Ramamoorthi and P. Hanrahan. Frequency space environment map rendering. *ACM Transactions on Graphics (SIGGRAPH 02 proceedings)*, 21(3):517–526, 2002.
- [92] Ruth Rosenholtz and Jan J. Koenderink. Affine structure and photometry. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 790–795, June 1996.

- [93] S. Rusinkiewicz. A new change of variables for efficient BRDF representation. In *EGRW 98*, pages 11–22, 1998.
- [94] Y. Sato and K. Ikeuchi. Temporal-color space analysis of reflection. *J. Optical Society of America A*, 11(11):2990–3002, 1994.
- [95] Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. Object shape and reflectance modeling from observation. In *Proc. SIGGRAPH*, pages 379–387, 1997.
- [96] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *Proc. SIGGRAPH*, pages 161–172, 1995.
- [97] S. Shafer. Using color to separate reflection components. *COLOR research and applications*, 10(4):210–218, 1985.
- [98] Larry S. Shapiro, Andrew Zisserman, and Michael Brady. 3D motion recovery via affine epipolar geometry. *Int. Journal of Computer Vision*, 16:147–182, 1995.
- [99] F. Sillion, J. Arvo, S. Westin, and D. Greenberg. A global illumination solution for general reflectance distributions. In *SIGGRAPH 91*, pages 187–196, 1991.
- [100] W.M. Silver. Determining shape and reflectance using multiple images. Master’s thesis, MIT, 1980.
- [101] D. Simakov and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *Proc. Int. Conf. on Computer Vision*, pages 1202–1207, October 2003.
- [102] W.C. Snyder. Reciprocity of the bidirectional reflectance distribution function (BRDF) in measurements and models of structured surfaces. *IEEE Trans. Geoscience and Remote Sensing*, 36(2):685–691, March 1998.
- [103] W.C. Snyder. Definition and invariance properties of structured surface BRDF. *IEEE Trans. Geoscience and Remote Sensing*, 40(5):1032–1037, May 2002.

- [104] R. Szeliski. Fast shape from shading. *Computer Vision, Graphics and Image Processing*, 53(2):125–153, 1991.
- [105] H.D. Tagare and R.J.P. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(2):133–152, February 1991.
- [106] C. Tomasi and T Kanade. Shape and motion from image streams under orthography: A factorization approach. *Int. Journal of Computer Vision*, 9(2):137–154, 1992.
- [107] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt Soc. Am.*, 57:1105–1114, 1967.
- [108] P. Tu and P. Mendonça. Surface reconstruction via Helmholtz reciprocity with a single image pair. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2003.
- [109] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.
- [110] G. Wahba. *Spline models for observational data*. Number 59 in CBMS-NSF regional conference series in applied mathematics. SIAM, 1990.
- [111] G.J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics*, 26:265–272, 1992.
- [112] G.-Q. Wei and G. Hirzinger. Parametric shape-from-shading by radial basis functions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(4):353–365, 1997.
- [113] D Weinshall. Model-based invariants for 3-D vision. *Int. Journal of Computer Vision*, 10(1):27–42, 1993.
- [114] S. Westin, J. Arvo, and K. Torrance. Predicting reflectance functions from complex surfaces. In *Proc. SIGGRAPH*, pages 255–264, 1992.

- [115] D. Wood, D. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Stuetzle. Surface light fields for 3D photography. In *Proc. SIGGRAPH*, pages 287–296, 2000.
- [116] R.J. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 17:117–140, 1981.
- [117] A. Yezzi and S. Soatto. Stereoscopic segmentation. In *Proc. Int. Conf. on Computer Vision*, volume 1, pages 59–66, 2001.
- [118] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *Proc. SIGGRAPH*, pages 215–224, 1999.
- [119] Li Zhang, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *Proc. Int. Conf. on Computer Vision*, pages 618–625, October 2003.
- [120] T. Zickler, S. Enrique, R. Ramamoorthi, and P. Belhumeur. Reflectance sharing: Image-based rendering from a sparse set of images. In Preparation.