Surface Reflectance Recognition and Real-World Illumination Statistics

by

Ron O. Dror

B.S., Electrical Engineering
B.A., Mathematics
Rice University, 1997

M.Phil., Biological Sciences
University of Cambridge, 1998

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September, 2002

Signature of Author: _____

Department of Electrical Engineering and Computer Science
August 30, 2002

Certified by: _____

Alan S. Willsky
Professor of Electrical Engineering
Thesis Supervisor

Certified by: _____

Edward H. Adelson
Professor of Vision Science
Thesis Supervisor

Accepted by: _____

Arthur C. Smith
Professor of Electrical Engineering
Chair, Committee for Graduate Students

Surface Reflectance Recognition and Real-World Illumination Statistics

by Ron O. Dror

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree
of Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Humans distinguish materials such as metal, plastic, and paper effortlessly at a glance. Traditional computer vision systems cannot solve this problem at all. Recognizing surface reflectance properties from a single photograph is difficult because the observed image depends heavily on the amount of light incident from every direction. A mirrored sphere, for example, produces a different image in every environment. To make matters worse, two surfaces with different reflectance properties could produce identical images. The mirrored sphere simply reflects its surroundings, so in the right artificial setting, it could mimic the appearance of a matte ping-pong ball. Yet, humans possess an intuitive sense of what materials typically "look like" in the real world. This thesis develops computational algorithms with a similar ability to recognize reflectance properties from photographs under unknown, real-world illumination conditions.

Real-world illumination is complex, with light typically incident on a surface from every direction. We find, however, that real-world illumination patterns are not arbitrary. They exhibit highly predictable spatial structure, which we describe largely in the wavelet domain. Although they differ in several respects from the typical photographs, illumination patterns share much of the regularity described in the natural image statistics literature.

These properties of real-world illumination lead to predictable image statistics for a surface with given reflectance properties. We construct a system that classifies a surface according to its reflectance from a single photograph under unknown illumination. Our algorithm learns relationships between surface reflectance and certain statistics computed from the observed image. Like the human visual system, we solve the otherwise underconstrained inverse problem of reflectance estimation by taking advantage of the statistical regularity of illumination. For surfaces with homogeneous reflectance properties and known geometry, our system rivals human performance.

Thesis Supervisors: Alan S. Willsky, Professor of Electrical Engineering
Edward H. Adelson, Professor of Vision Science

Acknowledgments

I consider myself fortunate to have completed this thesis under not one but two brilliant and dedicated advisors. Both of them contributed to my thesis research and taught me a great deal — Alan Willsky through his emphasis on mathematical rigor, his incredibly broad knowledge base, and his organizational skills; Ted Adelson through his amazing intuition, his scientific honesty, and his foresight in identifying important research problems. The collaboration was not always an easy one, given the difference in perspectives and approaches. Yet, even that proved to be a learning experience. I would like to thank Ted and Alan for agreeing to work not only with me, but also with each other.

I would also like to thank my other two committee members, Leonard McMillan and Bill Freeman. Both provided valuable feedback on this thesis. Bill offered much-needed encouragement during the difficult early stages of my thesis work. Leonard helped me understand the connections between my work and current trends in computer graphics.

My thesis research, as well as my experience in graduate school, benefited from extensive interaction with exceptional colleagues. Roland Fleming and I collaborated for over two years on human vision experiments. Originally inspired by early stages of the computational work described here, these experiments repeatedly suggested new avenues for further computational investigation. Marshall Tappen, my officemate for the past two years, acted as a sounding board for nearly all the ideas in this thesis. Marshall also implemented the surface inflation techniques described in Chapter 6 and taught me most of what I know about Linux. Thomas Leung not only collaborated in the initial investigation of illumination but also suggested the use of support vector machines and normalized convolution. Conversations with Antonio Torralba led to the development of much of the analysis presented in Chapters 5 and 6, including the dependence of wavelet coefficient distributions on image blur and on surface geometry. Polina Golland, Jason Rennie, and especially Ryan Rifkin guided my exploration of machine learning techniques; John Fisher did the same for feature selection. Samson Timoner taught me the basics of spherical wavelet transforms and the lifting scheme. Martin Wainwright repeatedly clarified points related to natural image statistics. Mike Schneider helped solidify my grasp of estimation theory, while Ilya Pollack did the same for information theory. Fredo Durand pointed out the relevance of understanding human perception when creating realistic computer graphics. Ray Jones proofread much of this thesis and helped clarify radiometric terminology. Finally, the members of

the Stochastic Systems Group and the Artificial Intelligence Laboratory deserve thanks for creating a supportive work environment, for answering a wide variety of technical questions, and for providing a continual stream of interesting ideas.

The systems and results presented here depend on code and data due to a number of other researchers, including Eero Simoncelli's matlabPyrTools, Ronan Collobert's SVMTorch, and Paul Debevec's light probes. Several people not only shared their research products but also spent a good deal of time helping me use them. Seth Teller, Neel Master, and Michael Bosse shared the data set from the MIT City Scanning Project and helped me use it to construct illumination maps. Julio Castrillon-Candas spent hours with me adapting his spherical wavelet code for use in analyzing illumination maps. Greg Ward repeatedly answered detailed questions about his *Radiance* software. Gunnar Färneback provided advice on the use of normalized convolution techniques. Rebecca Loh and Meredith Talusan acquired the data sets of photographs described in Chapter 6.

Perhaps the most rewarding aspect of my time at MIT was the freedom to explore a variety of research interests. My latest computational biology interests, involving gene arrays, grew out of a course taught by David Gifford, Rick Young, and Tommi Jaakkola. Jon Murnick proved an inspiring collaborator during the many hours we spent developing Bayesian methods for analysis of array data. Alex Hartemink provided helpful advice throughout that project. Nicola Rinaldi, Georg Gerber, and Gene Yeo agreed to co-teach a course in gene expression analysis with me.

My time at MIT would have been far less enjoyable and productive without the support of wonderful friends. Although I will not try to list them all, Mark Peng deserves mention for having shared a room with me during my entire time at MIT, after we were paired at random upon our arrival to Ashdown House.

Finally, I would like to thank my parents and my sister Daphna for their support over many years.

Contents

Abstract **3**

Acknowledgments **5**

List of Figures **11**

List of Tables **15**

1 Introduction **17**
1.1 Motivation 18
1.2 Thesis Organization and Contributions 20

2 Background **23**
2.1 Physics of Reflectance 23
2.1.1 Definition of Reflectance 23
2.1.2 Reflectance Properties of Real-World Materials 25
2.2 Previous Approaches to Reflectance Estimation 30
2.3 Reflectance Estimation in Human Vision 31
2.4 Summary and Discussion 37

3 Problem Formulation **39**
3.1 The Forward Problem: Rendering 39
3.1.1 The General Case 39
3.1.2 Reflectance as a Convolution 40
3.1.3 Image Formation 44
3.2 Reflectance Estimation as an Inverse Problem 46
3.2.1 Underconstrained Nature of the Problem 46
3.2.2 Statistical Formulation 48
3.3 Restricting the Reflectance Space 50
3.3.1 Parameter Estimation 50
3.3.2 Classification 50
3.4 Normalization for Overall Illumination Intensity 50

3.5 Summary and Discussion 51

4 Real-World Illumination Statistics 53
4.1 Measuring Illumination as an Image 53
4.2 Data Sets 54
4.3 Spherical Projection 55
4.4 Illumination Intensity Distribution 56
4.4.1 Marginal Distribution of Intensity 56
4.4.2 Non-Stationarity 58
4.4.3 Joint Distribution of Illumination from Adjacent Directions . . . 59
4.5 Spherical Harmonic Power Spectra 61
4.6 Wavelet Statistics 63
4.7 Summary and Discussion 68

5 From Illumination Statistics to Image Statistics 73
5.1 Dependence of Image Statistics on Reflectance 73
5.1.1 Statistics of Intensity Distribution 77
5.1.2 Statistics of Marginal Distributions of Wavelet Coefficients . . . 79
5.1.3 Statistics of Joint Distributions of Wavelet Coefficients 82
5.2 Reflectance Classification Based on Image Features 86
5.3 Summary and Discussion 88

6 Designing a Reflectance Classifier 89
6.1 Machine Learning Techniques for Reflectance Classification 89
6.1.1 Classification Techniques 89
6.1.2 Feature Selection and Classifier Performance 95
6.1.3 Misclassifications 101
6.2 Geometry 102
6.2.1 Handling Known Differences in Geometry: Remapping to the Equivalent Sphere 104
6.2.2 Robustness to Incorrect Assumed Geometry 110
6.2.3 Misalignment Artifacts for Non-Convex Surfaces 114
6.3 Summary and Discussion 116

7 Conclusions and Suggestions for Future Research 119
7.1 Thesis Contributions 119
7.1.1 Illumination Statistics 120
7.1.2 Relationship Between Surface Reflectance and Image Statistics . 120
7.1.3 Classification of Reflectance Based on Image Statistics 121
7.1.4 Effect of Surface Geometry 122
7.2 Suggestions for Future Work 122
7.2.1 Refinement of Reflectance Estimation Algorithms 122

7.2.2 An Alternative Approach to Compensating for the Effects of Surface Geometry . . . 123
7.2.3 Shadows, Unknown Geometry, and Spatially Varying Reflectance 124
7.2.4 Exploiting Additional Image Cues . . . 125
7.2.5 Relationship Between Reflectance and Texture . . . 125

A Expression of a Specular Image as a Convolution 129

B Rendering Under Photographically-Acquired Illumination 133

C Effect of Geometry on Image Statistics 135
C.1 Distortion Due to Geometry: Analysis in Three Dimensions . . . 135
C.1.1 Special Case of a Sphere . . . 138
C.2 Assumption of Stationarity on Gaussian Sphere . . . 139

Bibliography 141

List of Figures

1.1 Typical photographs including surfaces with different reflectance properties 17
1.2 Spheres of different reflectance under different illuminations 18

2.1 Geometry for definition of the BRDF 24
2.2 Cross-sections of Lambertian and specular BRDFs 26
2.3 Geometry used to define the specular component of the Ward model . . 28
2.4 Cross-sections of BRDFs specified by the Ward model with various parameter settings 29
2.5 Screen shot from the reflectance matching experiment 32
2.6 Range of reflectances in the reflectance matching experiment 33
2.7 Spheres rendered under each of the illuminations used in the experiments 34
2.8 Performance of individual subjects for specific real-world illumination conditions 35
2.9 Comparison of spheres rendered under point light source illumination and natural illumination. 36

3.1 Geometry for the Radiance Equation 41
3.2 Parameterization of surface location by surface normal 42
3.3 Local and global coordinates for a point on a surface 43
3.4 Rays corresponding to a fixed reflected direction in the local coordinate system 44
3.5 Imaging geometries for a viewer at a finite distance from the surface and for an infinitely distant viewer 46
3.6 A photograph of a matte sphere 48
3.7 Reflectance classification using photographs of surfaces under unknown real-world illumination 51

4.1 Examples of the illumination maps, shown in equal-area cylindrical projection 55
4.2 Equal-area cylindrical projection 56
4.3 Illumination intensity distributions for real-world illumination maps . . 57
4.4 Dependence of illumination on elevation 58

4.5 Pixelwise mean of over 300 images of outdoor scenes containing a person, aligned with respect to the person's head before averaging 59
4.6 Joint histograms of horizontally adjacent pixels in illumination maps . . 60
4.7 Spherical harmonic power spectra of several illumination maps 62
4.8 Spherical harmonic power spectrum of an illumination map with pixel values linear in luminance, and with the largest pixel values clipped . . 63
4.9 Mean power spectra of the Teller illumination maps 64
4.10 Distributions of spherical wavelet coefficients at successive scales, along with generalized Laplacian fits, for the Teller illumination maps 66
4.11 Variation in marginal distributions of wavelet coefficients from one image to another 67
4.12 Distributions of planar quadrature mirror filter wavelet coefficients at successive scales, along with generalized Laplacian fits, for the Teller illumination maps 68
4.13 Conditional histograms for a horizontal wavelet coefficient given the values of its neighbors 69
4.14 Spheres of identical reflectance properties rendered under synthetic illumination maps generated using texture synthesis algorithms 70

5.1 A chrome sphere and a rough metal sphere, each photographed under two different real-world illuminations 74
5.2 A photographically-acquired illumination map, illustrated on the inside of a spherical shell, and three surfaces of different geometry and reflectance rendered under that illumination 75
5.3 Effect of changes in Ward model reflectance parameters on images of a sphere under fixed illumination 75
5.4 View geometry for specular reflection from a planar surface 76
5.5 Effect of changes in Ward model reflectance parameters on an image of a flat surface and its pixel intensity histogram 77
5.6 Sensitivity of pixel histogram statistics to Ward reflectance parameters . 78
5.7 Sensitivities of wavelet coefficient statistics with respect to Ward model reflectance parameters. 80
5.8 Frequency response of a wavelet basis function before and after Gaussian blur 81
5.9 Images of a rough metal sphere, corrupted by white and pink Gaussian noise 83
5.10 Two painted clay surfaces with visible three-dimensional texture 84
5.11 Location in a two-dimensional feature space of images corresponding to several surface reflectances under several illumination conditions 85
5.12 Regions of a feature space assigned to different reflectance classes by a support vector machine classifier on the basis of training images 87

6.1 Class boundaries assigned by three different discriminative classifiers on the basis of the same training data points 91
6.2 Unwrapping an annulus with a polar-to-rectangular coordinate transformation 96
6.3 Synthetic spheres representing 11 reflectances, each rendered under one of Teller's illumination maps 97
6.4 Photographs of nine spheres, all under the same illumination 98
6.5 A chrome sphere photographed under each of seven illuminations 99
6.6 Elimination of uninformative features improves classifier performance . . 101
6.7 Examples of classification errors for the set of photographs 103
6.8 Classification for surfaces of arbitrary known geometry 104
6.9 Spheres of different sizes, rendered under the same illumination 105
6.10 Two differently shaped surfaces with the same reflectance photographed under the same illumination 105
6.11 Images of surfaces of three geometries, each mapped to a Gaussian sphere of the same reflectance and illumination, then "unwrapped" to a cylindrical projection of the sphere 107
6.12 Synthetic images of surfaces of three different geometries, each rendered with 6 reflectances under one of Teller's illumination maps 109
6.13 Images of a sphere and an ellipsoid with identical reflectance, rendered under the same illumination 111
6.14 Sample surface geometry in a two-dimensional world 112
6.15 Reconstruction of a Gaussian sphere for a non-convex shape, using approximate geometry 115
6.16 Painted clay shapes with convex, approximately spherical geometries, all photographed under the same illumination 117
6.17 Painted clay shapes with non-convex geometries 118

7.1 Sample images from the Columbia-Utrecht Reflectance and Texture database 126

A.1 View geometry for a specular surface 129

C.1 Spheres with both specular and diffuse reflectance components, illuminated by point sources in two different directions 140

List of Tables

3.1 Notation for Chapter 3 40

6.1 Classification performance for different feature sets 96
6.2 Classification performance for different machine learning techniques . . . 102
6.3 Classification performance for surfaces with different geometries 108

Chapter 1

Introduction

Humans effortlessly recognize surfaces with different optical reflectance properties at a glance. In the images of Figure 1.1, we recognize the shiny metal of the bowl, the rough metal of the pie tin, the white matte tabletop, the glossy apple skins, and the wet glistening ice cream. This ability to characterize reflectance properties from images in uncontrolled real-world environments is impressive for two reasons. First, images of identical surfaces in various settings can be very different. Figure 1.2 shows four spheres, each photographed in two locations. Images of different spheres in the same setting are more similar in a pixelwise sense than images of the same sphere in different settings. Second, two identical images may represent surfaces with different reflectance properties. Any of the images in Figure 1.2 could in principle be a photograph of a chrome surface; a chrome sphere simply reflects its environment, so it could, in principle, take on an arbitrary appearance.

In a typical real-world setting, however, distinctive image features characterize the appearance of chrome. We know what the real world typically "looks like," so we recognize its reflection in the surface. The visual world contains sharp edges, for example, so we expect to see sharp edges in the image of the chrome sphere. Estimation of surface

Figure 1.1. Typical photographs including surfaces with different reflectance properties

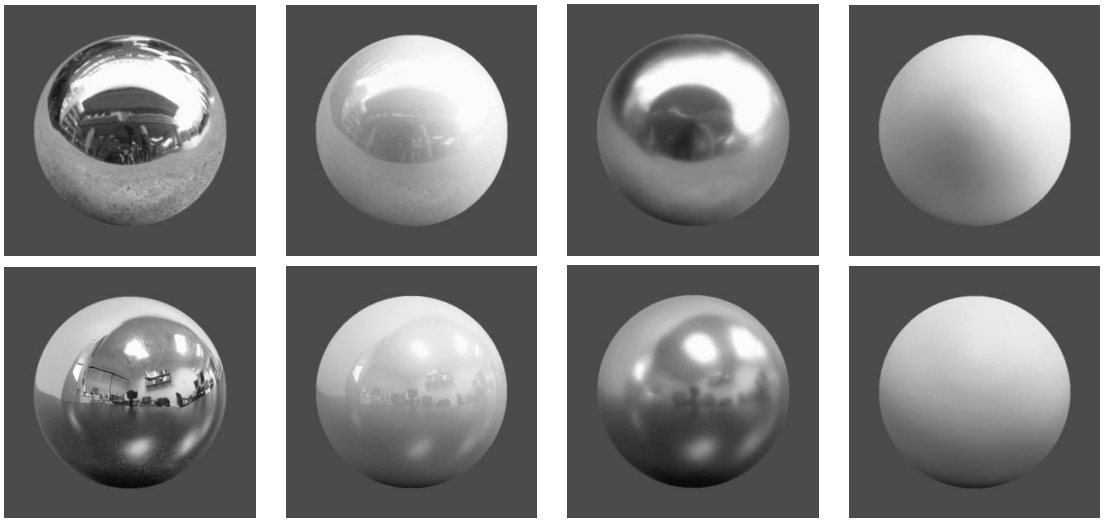


Figure 1.2. The two images in each column are photographs of the same sphere, shown over a standard gray background. The images in each row were photographed in the same location, under the same illumination.

reflectance properties involves recognition of patterns due not only to the physical laws governing electromagnetic reflectance, but also to the visual appearance of real-world environments.

This thesis explores the relationship between the reflectance properties of a surface and the statistics of an image of that surface under complex, real-world illumination. We show that the spatial structure of real-world illumination patterns exhibits a great deal of statistical regularity. We identify statistical features that a vision system can use to identify reflectance properties from an image. We use these features to design image-based reflectance classifiers for surfaces under unknown, uncontrolled illumination.

The analysis and results of this thesis rely only on a single monochrome image of the surface of interest. One could undoubtedly improve the performance of the proposed reflectance recognition algorithms by exploiting image context, motion, and color. However, we wish to determine what information the basic image structure captures about reflectance, even in the absence of these additional cues. We have found that humans can estimate certain surface reflectance properties given only a single isolated image of a surface [33, 34]. A computational vision system should also be able to accomplish this task.

■ 1.1 Motivation

Humans take for granted their ability to recognize materials, from gold to skin to ice cream, in a wide variety of environments. We can often judge whether a surface is wet or dry, rough or smooth, clean or dirty, liquid or solid, even soft or hard. We sometimes

make mistakes; one can design a foam object that appears to be a rock, but these are sufficiently unusual to be sold as novelties in curiosity shops. The ability to recognize and characterize materials is essential to interaction with the visual world. We rely on that ability to identify substances (*e.g.*, food), to judge their state (*e.g.*, fresh or rotten), and to identify the objects and scenes that they comprise. We recognize the human form not only by its geometry, but by the material properties of flesh and clothing. We recognize coins as much by their metallic reflectance properties as by their disk-like shapes.

Current computer vision systems, on the other hand, are typically powerless to distinguish materials accurately in uncontrolled, real-world environments. A robot may not eat ice cream, but many machine vision applications still demand material recognition abilities. An autonomous vehicle should be able to recognize a wet road or mud before driving onto it. An industrial inspection system should be able to recognize a dirty surface. A surgical robot should be able to distinguish different tissues. An industrial robot should be able to distinguish solids from powders and liquids. A face recognition system should be able to distinguish a human from a mannequin.

The desire to build vision systems capable of recognizing materials provides the primary motivation for the present work. Reflectance and texture both differentiate materials. Over the past several years, researchers have taken significant strides toward characterization and recognition of texture [43, 81]. We wish to do the same for reflectance, creating vision systems capable of distinguishing materials based on their reflectance properties.

While this thesis focuses on analysis of visible-spectrum photographs, other imaging modalities pose analogous material recognition problems. One may wish to distinguish terrain types from a remote radar image, or tissue types in a medical image. Although the physics of image formation depends on the specific modality, the difference in appearance of various materials often stem from their reflectance properties.

Additional motivation for our work stems from computer graphics. Modern-day graphics is constrained as much by the acquisition of realistic scene models as by rendering requirements. One often wishes to build a model of a real-world scene from photographs of that scene. If one plans to render the scene from a different viewpoint, under different illumination, or with synthetic objects inserted, one needs to recover not only surface geometry, but also surface reflectance properties. Graphics researchers have also observed that one can increase the realism of synthetic scenes by rendering them under illumination patterns measured in the real world [23]. Models for the statistical regularities of real-world illumination may allow us to reconstruct real-world illumination patterns from sparse samples, or to synthesize illumination patterns that provide the impression of image realism.

An understanding of real-world illumination and an ability to recognize reflectance properties are also necessary to overcome limitations of traditional computer vision algorithms for shape and motion estimation. Shape-from-shading algorithms, for example, require a known relationship between surface orientation and image irradiance

(brightness) [45]. This relationship depends on the reflectance and illumination of the surface, exhibiting particularly dramatic changes as a specular surface moves from one illumination to another. One may be able to recover shape for such a surface in an unknown setting by exploiting the statistical properties of real-world illumination patterns as well as the relationships between illumination, reflectance, geometry, and the observed image. Classical methods for optical flow estimation and stereo reconstruction pose a similar problem. These techniques are based on the constant brightness assumption, which holds only for Lambertian surfaces [46,61]. The specularities moving across a smooth metallic surface therefore produce incorrect depth and motion estimates. To remedy these problems, one must be able to distinguish image changes due to non-Lambertian surface reflectance from changes due to an actual change in the position of surfaces with respect to the viewer.

Finally, analysis of the relationship between real-world images and surface reflectance facilitates investigation of perceptual mechanisms for surface recognition [33,34]. Experimental studies of these mechanisms not only contribute to our scientific understanding of the human visual system but also have practical implications for computer graphics. The relevant measure of realism for most graphics applications is perceptual. To design reflectance models and implement efficient rendering techniques that produce images with a realistic appearance, one must understand which image features the visual system uses in recognizing surface geometry and reflectance.

■ 1.2 Thesis Organization and Contributions

The following two chapters cover background material and formulate the reflectance recognition problem mathematically. Chapter 2 describes prior work in physics, computer graphics, computer vision, and human perception that frames the developments of this thesis. We define surface reflectance and discuss the reflectance properties of real-world materials. We describe previous image-based methods for measuring surface reflectance. We also summarize studies of the human visual system's ability to recognize surface reflectance properties, including our own experiments performed in conjunction with the computational work of this thesis.

Chapter 3 poses surface reflectance estimation under unknown illumination as a blind inverse problem. The process by which a surface interacts with incident light to produce an image constitutes the forward problem and serves as the basis for rendering in computer graphics. Even when one assumes that surface geometry is known and that reflectance properties do not vary across a surface, the problem of recovering reflectance from an image under unknown illumination is underconstrained. In order to solve it, we must exploit prior information about the real world.

Chapters 4, 5, and 6 cover the major results of this thesis. Chapter 4 presents an empirical study of the statistical properties of real-world illumination. We use a spherical image, or *illumination map*, to describe the amount of light incident from every direction at a point in the real world. Illumination maps exhibit statistical properties

that vary little from one location to another. We describe these properties in terms of illumination intensity distributions, power spectra, and especially wavelet coefficient distributions, comparing our results to those reported for typical photographs in the natural image statistics literature. Although the remainder of this thesis focuses on reflectance recognition, the properties of real-world illumination are relevant to a variety of problems in computer vision and computer graphics.

Chapter 5 shows that the regularity of real-world illumination leads to predictable relationships between the reflectance of a surface and the statistics of an image of that surface. We explore these relationships using a parameterized reflectance model. Certain image statistics vary significantly with changes in reflectance but little from one real-world illumination to another. One can build a reflectance classifier for images under unknown illumination by partitioning a feature space based on these statistics into regions corresponding to different reflectance classes. We use machine learning techniques to train such a classifier, either from photographs of surfaces or from synthetic images rendered under photographically-acquired illumination.

Chapter 6 considers the design of an effective reflectance classifier in more detail. We consider the choice of machine learning techniques and the selection of specific image statistics as classification features. We analyze the effects of surface geometry on image statistics and describe a method to classify surfaces of different known geometry according to their reflectance. We also examine the robustness of our reflectance classification techniques to incorrect geometry estimates. This chapter includes multiple examples of classifiers applied to both synthetic images and real photographs. When geometry is known in advance and reflectance properties are homogeneous across the surface, the accuracy of our classification algorithms rivals that of the human visual system.

The concluding chapter summarizes the contributions of the thesis in more detail and proposes a number of avenues for future research. This chapter also relates our work on reflectance and illumination to the broader context of material recognition.

Chapter 2

Background

This chapter combines material from physics, computer graphics, computer vision, and human vision as relevant to this thesis. We start by defining surface reflectance (Section 2.1.1) and discussing the reflectance properties of real-world materials (Section 2.1.2). We then describe previous work on recognition of reflectance properties from images (Section 2.2). Finally, we discuss studies of the human visual system's ability to recognize surface reflectance properties under unknown illumination (Section 2.3).

■ 2.1 Physics of Reflectance

■ 2.1.1 Definition of Reflectance

The bidirectional reflectance distribution function (BRDF) of an opaque surface patch defines its reflectance by specifying what proportion of the light incident from each possible illumination direction is reflected in each possible observation or view direction [39]. Figure 2.1 shows a surface patch with normal $\mathbf{N}$ illuminated by a directional light source in direction $\mathbf{S}$ and observed by a viewer in direction $\mathbf{V}$. In a three-dimensional world, two angles are necessary to uniquely specify the illumination direction $\mathbf{S}$ and two more to specify the view direction $\mathbf{V}$. The BRDF is therefore a function of four continuous angular variables. We denote it by $f(\theta_i, \phi_i; \theta_r, \phi_r)$, where θ_i and θ_r are the angles of $\mathbf{S}$ and $\mathbf{V}$, respectively, from the surface normal $\mathbf{N}$, and ϕ_i and ϕ_r are their respective azimuthal angles. This function is defined for θ_i and θ_r in the range $[0, \pi/2]$ and for ϕ_i and ϕ_r in the range $(-\pi, \pi]$.[1] Because surface radiance depends linearly on the amount of light incident from every direction, the BRDF of a surface patch

[1] To define the BRDF precisely, we introduce several radiometric terms following the formulation of Nicodemus [70]. Light incident on a surface is typically measured in terms of irradiance, the power per unit surface area of radiant energy ($\frac{W}{m^2}$) . Light reflected by a surface in a particular direction is measured in terms of radiance, or power per unit foreshortened area emitted into a unit solid angle ($\frac{W}{sr \cdot m^2}$); foreshortened surface area is equal to the actual surface area times the cosine of θ_i. Irradiance corresponds to the concept of image brightness, and radiance to scene brightness. If one takes a photograph of a scene, the irradiance of a point on the film is proportional to the radiance of the corresponding point in the scene. The BRDF $f(\theta_i, \phi_i; \theta_r, \phi_r)$ is the ratio of the reflected radiance in a particular direction to incident irradiance from a differential solid angle centered on the incident direction. Since irradiance has units $\frac{W}{m^2}$ and radiance has units $\frac{W}{sr \cdot m^2}$, the BRDF has units $\frac{1}{sr}$ and can take on values from 0 to ∞.

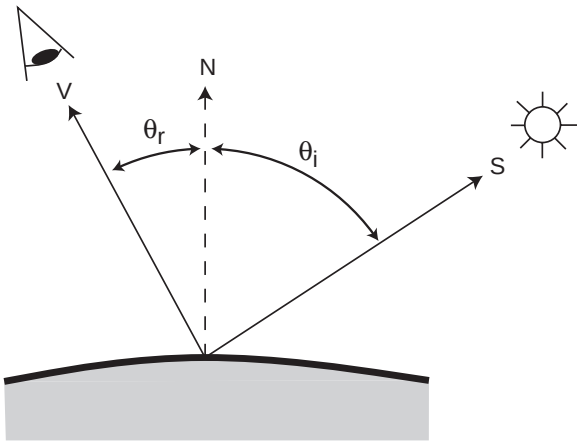


Figure 2.1. A surface patch with normal **N**. The BRDF is a function of light source direction **S** and view direction **V**.

determines its appearance under any illumination distribution.

Reflectance also depends on the wavelength of the incident light. Because most materials reflect light of a given wavelength at the same wavelength, one can describe the effect of color on reflectance by writing the BRDF as a function of an additional variable representing wavelength.[2] The intensity and polarization of light reflected by a surface both depend on the polarization of the incident light. One could capture these dependencies in the BRDF as well, although this is rarely done in practice because polarization effects are generally minor in typical real-world settings.[3] Likewise, one could add a temporal variable to the BRDF for phosphorescent surfaces that absorb incident radiation and then emit radiation after a delay.

The expression of reflectance as a BRDF presupposes an opaque surface. Many real-world materials exhibit some degree of translucency, meaning that light incident at one point on the surface may be emitted at nearby points. Skin, milk, and wax are all highly translucent. One can capture their appearance properties with a bidirectional scattering-surface reflectance distribution function (BSSRDF), a generalization of the BRDF that specifies a proportionality constant dependent on incident and exitant locations as well as directions. The BSSRDF is a function of two positions on the surface as well as two directions in the three-dimensional world, so it depends on eight spatial variables. If the surface is homogeneous and isotropic, one need only specify the distance between the points of incidence and exitance, rather than their locations. Recent computer graphics worked has used this simplified form of the BSSRDF for rendering purposes [50, 51]. This thesis focuses on reflectance as described by the BRDF,

[2] If the incident and reflected wavelengths differ, as for fluorescent surfaces, one must add two wavelength variables to the BRDF.

[3] To capture polarization effects, one would augment the BRDF with two additional binary-valued variables, representing incoming and outgoing polarization. Each of these variables takes on values corresponding to horizontal and vertical polarization.

although the experiments described in Chapter 6 include materials with some degree of translucency.

The BRDF is defined locally, for an infinitesimal patch of a surface. It may vary from one point on a surface to another. Two types of reflectance variation are commonplace. The first occurs at a boundary between two surfaces or between distinct materials within a surface. In these cases, reflectance changes help distinguish two or more different materials. The second involves regular variation of reflectance within a surface, associated with surface texture. In this case, the patterns of variation themselves are an important characteristic of the surface.

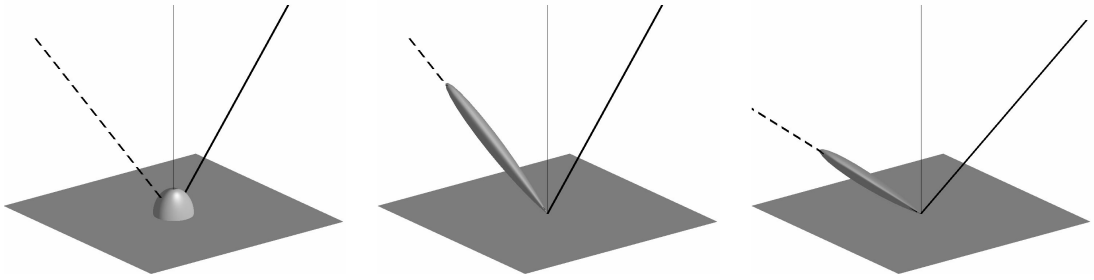
While the computer graphics community tends to regard reflectance and texture as complementary properties [26], material recognition tasks may demand that they be considered jointly. Whether a texture results from fine-scale geometry or from actual reflectance variation, it will have a different appearance when viewed from different angles. A Bidirectional Texture Function (BTF) captures these properties by specifying the two-dimensional texture visible from each viewing angle for each lighting angle [20]. One might model a BTF as a random field of BRDFs. The distinction between reflectance and texture is a matter of scale; as one views a surface from increasing distance, fine scale geometry variations will no longer be resolvable, but they will influence the measured surface BRDF. Although this thesis focuses on reflectance recognition, we further discuss the relationship between recognition of surface texture and reflectance in Sections 5.1.3 and 7.2.5.

■ 2.1.2 Reflectance Properties of Real-World Materials

Maxwell's equations impose two constraints on the BRDF of a passive surface. First, the BRDF must obey energy conservation or *normalization*; for any illumination, total reflected energy must be less than or equal to total incident energy. Second, the BRDF must satisfy the Helmholtz reciprocity principle, which guarantees symmetry between incident and reflected directions. Reciprocity requires that $f(\theta_i, \phi_i; \theta_r, \phi_r) = f(\theta_r, \phi_r; \theta_i, \phi_i)$.

Although any function of four variables satisfying the reciprocity and normalization properties constitutes a physically realizable BRDF, some reflectances are much more common than others in the real world. For example, the reflectance of this paper is more common than that of a particular point in a hologram. Visual reflectance estimation is feasible partly because physical materials tend to produce certain types of BRDFs. In other words, the frequency distribution of surface BRDFs in the real world is not uniform.

A great deal of research has focused on approximating common BRDFs by models with only a few free parameters. These parameterized models play an important role in computer graphics, where they are used to implement efficient shading algorithms that can be effectively controlled by a user [39]. The graphics and applied physics literatures include models derived from the physics of light reflection as well as models designed empirically to fit experimental BRDF data or to produce appealing renderings.



(a) (b) (c)

Figure 2.2. These diagrams illustrate BRDFs by showing the distribution of emitted radiation for a particular incident direction. The incident illumination direction is indicated by the thick solid line, which can be uniquely specified by the two angles θ_i and ϕ_i. The curved surface is a plot of the outgoing radiance for all directions over the hemisphere spanned by θ_r and ϕ_r. This is a two-dimensional slice of a BRDF. The distance from the origin to this surface in any direction is proportional to the reflected radiance in that direction. The thin solid line indicates the surface normal, while the thick dashed line indicates the direction of ideal specular reflection given the incident illumination direction. (a) Lambertian BRDF. (b) Specular BRDF, described by the Ward model, for light incident at 45° to the normal. (c) The same specular BRDF, for light incident at 60° to the normal. The illustrations of this figure and Figure 2.4 were inspired by Rusinkiewicz [91].

These studies have focused on two general reflectance phenomena, diffuse and specular reflectance.

Diffuse reflectance is associated with matte surfaces such as plaster or uncoated paper. An ideal diffuse, or Lambertian, surface has equal radiance in all directions regardless of the incident light direction. Matte paint consisting of reflective patches suspended in a clear matrix approximates a Lambertian reflector, because light will emerge after multiple reflections in a more or less random direction. An ideal Lambertian reflector that emits a constant proportion ρ_d of the incident energy and absorbs the rest has a constant BRDF of the form

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi}, \tag{2.1}$$

where $0 \leq \rho_d \leq 1$. Figure 2.2(a) illustrates this BRDF.

Real diffuse reflectors deviate from this ideal behavior. Oren and Nayar [69, 74, 75] and Koenderink *et al.* [54] studied locally Lambertian surfaces with a fine-scale texture consisting of V-shaped grooves or spherical pits. From a typical viewing distance, such a surface appears homogeneous because the spatial variations due to texture are not visible. The fine-scale geometry has a direction-dependent effect on average reflectance, however, so the measured BRDF is not Lambertian. The BRDF models derived from these physical assumptions provide an accurate fit to measured BRDFs of natural surfaces such as plaster, chalk, and clay.

Specular reflectance is typified by a mirror. An ideal specular surface reflects all energy in an incident light ray such that the incident and reflected directions are bisected

by the surface normal. Such a surface has a BRDF

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\delta(\cos\theta_i - \cos\theta_r)}{-\cos\theta_r}\delta(|\phi_r - \phi_i| - \pi). \tag{2.2}$$

Specular surfaces such as metals also typically exhibit some fine-scale variation in surface geometry. This roughness causes the specular reflection of a point source to be distributed in a small region around the ideal mirrored direction, as shown in Figure 2.2(b) and (c).

Diffuse and specular reflectance differ fundamentally in two ways. First, specularities are typically visible over a narrow view angle, so that specular reflection tends to be sharper than diffuse reflection. Second, even a rough specular surface differs from a diffuse surface in the direction of dominant reflectance. As a fixed observer views a surface illuminated by a moving light source, the diffuse component will peak for illumination normal to the surface, while the specular component will peak when the surface normal bisects the illumination direction and the view direction.

A number of parameterized models of specular reflectance take into account the width of the specular lobe. The earliest of these, still common in computer graphics, is the Phong model [8, 79], which uses one parameter to describe the strength of the specular reflectance and another to specify surface smoothness, which is inversely correlated to roughness and the width of the specular lobe. The Phong model leads to a computationally efficient implementation but is not constrained by the reciprocity and normalization requirements of a passive physical surface.

Ward proposed a variant of the Phong model that largely overcomes these problems [56, 118]. The BRDF for the specular component of the Ward model takes the form

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \rho_s \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \frac{\exp(-\tan^2\delta/\alpha^2)}{4\pi\alpha^2}, \tag{2.3}$$

where δ is the angle between the surface normal and a vector bisecting the incident and reflected directions, the so-called "half-angle vector" (Figure 2.3). The specular component is spread out about the ideal specular direction in a Gaussian distribution with standard deviation α. The larger α is, the wider the specular lobe and the blurrier the specular reflection. The other free parameter, ρ_s, specifies the proportion of incident energy reflected by the specular component.[4]

Most surfaces reflect light by several physical mechanisms, including both specular and diffuse reflection. BRDFs are therefore typically modeled as a sum of a specular and a diffuse component. For example, the isotropic Ward model combines the specular

[4]The $4\pi\alpha^2$ normalization factor in the denominator ensures that the total energy reflected remains independent of α. This normalization factor, computed using the small-angle approximation $\tan x \approx x$, is accurate as long as α is not much greater than 0.2 [118]. The $\frac{1}{\sqrt{\cos\theta_i \cos\theta_r}}$ term satisfies the reciprocity principle while ensuring that the amount of energy reflected does not depend on the position of the light source.

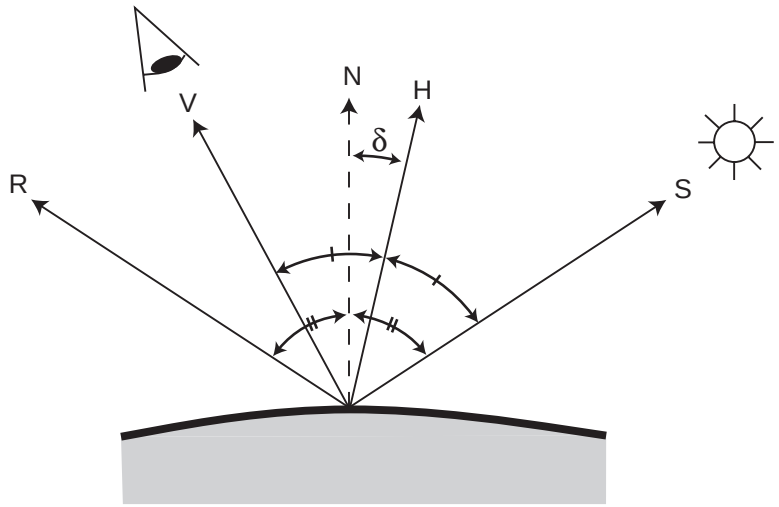


Figure 2.3. Geometry used to define the specular component of the Ward model. As in Figure 2.1, **N** is the surface normal, **S** is the light source direction, and **V** is the view direction. The half-angle vector **H** bisects **S** and **V**. The direction of ideal specular reflection **R** is such that **N** bisects **R** and **S**.

component described by Equation (2.3) with the Lambertian component described by Equation (2.1):

$$f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi} + \rho_s \frac{1}{\sqrt{\cos\theta_i \cos\theta_r}} \frac{\exp(-\tan^2\delta/\alpha^2)}{4\pi\alpha^2}. \tag{2.4}$$

The sum $\rho_d + \rho_s$ specifies the fraction of incident energy reflected by the surface, so normalization requires $\rho_d + \rho_s \leq 1$. Figure 2.4 illustrates the effect of each of the Ward model parameters on the BRDF.

Many common materials have isotropic reflectance functions with no particular orientation [44]. Such a BRDF can be written as a function of the difference between the incident and reflected azimuthal angles, $\phi_i - \phi_r$, rather than the actual values of ϕ_i and ϕ_r. Certain materials, such as brushed aluminum or fine-grained wood, have anisotropic BRDFs; rotating a patch of such a surface about its own normal may change its appearance. A number of parameterized reflectance models, including a more general form of the Ward model, can capture anisotropic specular reflection.

The empirically derived Ward model fits measured data well for certain materials, such as latex paints [118]. However, it fails to capture a variety of real-world reflectance properties. Some of these properties, including Fresnel effects, specular spikes, and off-specular peaks, are captured by more complicated models derived directly from physical principles [39]. Most physically-based models in the computer graphics and applied physics literatures assume a simple but random surface micro-structure, such as a flat substrate with randomly oriented V-shaped grooves. They predict surface reflectance based on geometrical considerations and physical optics. Perhaps the most physically complete model to date is that of He, Torrance, Sillion, and Greenberg (HTSG) [42],

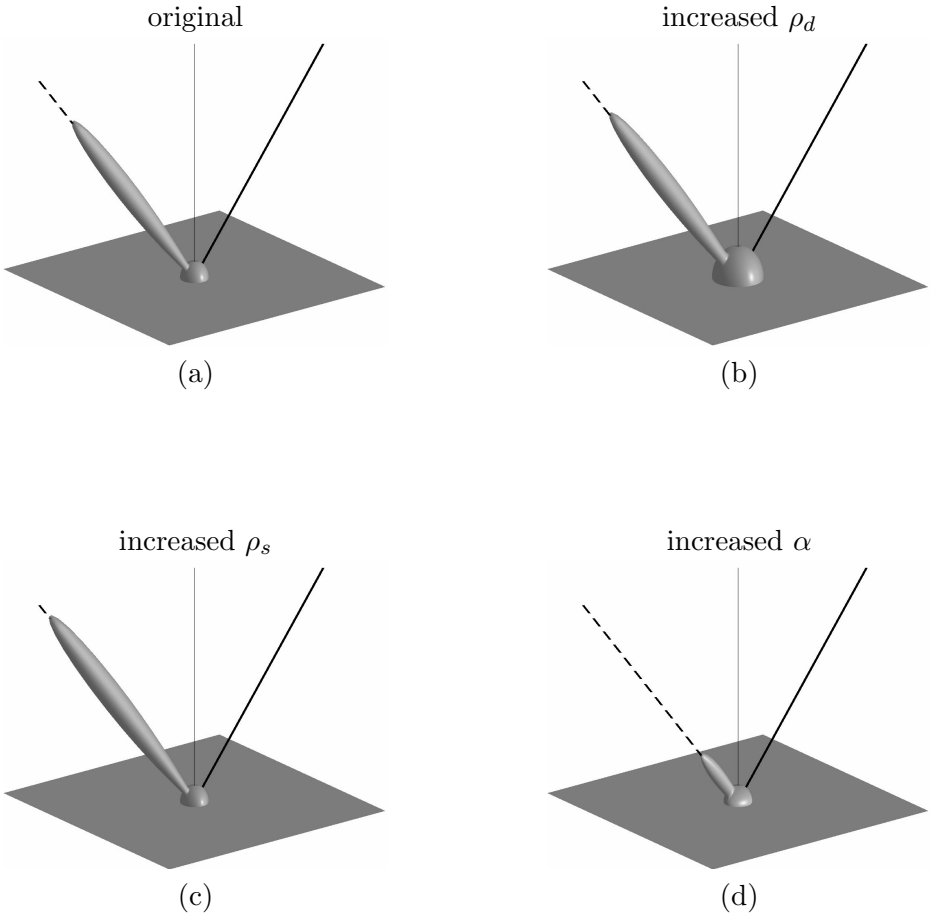


Figure 2.4. Effects of the Ward model parameters on the model BRDF. Emitted radiance distributions are illustrated for light incident at 45° to the normal. (a) BRDF with $\rho_d = .5$, $\rho_s = .05$, $\alpha = .05$. (b) Same as (a), except that ρ_d has been increased to .9. (c) Same as (a), except that ρ_s has been increased to .6. (d) Same as (a), except that α has been increased to .1.

parameterized by a complex index of refraction, a spectral reflectivity function, and two values which characterize surface roughness as an RMS deviation from the plane and an autocorrelation length. This model has been verified experimentally for a wider class of surfaces than the Ward model. However, it still cannot accurately model anything near the full range of real-world surface reflectances. The HTSG model, like most comparable physically-based models, is far more analytically complex and computationally expensive than the Ward or Phong models. As a result, it is almost never used for rendering applications.

In general, parameterized reflectance models capture a range of common reflectances, but they fail to capture many of the reflectances encounterd in the real world. Even within the range of reflectances they accommodate, they do not describe the relative frequency with which different reflectances are observed.

■ 2.2 Previous Approaches to Reflectance Estimation

The importance of reflectance models in computer graphics has motivated several researchers to develop image-based reflectance estimation techniques. The majority of these techniques assume a controlled laboratory setting similar to that employed by traditional gonioreflectometers, devices that measure a BRDF by illuminating a material sample with a movable point light source and measuring its radiance in every direction for each illumination direction. To accelerate the task of BRDF measurement, Ward [118] developed an "imaging gonioreflectometer," which for each illumination direction captures radiance in all directions simultaneously as a single image of a silvered hemisphere taken with a fisheye lens. Marschner *et al.* [66] developed a laboratory technique for measuring BRDFs from multiple images of a curved surface such as skin; instead of assuming a flat sample, they take advantage of the additional information provided by known surface curvature. Debevec *et al.* [24] also acquired BRDFs of skin under controlled point source illumination, using color space techniques to separate specular and diffuse reflections. Tominaga *et al.* [110] present a method for estimating Phong model parameters from an image of a uniform cylindrical surface. They require illumination by a point light source, although they estimate the exact location of the light source from image data. Sato *et al.* [92] as well as Marschner [65] develop similar techniques that accommodate spatial variation in the diffuse reflectance coefficient as well as a more general geometry acquired through laser range scanning. Love [60] measured reflectance from multiple photographs under illumination by the sun and sky, using a model that specifies the amount of light incident from each direction at a particular location, season, and time of day. None of these methods recover reflectance from photographs acquired in the real world under unknown lighting conditions.

Several authors have recently estimated both illumination and reflectance from a set of photographs under real-world illumination [9,73,86,123,124]. They all assume known geometry and a Phong- or Ward-like specular plus diffuse reflectance model. They all apply an iterative estimation technique to deduce both illumination and reflectance, matching resynthesized images to the observed images. These techniques assume that enough information is available to guarantee that this optimization will converge to a unique solution. Multiple combinations of illumination and reflectance can explain the entire reflected light field, even when reflectance is restricted to a Phong- or Ward-like model (Section 3.2.1). One must therefore introduce additional information to guarantee that the joint optimization will converge to a unique solution. All of these approaches require a complete geometric model of the surrounding scene and a reasonable initial estimate for illumination. Yu and Malik [124] measure the illumination incident on the scene from each direction photographically, constructing an illumination map such as those described in Chapter 4. Yu *et al.* [123] explicitly specify the location of primary light sources. Ramamoorthi and Hanrahan [86] assume the presence of a point source in a known direction. Nishino *et al.* [72,73] introduce a regularization term on illumination motivated by computational efficiency, and also assume that all illumination has the same color and that color images of the surface are available. Boivin and

Gagalowicz [9], the only authors to estimate both illumination and reflectance based on a single photograph, rely on human interaction in the estimation process.

We wish to avoid estimating illumination explicitly by characterizing it statistically. In this sense, our approach has something in common with that of Weiss [119], who decomposed a set of images of the same scene under different illumination into intrinsic "illumination" and "reflectance" images by assuming statistics on the illumination images. We also draw on Freeman's observation that one can select between different reflectance functions that perfectly explain an image by integrating the posterior probability of each reflectance over possible illuminations [37]. Freeman demonstrated that this "generic viewpoint" approach favors image explanations that are relatively insensitive to changes in illumination and reflectance parameters.

■ 2.3 Reflectance Estimation in Human Vision

The human ability to recognize surface reflectance properties in real-world circumstances provides motivation for our investigation into computational reflectance estimation problems. Human vision researchers have conducted a variety of psychophysical and physiological experiments to investigate the algorithmic strategy and efficacy of the human reflectance estimation process.

Most of this research has assumed Lambertian surfaces, focusing on estimation of diffuse surface reflectance (albedo) and color. A gray surface under bright illumination may have exactly the same luminance as a white surface under dim illumination. Humans possess a surprising ability to recognize the intrinsic albedo of surfaces under realistic and varied illumination conditions. This ability, termed *lightness constancy*, depends on the spatial arrangement of luminances within a scene.

Vision researchers have studied the lightness constancy problem since the 19th century. Herring emphasized low-level visual effects that could correspond to basic retinal mechanisms, such as the fact that the perceived reflectance of an image region depends on the luminance of its immediate surroundings [1]. Helmholtz, on the other hand, described lightness constancy as a high-level process of unconscious inference, whereby an observer deduces the most likely explanation of a visual image by drawing upon prior experience [1]. More recently, psychophysicists have found evidence for a variety of mid-level lightness perception mechanisms based on image features such as contours, edge junctions, and local brightness distributions [1, 38]. These mechanisms do not require a high-level understanding of the image, but they may be viewed as statistical estimation algorithms in that their success depends on statistical assumptions about the real visual world. Several researchers have focused on the development of computational approaches to estimating surface albedo [10, 55, 64].

Two surfaces of different intrinsic colors may produce exactly the same image color when viewed under differently colored light sources. The human visual system also exhibits approximate *color constancy*, the ability to recognize intrinsic surface color under a wide variety of viewing conditions [11, 29]. A number of authors have proposed

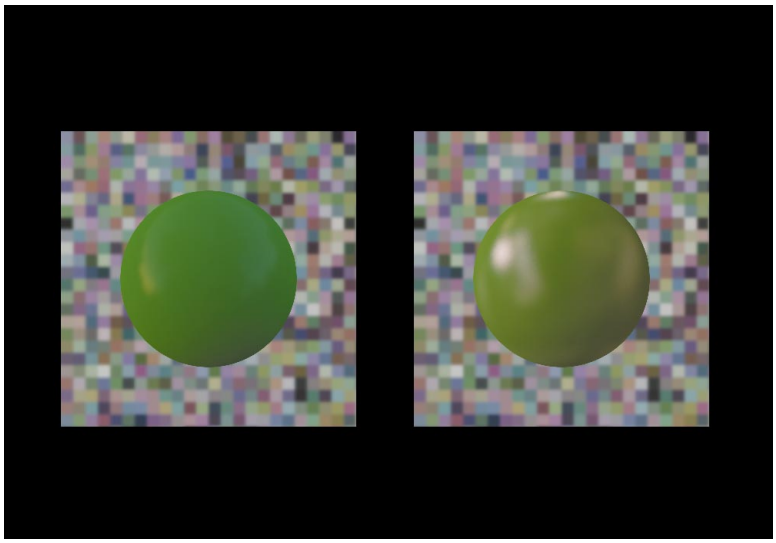


Figure 2.5. A sample screen from the matching experiment. Subjects adjusted two parameters of the sphere on the right until it appeared to be made of the same material as the sphere on the left. The two spheres pictured here have different reflectance properties.

computational techniques to achieve color constancy [12, 14, 28, 35, 63], some of which are used for color balance in photographic systems.

Although psychophysicists have long been aware that humans can recognize non-Lambertian reflectance properties reliably, investigation of the extent of this ability and the mechanisms underlying has been limited. Beck [5] observed that eliminating all the highlights in an image of a glossy vase could make the entire vase look matte, suggesting that gloss perception involves propagation of local cues over a surface. However, Beck and Prazdny [6] performed further experiments suggesting that gloss perception involves responses to low- and mid-level visual cues rather than high-level inference that the surface is reflecting light specularly.

Pellacini *et al.* [77] established a "perceptually uniform gloss space." They applied multi-dimensional scaling to human judgments of gloss differences in order to establish a nonlinear reparameterization of the space spanned by the Ward model. Equal distances in this reparameterized space correspond to equal perceptual differences.

We carried out a series of experiments to measure the human ability to match non-Lambertian reflectances under unknown real-world illumination conditions. This experimental work, which involved a collaboration with Roland Fleming, is summarized in the present section as background to the remainder of the thesis. A more detailed account has been published elsewhere [33,34]. We wished to ascertain the accuracy with which humans could judge gloss properties from single images of isolated surfaces, in the absence of motion, stereo, or contextual information. We also wished to determine under what range of illuminations humans can perform the task, and what image cues they use to solve it.

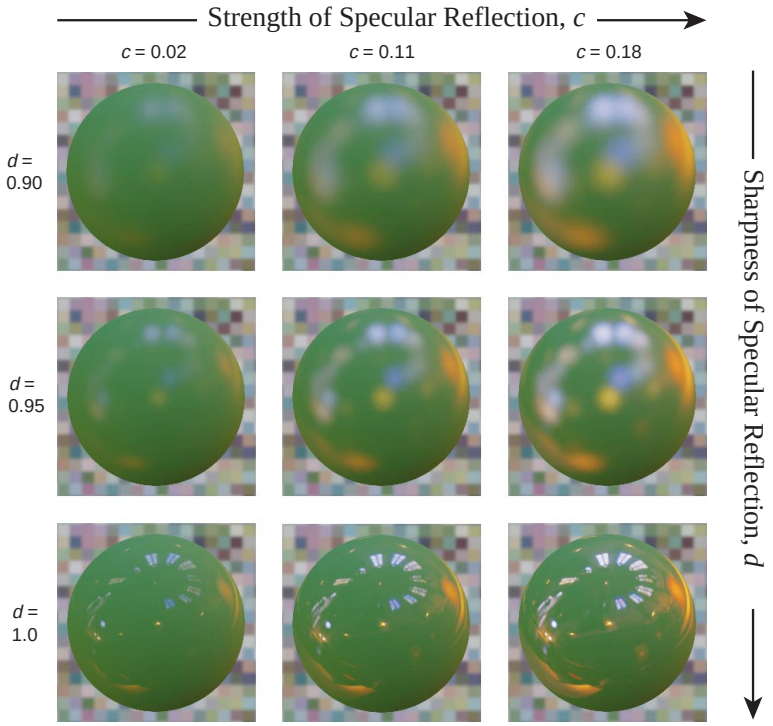


Figure 2.6. Grid showing range of reflectance properties used in the experiments for a particular real-world illumination map. All the spheres shown have an identical diffuse component. In Pellacini's reparameterization of the Ward model, the specular component depends on the c and d parameters. The strength of specular reflection, c, increases with ρ_s, while the sharpness of specular reflection, d, decreases with α. The images were rendered in *Radiance*, using the techniques described in Appendix B.

To investigate these issues, we used the experimental setup pictured in Figure 2.5. The subject was presented with two images of spheres rendered by computer under different illuminations. The subject was instructed to adjust two reflectance parameters of one sphere (the "Match" sphere) until it appeared to be made of the same material as the other sphere (the "Test" sphere).

The sphere reflectances were restricted to the space covered by the Ward model, with the diffuse reflectance fixed. Subjects were given two knobs, corresponding to two parameter values in Pellacini's reparameterization of the Ward model, with which to navigate in this space (Figure 2.6). Because the illuminations of the two spheres differed, no reflectance parameter setting would achieve identical images. Instead, subjects tried to adjust the Match image so that it might represent the same sphere as the Test, but viewed in a different location. All spheres were shown over the same checkered background, and the illumination maps used to render the spheres were not disclosed to the subjects.

We used a variety of illuminations, both real-world and synthetic, to render the Test spheres. The real-world illuminations consisted of eight photographically-acquired

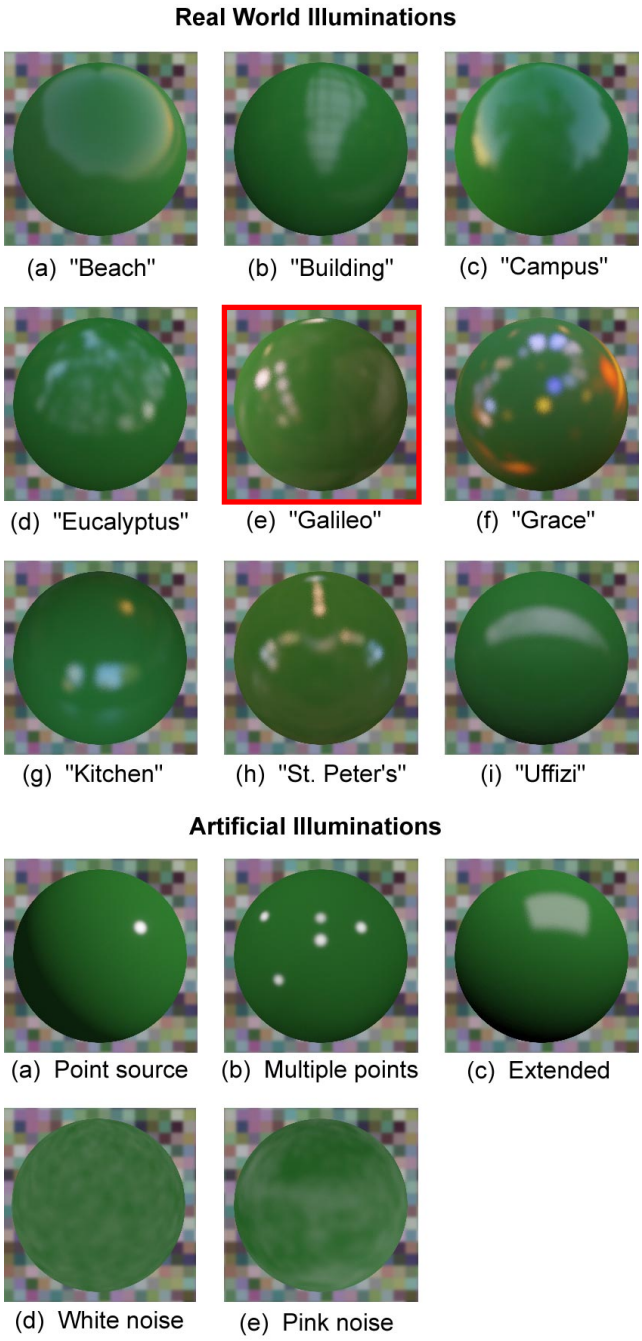


Figure 2.7. Spheres rendered under each of the illuminations used in the experiments. All spheres have the same surface reflectance. Real-world illumination (e), highlighted with a perimeter, was the standard Match illumination.

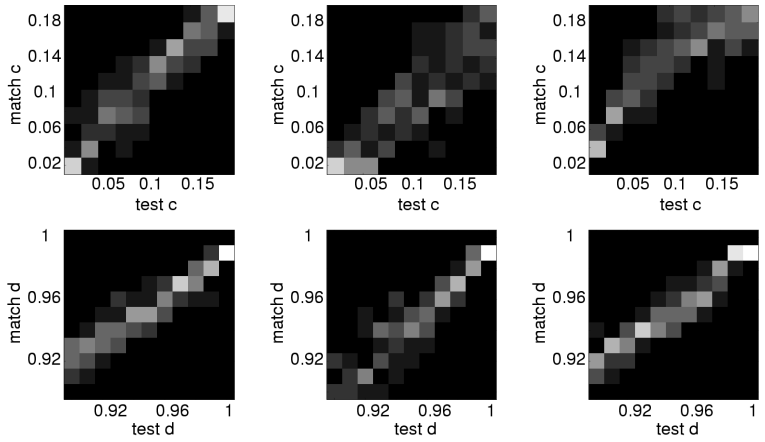


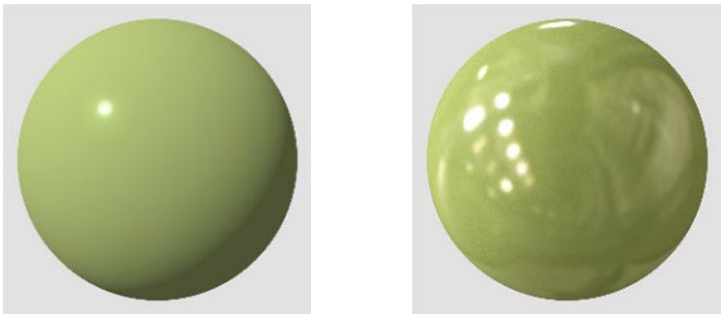
Figure 2.8. Match values plotted as a function of Test values for individual subjects. Graphs in the top row are matches for the strength of specular reflection, c; graphs in the bottom row are for sharpness of specular reflection, d. The gray value represents the density of responses for a given Test value. Thus, if subjects always responded with the same Match value to a given Test value, the corresponding sample is white; the rarer the response, the darker the gray. The graphs in (a) are subject RF's matches for spheres under the "St. Peter's" illumination; (b) shows RA's matches for spheres under the "Eucalyptus" illumination; (c) shows subject MS's matches for spheres under the "Grace" illumination.

illumination maps due to Debevec [24], described further in Section 4.2. The synthetic illuminations included a single point source, multiple point sources, a single extended rectangular source, Gaussian white noise, and Gaussian noise with a $1/f$ amplitude spectrum (pink noise).[5] The Match sphere that the subject adjusted was always viewed under the same real-world illumination. Figure 2.7 shows a sphere of a fixed surface reflectance rendered under each of the illuminations used in the experiments.

These experiments resulted in several findings:

- For spheres viewed under photographically-acquired real-world illumination, humans perform the task with high accuracy. The reflectance matching task is underconstrained — if one makes no assumptions about illumination, a range of reflectance parameters could produce the observed images. In practice, however, subjects' parameter settings for the Match sphere correspond closely to the parameters used to render the Test sphere. This serves as a demonstration of feasibility for our goal of reflectance estimation from a single image under unknown

[5] We generated the white noise illumination map by summing spherical harmonics whose coefficients up to a fixed order were chosen from independent Gaussian distributions of equal variance. For the pink noise, the spherical harmonic coefficients were chosen independently from Gaussian distributions with standard deviation inversely proportional to the spherical harmonic order, which is analogous to frequency. This process produces noise whose power spectrum is similar to that of many real-world illuminations and natural images (see Chapter 4).



(a) (b)

Figure 2.9. (a) A shiny sphere rendered under illumination by a point light source. (b) The same sphere rendered under photographically-acquired real-world illumination. Humans perceive reflectance properties more accurately in (b).

illumination. Figure 2.8 shows example data from the matching experiments for individual subjects under individual illumination conditions.

- Subjects estimate reflectance more consistently and accurately under real-world illumination than under simple synthetic illuminations such as a point light source or Gaussian noise. Figure 2.9 shows two identical shiny spheres, one rendered under a point light source, and the other under photographically-acquired real-world illumination. Even though the point source illumination is "simpler," the perception of realistic reflectance properties is much stronger under complex real-world illumination. This result is consistent with the observation that computer graphics scenes rendered under photographically-acquired illumination (image-based lighting) appear more realistic than those rendered under traditional simple illumination [23].

- Even though subjects match reflectances accurately under unknown real-world illumination, they exhibit biases dependent on illumination. These biases are statistically significant and are similar from one subject to the next. In other words, certain illumination maps make surfaces viewed under those illuminations appear to have higher or lower specular contrast or distinctness.

All of these observations suggest that subjects use stored assumptions about illumination in estimating reflectance. These assumptions seem to be valid for most real-world illumination conditions, but less so for synthetic illuminations. Experimental work has not yet pinpointed what assumptions the human visual system makes about illumination, or which image features actually cue humans to reflectance properties. One of the goals of our computational work is to determine what information different image features capture about reflectance under real-world illumination.

Nishida and Shinya found that humans failed to match reflectance accurately for surfaces of different geometry rendered under point source illumination [71]. They found

that subjects' matches related strongly to luminance histograms of the observed images. Their results also suggest that human reflectance recognition depends on stored assumptions about the real world. For arbitrary illumination and geometry, these assumptions may not be valid.

■ 2.4 Summary and Discussion

Opaque surfaces possess a wide range of reflectance properties, described by the bidirectional reflectance distribution function. Although a number of authors, particularly in the computer graphics community, have recently developed methods to recover surface reflectance from images, they have assumed either that reflectance is known in advance or that enough information is available about the scene to explicitly recover both illumination and reflectance. By contrast, humans display an ability to recognize reflectance properties from an image of a surface under unknown illumination, as long as that illumination is somehow typical of the real visual world. In the following chapter, we describe the process of image formation from reflectance and illumination mathematically, and show how the problem of recovering reflectance under unknown illumination is underconstrained. Later chapters discuss the statistical regularity of real-world illumination and the relevance of this regularity to the reflectance recognition problem.

Chapter 3

Problem Formulation

The illumination, reflectance, and geometry of a surface determine its appearance from any viewpoint. While decades of computer graphics research have focused on rendering images efficiently given this information, the process is conceptually straightforward. Inferring surface reflectance from one or more images under unknown illumination is more difficult. More than one combination of illumination and reflectance could explain the observed data, so the problem is underconstrained. We wish to select the most likely reflectance properties given the available image data and available prior information about the real world. In this chapter, we pose the reflectance estimation problem mathematically as a blind inverse problem. We also consider several simplified formulations of this problem. Table 3.1 defines the notation of this chapter.

We will assume in this chapter, as in most of this thesis, that surface geometry is known in advance. Chapter 6 discusses relaxation of this assumption.

■ 3.1 The Forward Problem: Rendering

■ 3.1.1 The General Case

Given the BRDF of a surface, the position of a viewer, and the illumination incident at a point on the surface from each direction, we can compute the reflected radiance from that point to the viewer's direction by summing the contributions of incident light from all directions. To make this statement more precise, consider a small surface patch with normal $\mathbf{N}$ (Figure 3.1). We define angles with respect to $\mathbf{N}$ as in Section 2.1.1 and consider a distant observer in direction (θ_r, ϕ_r). If $L(\theta_i, \phi_i)$ gives the radiance of incident illumination from direction (θ_i, ϕ_i), the total reflected radiance B of the surface patch in the direction (θ_r, ϕ_r) is given by

$$B(\theta_r, \phi_r) = \int_{\phi_i=0}^{2\pi} \int_{\theta_i=0}^{\pi/2} L(\theta_i, \phi_i) f(\theta_i, \phi_i; \theta_r, \phi_r) \cos\theta_i \sin\theta_i \, d\theta_i \, d\phi_i, \tag{3.1}$$

θ_i, θ_i'	Incident elevation angle in *global, local* coordinates
ϕ_i, ϕ_i'	Incident azimuthal angle in *global, local* coordinates
θ_r, θ_r'	Reflected elevation angle in *global, local* coordinates
ϕ_r, ϕ_r'	Reflected azimuthal angle in *global, local* coordinates
γ, β	Surface normal parameterization (*elevation* and *azimuthal*) angles
$R_{\gamma,\beta}$	Rotation operator for surface normal (γ, β)
L	Incoming radiance (illumination)
B	Reflected radiance, either as a function $B(\gamma, \beta; \theta_r', \phi_r')$ of reflected direction and surface normal or as a function $B(\theta_r', \phi_r')$ of reflected direction only
f	Surface BRDF
$\tilde{f}$	BRDF multiplied by cosine of incident elevation angle and defined as 0 for incident elevations larger than $\pi/2$
$\hat{f}$	Estimated BRDF
S^2	A unit sphere, used as an integration region
$d\omega$	Differential area element on the sphere

Table 3.1. Notation used in this chapter. Our notation follows that of Ramamoorthi and Hanrahan [86], but we use r rather than o subscripts to denote reflected (outgoing) radiation, γ rather than α to denote surface normal elevation angle, and f rather than ρ to denote the BRDF. We use $\hat{f}$ to denote the estimated BRDF, while Ramamoorthi and Hanrahan use $\hat{\rho}$ to denote a modified transfer function similar to our $\tilde{f}$.

where f is the surface BRDF.[1] Equation (3.1) is a form of the Radiance Equation, which serves as the basis for rendering algorithms in computer graphics[2] [39,56,118]. Because image irradiance is proportional to scene radiance, Equation (3.1) gives the image intensity associated with the surface patch [44]. Ideally, the integration of Equation (3.1) should be carried out separately for each wavelength of light. One can approximate the results by performing one integration for each of three color channels.

■ 3.1.2 Reflectance as a Convolution

Equation 3.1 describes a linear relationship between the illumination of a surface point and the amount of light it reflects in each direction. If we know the geometry of an entire surface, we can determine a linear relationship between the amount of light incident on each surface point from each direction and the irradiance of each point of the resulting image. In fact, one can express the relationship between illumination and reflected

[1] The $\cos\theta_i$ term in this equation accounts for the fact that the radiance used to measure the illumination $L(\theta_i, \phi_i)$ is defined in terms of foreshortened area. The $\sin\theta_i$ term is the standard integration factor for spherical coordinates.

[2] Equation (3.1) assumes that $L(\theta_i, \phi_i)$ is measured near the surface, such that it takes all indirect reflections and blocking effects into account. The equation can be amended to describe the effects of a participating medium such as fog on the observed image, and to account for radiation emitted by a luminous surface.

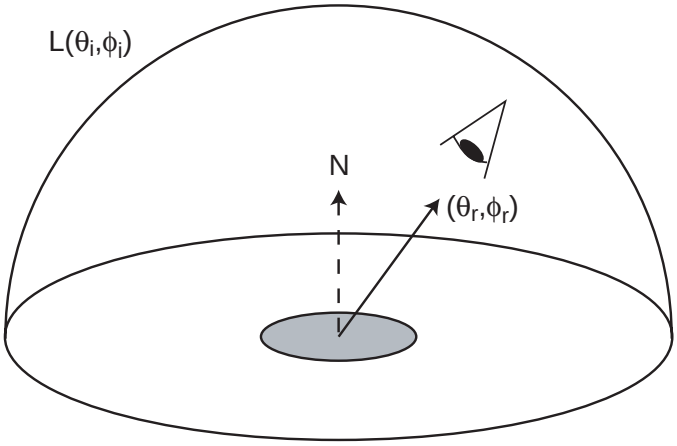


Figure 3.1. A viewer observes a surface patch with normal **N** from direction (θ_r, ϕ_r). $L(\theta_i, \phi_i)$ represents radiance of illumination from direction (θ_i, ϕ_i). The coordinate system is such that **N** points in direction $(0, 0)$.

light as a spherical convolution whose kernel is determined by the BRDF. Although this observation is not novel (*e.g.*, [27,68]), it was recently formalized by Ramamoorthi and Hanrahan [85,86] and by Basri and Jacobs [4].

To express the reflection process as a spherical convolution, we make the following assumptions:

- The surface is curved and convex.
- Sources of illumination, both direct and indirect, are distant relative to the size of the surface.
- The surface is made of a homogeneous material, such that its BRDF is the same everywhere. To simplify the derivations of this section, we will also assume that the BRDF is isotropic.

The distant illumination assumption implies that the amount of light incident from a particular direction is identical at nearby points. In other words, we can imagine that all illumination comes from the inside of an infinitely large sphere centered on the surface of interest. If $L(\theta_i, \phi_i)$ denotes the radiance of illumination incident from direction (θ_i, ϕ_i), then any point whose surface normal lies within the hemisphere centered at (θ_i, ϕ_i) receives illumination $L(\theta_i, \phi_i)$ from that direction. The surface itself occludes illumination from direction (θ_i, ϕ_i) at any surface point whose normal lies outside this hemisphere, giving rise to attached shadows.

Because the surface is convex and curved, we can parameterize it by the global spherical coordinates of its surface normal (γ, β), where $(\gamma, \beta) = (0, 0)$ points vertically upward (Figure 3.2). To simplify the formulas relating illumination to reflected light, we define local coordinates with respect to the surface normal (γ, β), as shown

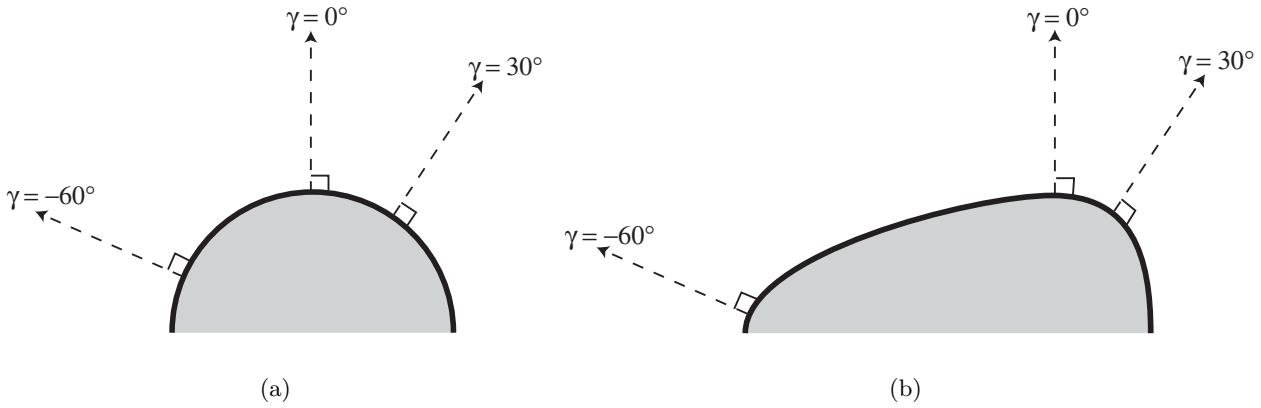


Figure 3.2. Parameterization of surface location by surface normal, for (a) a circular surface and (b) an egg-shaped surface. Only the top halves of each surface are shown. For illustrative purposes, we show two-dimensional surfaces parameterized by a single angle γ ranging from -180° to 180°. In the three-dimensional case, the elevation angle γ ranges from 0° to 180°, while the azimuthal angle β ranges from 0° to 360°.

.

in Figure 3.3. Global coordinates are indicated by unprimed angles, while local coordinates are indicated by primed angles. The BRDF of a surface point is most naturally expressed in local coordinates, because it is defined relative to the surface normal. Under the distant lighting assumption, illumination is more naturally expressed in global coordinates.

For any particular surface normal (γ, β), the local and global coordinate systems are related by rotation. In local coordinates, $(0', 0')$ is the surface normal, corresponding to (γ, β) in global coordinates. We define $R_{\gamma,\beta}$ to be the rotation operator that maps local coordinates to global coordinates. We can decompose this three-dimensional rotation as $R_{\gamma,\beta} = R_z(\beta)R_y(\gamma)$, where $R_z(\beta)$ denotes rotation about the z axis by angle β and $R_y(\gamma)$ denotes rotation about the y axis by angle γ.[3] We can now convert between local and global coordinates with the following formulas:

$$(\theta_i, \phi_i) = R_{\gamma,\beta}(\theta_i', \phi_i') = R_z(\beta)R_y(\gamma)(\theta_i', \phi_i') \tag{3.2}$$

$$(\theta_i', \phi_i') = R^{-1}_{\gamma,\beta}(\theta_i, \phi_i) = R_y(-\gamma)R_z(-\beta)(\theta_i, \phi_i). \tag{3.3}$$

We denote by B the reflected light field, the amount of light reflected in each direction at each point on the surface. In particular, $B(\gamma, \beta; \theta_r', \phi_r')$ is the radiance reflected by the surface patch with normal (γ, β) in a direction (θ_r', ϕ_r') relative to the surface normal. Because the reflected angles are parameterized in local coordinates, the analog

[3]The z axis points vertically upward with direction given by $(0, 0)$ in global coordinates. The y axis is horizontal, pointing in direction $(\pi/2, \pi/2)$. We are assuming an isotropic BRDF; otherwise, we would include three components in this decomposition rather than two.

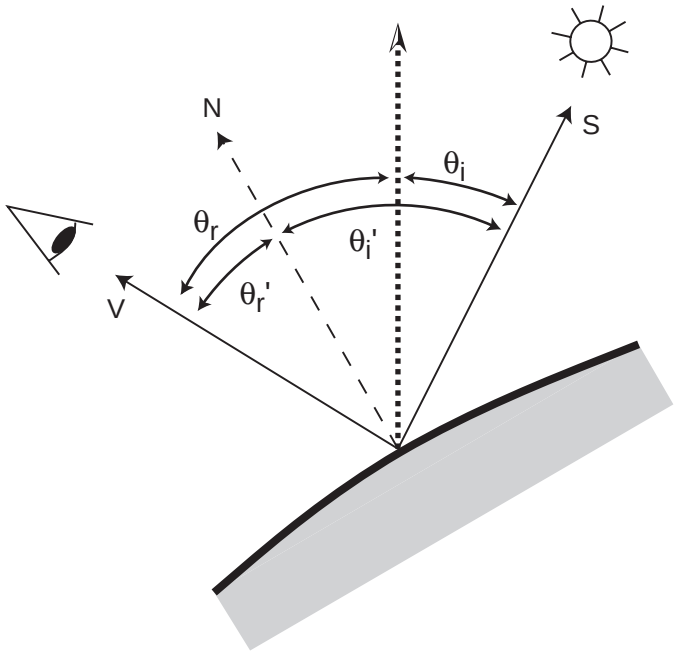


Figure 3.3. Local and global coordinates, for a particular point with surface normal **N**. The local coordinates θ_i' and θ_r' specify the illumination and view directions with respect to the local surface normal, while the global coordinates θ_i and θ_r specify the same directions with respect to a global vertical reference direction. **S** and **V** are the incident and reflected directions, as in Figure 2.1.

of Equation (3.1) for a surface patch with normal (γ, β) is simply

$$B(\gamma, \beta; \theta_r', \phi_r') = \int_{\phi_i'=0}^{2\pi} \int_{\theta_i'=0}^{\pi/2} L(\theta_i, \phi_i) f(\theta_i', \phi_i'; \theta_r', \phi_r') \cos\theta_i' \sin\theta_i' \, d\theta_i' \, d\phi_i'. \tag{3.4}$$

To simplify Equation (3.4), we define a modified reflectance function $\tilde{f}$ as the product of the BRDF and the cosine of the incident elevation angle. We also replace the double integral of Equation (3.4) by integration over the sphere S^2 with respect to a differential area element $d\omega$. The $d\omega$ term replaces the $\sin\theta_i' \, d\theta_i' \, d\phi_i'$ terms of Equation (3.4). Because we will be integrating over the full sphere rather than only the hemisphere surrounding the surface normal, we define $\tilde{f}$ to be 0 for incident elevation angles between $\frac{\pi}{2}$ and π, where the BRDF f is by convention not defined. This leads to

$$\tilde{f}(\theta_i, \phi_i; \theta_r, \phi_r) = \begin{cases} f(\theta_i, \phi_i; \theta_r, \phi_r) \cos\theta_i & \text{if } \theta_i \in [0, \frac{\pi}{2}] \\ 0 & \text{if } \theta_i \in (\frac{\pi}{2}, \pi]. \end{cases} \tag{3.5}$$

We can then write Equation (3.4) as

$$B(\gamma, \beta; \theta_r', \phi_r') = \int_{S^2} L(\theta_i, \phi_i) \tilde{f}(\theta_i', \phi_i'; \theta_r', \phi_r') d\omega.$$

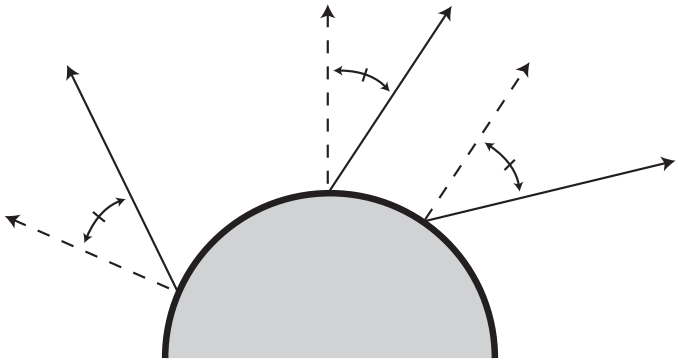


Figure 3.4. The solid arrows indicate a set of reflected rays with identical values of the reflected angle θ'_r in the local coordinate system. Angles in the local coordinate system for a given surface point are measured with respect to the surface normal, shown as a dashed arrow. The surface normals in this figure are identical to those of Figure 3.2(a).

Converting the dependence on incident angles to global coordinates gives

$$B(\gamma, \beta; \theta'_r, \phi'_r) = \int_{S^2} L(\theta_i, \phi_i) \tilde{f} \left(R^{-1}_{\gamma,\beta}(\theta_i, \phi_i); \theta'_r, \phi'_r \right) d\omega. \tag{3.6}$$

If one regards θ'_r and ϕ'_r as fixed, Equation (3.6) is a convolution of L with $\tilde{f}$. This convolution is based on a rotation operator over a spherical data domain, while "standard" convolution involves a translation operator over a Euclidean domain. For a fixed reflected direction in the local coordinate system, the reflected radiance as a function of surface normal is a convolution of the illumination and a two-dimensional cross-section of $\tilde{f}$ corresponding to that reflected direction. Figure 3.4 shows the reflected rays corresponding to a fixed reflected direction in the local coordinate system.

A planar convolution corresponds to multiplication of coefficients in the Fourier transform domain. Likewise, a spherical convolution corresponds to multiplication of coefficients in the spherical harmonic domain. Ramamoorthi and Hanrahan [86] transform B, L, and $\tilde{f}$ to the spherical harmonic domain, where each coefficient of the reflected light field can be written as a product of a coefficient of the illumination and a coefficient of the modified reflectance function $\tilde{f}$. If the BRDF is isotropic, then only a three-dimensional subspace the spherical harmonic coefficients of $\tilde{f}$ contains non-zero coefficients. The reciprocity condition on the BRDF (Section 2.1.2) translates into an additional symmetry property on the coefficients of $\tilde{f}$.

■ 3.1.3 Image Formation

An image is a cross-section through the light field B consisting of a set of rays that converge to a single viewpoint. Figure 3.5 shows two such sets of rays. Figure 3.5(a) illustrates the typical case, in which converging rays form an image on a view plane at a finite distance from the surface. Figure 3.5(b) illustrates the limiting case of an observer at infinite distance. In this case, the rays are parallel.

The rays corresponding to a fixed reflected direction (θ'_r, ϕ'_r) in the local coordinate system diverge from a convex surface (Figure 3.4), so they do not correspond to the image observed by a viewer in any position. An observable image, on the other hand, represents a two-dimensional cross section[4] through the light field $B(\gamma, \beta; \theta'_r, \phi'_r)$ where θ'_r and ϕ'_r vary along with γ and β. Such an image cannot, in general, be represented as a convolution over the illumination.

In a few specific cases, however, the observed image does represent a convolution of the illumination. One such case is that of a surface with a Lambertian BRDF. A Lambertian surface reflects equal radiance in all directions, so the observed image is independent of the relationship between (θ'_r, ϕ'_r) and (γ, β). Thus any image parameterized by (γ, β) is a convolution of the illumination $L(\theta_i, \phi_i)$ with a clamped cosine kernel.[5]

For a non-Lambertian surface, the radiance of an image point depends not only on the illumination, but also on the position of the viewer. In general, the global coordinates of the direction toward the viewer vary as one moves across the surface. The appearance of a surface patch with a given BRDF under fixed illumination therefore depends on its position as well as its surface normal. In the case of an infinitely distant observer, however, the reflected rays are parallel, so the appearance of such a surface patch depends only on its normal. This simplifies the analysis, because we can parameterize the observed image as a function of γ and β irrespective of the surface geometry. We will make use of this property when dealing with different surface geometries in Chapter 6. While the property holds strictly only for an infinitely distant observer, it holds approximately whenever an observer is distant relative to the radius of curvature of the surface. It breaks down for flat surfaces, whose appearance would be uniform to an infinitely distant observer.

For a distant viewer, the image produced by the specular component of the Ward or Phong reflectance models can also be described approximately as a convolution over illumination. This convolution involves a different rotation operator from that of Equation (3.6). Appendix A presents a derivation of this result.

[4]This "cross section" is not typically planar. It corresponds to the values of the light field along a two-dimensional manifold, parameterized by γ and β. For example, an infinitely distant viewer sees an image comprised of rays whose direction (θ_r, ϕ_r) is constant in the global coordinate system. Therefore θ'_r and ϕ'_r depend on γ and β as $(\theta'_r, \phi'_r) = R^{-1}_{\gamma,\beta}(\theta_r, \phi_r)$. The image point corresponding to the surface point with normal (γ, β) therefore has radiance given by

$$B\left(\gamma, \beta; R^{-1}_{\gamma,\beta}(\theta_r, \phi_r)\right) = \int_{S^2} L(\theta_i, \phi_i) \tilde{f}\left(R^{-1}_{\gamma,\beta}(\theta_i, \phi_i); R^{-1}_{\gamma,\beta}(\theta_r, \phi_r)\right) d\omega.$$

[5]For a Lambertian reflector, the BRDF f is a constant ρ, so the convolution kernel is

$$\tilde{f}(\theta_i, \phi_i; \theta_r, \phi_r) = \begin{cases} \rho \cos\theta_i & \text{if } \theta_i \leq \frac{\pi}{2} \\ 0 & \text{if } \theta_i > \frac{\pi}{2}. \end{cases}$$

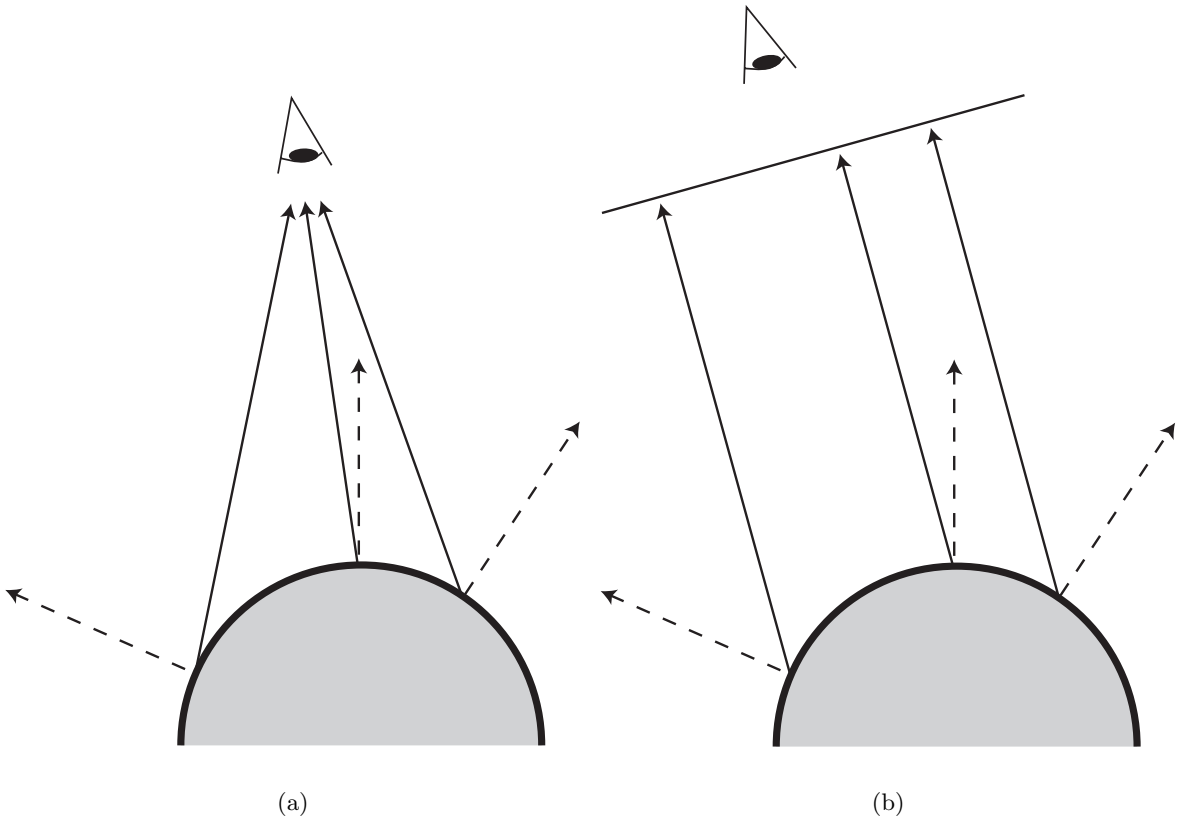


Figure 3.5. Imaging geometries for (a) a viewer at a finite distance from the surface and (b) an infinitely distant viewer. In each case, the dashed arrows represent surface normals, and the solid arrows represent light rays reflected in the direction of the viewer. In (b), the rays are parallel to one another and perpendicular to the distant image plane. The surface normals in this figure are identical to those of Figure 3.2(a) and Figure 3.4.

■ 3.2 Reflectance Estimation as an Inverse Problem

■ 3.2.1 Underconstrained Nature of the Problem

Reflectance estimation amounts to an inversion of the rendering process. Instead of generating images from illumination and reflectance, we wish to recover reflectance given image data. We wish to recover reflectance from a single image under unknown illumination, but we first consider an easier problem, that of recovering reflectance from the entire reflected light field under known illumination. That is, we assume we know the illumination $L(\theta_i, \phi_i)$ incident on every point from every direction, and that we have access to images of the surface from all possible directions.

If we place no constraints on the BRDF and allow it to vary arbitrarily from one point on the surface to another, the reflectance recovery problem is hopeless despite

all the information we have available. For each point on the surface, we have a two-dimensional array of measurements, corresponding to variations in θ_r' and ϕ_r', but we wish to recover a three- or four-dimensional array of values specifying an isotropic or anisotropic BRDF. One can make the problem more tractable by assuming that the BRDF is the same everywhere on the surface [65, 66]. Then, if the illumination consists simply of a point light source, each image point constitutes a separate BRDF measurement, and variations in the surface normal across the surface lead to BRDF samples over a wide range of incident illumination angles.

Even when the reflected light field is available for all surface normals and illumination at every point is known, however, the BRDF is only recoverable under certain illuminations. In the case of a convex surface under distant illumination, the reflected light field is a convolution of the illumination and the BRDF. If the illumination field contains no energy at certain frequencies (*i.e.*, at certain spherical harmonic orders), modes of the BRDF at those frequencies cannot be recovered [86]. The BRDF can be recovered completely under illumination by points sources, which contain energy at all frequencies, but not under illumination by a slowly-varying illumination field that contains no power at high frequencies.

Next, we consider the more difficult case when the entire reflected light field is available but the illumination is unknown. Again we assume that the BRDF is constant across the surface. Ramamoorthi and Hanrahan examine the combinations of BRDFs and illuminations that could produce a given reflected light field under the assumptions of Section 3.1.2 [86]. The symmetry of the BRDF spherical harmonic coefficients required by reciprocity ensures that the BRDF is unique up to a global scaling factor when all spherical harmonic coefficients of the reflected light field are nonzero. This factorization breaks down when certain coefficients of the reflected light field vanish. This causes problems in reflectance recovery, even when reflection is restricted to a simple parameterized reflectance model such as the Phong or Ward models. In particular, increasing the specular roughness parameter (α in the Ward model) has a virtually identical effect on the reflected light field as blurring the illumination $L(\theta_i, \phi_i)$ with an appropriate filter.

Finally, we consider the case where only one image of the surface is available, and the illumination is unknown. The light field factorization technique of Ramamoorthi and Hanrahan no longer applies. One cannot recover a three- or four-dimensional array of BRDF coefficients from a two-dimensional cross-section of the light field. Because the space of physically realizable BRDFs is much larger than the space of possible images, many different combinations of BRDFs and illuminations will produce identical images.

Even when we constrain the BRDF to a reflectance described by the Ward model, the reflectance estimation problem remains underconstrained. As in the case when the entire light field was available, reflectance recovery suffers from the ambiguity between increasing specular roughness and blurring the illumination, as well as unknown global scaling factors for illumination and reflectance. Appendix A describes an additional ambiguity that confounds diffuse reflectance under one illumination with specular re-

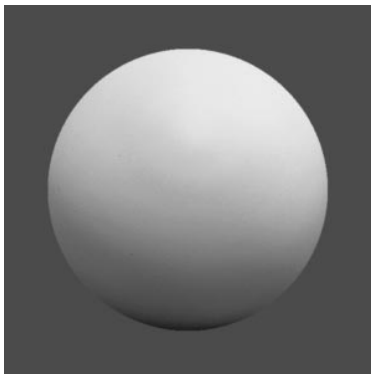


Figure 3.6. A photograph of a matte sphere, shown against a uniform gray background. This image could also be produced by a chrome sphere under appropriate illumination, but that scenario is highly unlikely.

flectance under another.

■ 3.2.2 Statistical Formulation

Because multiple reflectances could explain an observed image under unknown illumination, we wish to choose the *most likely* reflectance given the image. That is, we wish to regularize the underconstrained estimation problem using prior information about illumination and reflectance. For example, the image of Figure 3.6 could be explained either as a highly specular sphere under perfectly diffuse illumination or as a matte sphere under more typical illumination. The latter explanation is more likely, if only because real-world illumination typically contains some sharp edges.

Ideal Bayesian estimation of reflectance would require marginalizing over all possible illuminations to find the most likely BRDF $\hat{f}$ for a given observed image. If L denotes illumination from every direction at each point on the surface, and I denotes the observed radiance of each point in the image, then

$$\begin{aligned}
\hat{f} &= \arg\max_f P(f|I) \\
&= \arg\max_f \int_L P(f, L|I) dL \\
&= \arg\max_f \int_L P(f, L) P(I|f, L) dL \\
&= \arg\max_f P(f) \int_L P(L) P(I|f, L) dL. \qquad (3.7)
\end{aligned}$$

The prior probability over reflectances $P(f)$ captures the fact that some BRDFs are more common than others in the real world; for example, white matte surfaces are more common than holograms. Likewise, $P(L)$ captures the statistical structure of real-world illumination patterns. The last equality of Equation (3.7) is justified by the

independence of illumination (L) and surface reflectance (f).

As we showed in Section 3.1, the image I depends linearly on f. The estimation problem described by Equation (3.7) is nevertheless difficult for several reasons. First, it requires explicit formulations of $P(L)$ and $P(f)$, the prior probabilities of any given illumination and reflectance. Even under the assumptions of Section 3.1.2, where illumination and reflectance remain constant across a surface, illumination corresponds to a function of two variables and reflectance to a function of four variables. Although we demonstrate in Chapter 4 that real-world illumination displays a great deal of statistical regularity, an explicit probability distribution over all possible illuminations remains difficult to specify. Likewise, a general distribution over reflectances in the real world remains unknown, although some authors have studied the color distributions of Lambertian reflectors [12].

Evaluation of Equation (3.7) is also difficult because the expression to be maximized involves an integral over all possible illuminations. The unknown illumination constitutes a high-dimensional set of nuisance parameters that strongly affect our observed data but that we do not wish to estimate explicitly.

Replacing the integration with a simple maximum over illumination is not desirable, because the most likely combination of illumination and reflectance may not correspond to the most likely reflectance. Consider a photograph of a white matte sphere (*e.g.*, Figure 3.6), corrupted by slight high-frequency imaging noise. One could explain this image approximately as a white matte sphere under any of a number of illuminations, but none of these would predict the noisy image exactly. On the other hand, one could explain the photograph precisely as a chrome sphere under just the right illumination. This combination of illumination and reflectance may well be the single most likely explanation for the image. Integrating over all possible illuminations, however, would reveal that a more likely reflectance is white matte, because for that reflectance a large number of illuminations produce approximately the observed image.[6]

We wish to avoid integration over all illuminations and instead proceed by determining a set of informative statistics, computable from the observed image, that capture the information relevant to estimating reflectance.[7] Ideally, this set of statistics would be small and insensitive to variations in the illumination nuisance parameters.

We propose a solution along these lines in Chapters 5 and 6, in the form of a classifier that learns the relationship between reflectance and certain image statistics empirically. This circumvents the need for an explicit probability distribution over illuminations, and also avoids large-scale integration. We cope with the problem of explicitly specifying probabilities for all possible BRDFs by restricting our estimated reflectance to a set of common reflectances.

[6] This distinction between integration over illumination and maximization over illumination is related to the classic "generic viewpoint" assumption. Freeman [37] showed that integration over "generic" variables such as illumination strongly affects the outcome of certain reflectance estimation problems.

[7] An optimal solution might involve a set of *sufficient statistics* that capture all information relevant to reflectance estimation, but deriving a concise set of sufficient statistics requires a precise probability distribution over illuminations and proves impractical.

■ 3.3 Restricting the Reflectance Space

One can naturally restrict the space of reflectances among which an estimator must choose in one of two ways, leading to either a parameter estimation problem or a classification problem. In either case, we would like the estimator to choose the "best fit" to the observed surface reflectance among its set of candidate reflectances.

■ 3.3.1 Parameter Estimation

Previous authors have typically restricted the reflectance space by using a parameterized reflectance model from computer graphics, such as the Ward, Phong, or Torrance-Sparrow models [9, 65, 86, 92, 123]. We use this approach in analyzing the relationships between reflectance, illumination statistics, and image statistics (Chapter 5). We also used a parameterized reflectance space in the psychophysical experiments described in Section 2.3. In a practical estimation system, this approach suffers from the problem that no parameterized reflectance model provides a good fit to the entire space of real-world reflectances (Section 2.1.2).

■ 3.3.2 Classification

An alternative, non-parametric approach is to restrict reflectance to a finite set of BRDFs. In this case, the estimator acts as a classifier, selecting the best fit among the candidate BRDFs. We emphasize classification in the reflectance recognition system of Chapters 5 and 6. An advantage of this approach is that the candidate reflectances can be completely arbitrary. In fact, one does not even need to know their BRDFs. Using machine learning techniques, one can train such a classifier on sets of photographs of different materials (Figure 3.7). Chapter 6 provides several such examples.

■ 3.4 Normalization for Overall Illumination Intensity

The analysis of this thesis relies solely on a monochrome image of the surface of interest. Because we ignore image context, we are unable to resolve the ambiguity between overall strength of illumination and overall albedo of the surface. A white matte surface under dim illumination and a gray matte surface under bright illumination will produce identical images. Statistical characterization of real-world illumination distributions does not suffice to resolve this ambiguity, because similar illumination patterns occur at a wide range of overall intensities. One can, however, estimate overall intensity of illumination from the ensemble of surfaces visible in an image. The mean luminance of an image containing a number of surfaces provides an elementary estimate of this intensity. One can improve the estimate by taking the typical structure of real-world scenes into account [1, 10]. Because we wish to focus on the relationship between an image of a surface and its reflectance properties rather than overall scene structure, we eliminate the overall illumination ambiguity from the current study by assuming that an accurate estimate of overall illumination strength is available. In particular, we normalize our

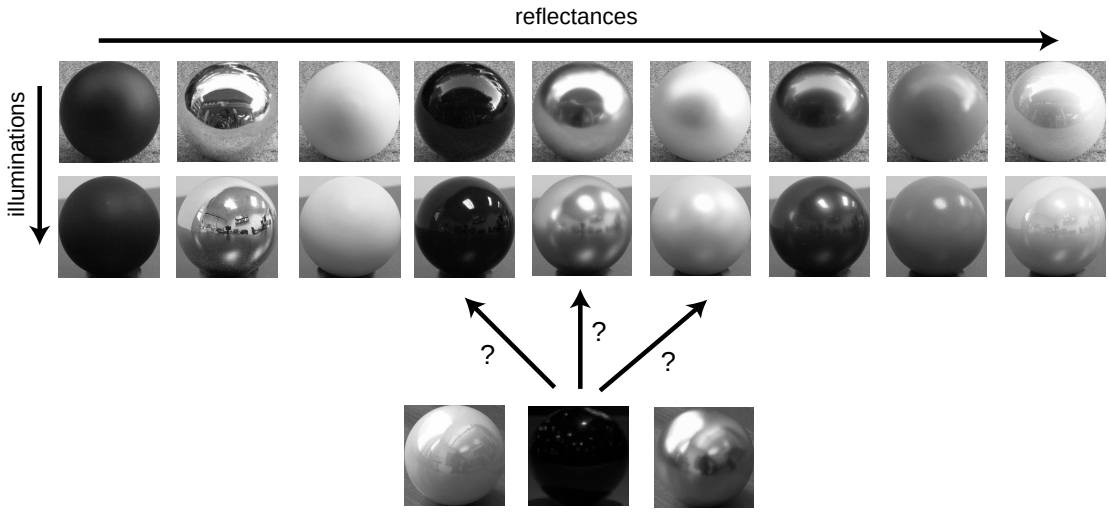


Figure 3.7. The problem addressed by a classifier of Chapter 6, illustrated using a database of photographs. Each of nine spheres was photographed under seven different illuminations. We trained a nine-way classifier using the images corresponding to several illuminations, and then used it to classify individual images under novel illuminations.

images by the luminance of a standard white surface positioned perpendicular to the viewer near the surface under observation.

■ 3.5 Summary and Discussion

The image of a surface depends on the surface's reflectance properties, its geometry, and the amount of light incident on it from every direction. One can express the light field reflected by a surface with homogeneous reflectance properties as a convolution of the illumination and the surface reflectance function. An observed image represents a cross-section through the reflected light field. Inverting this rendering process to recover reflectance is difficult. Even when illumination is known and images of the surface are available from all directions, one cannot always recover the BRDF. When illumination is unknown and only one image of the surface is available, reflectance recovery is even more underconstrained; multiple combinations of illumination and reflectance could explain the observed image. We wish to resolve the ambiguity by finding the most likely reflectance given the observed image. An explicit Bayesian solution of this problem is daunting, but this thesis develops an approximate, computationally tractable solution. The following chapter takes a step in that direction by describing the statistical properties of real-world illumination.

Chapter 4

Real-World Illumination Statistics

Computer vision systems have traditionally relied on idealized models of illumination, such as a single point light source or a uniform hemispherical source. Real-world illumination, on the other hand, is highly complex. Surfaces are illuminated not only by luminous sources such as the sun, sky, or indoor lights, but also by light reflected from other surfaces in the environment. Figure 2.9 illustrates the difference in appearance between a realistically illuminated surface and a surface under point source illumination.

The variability of real-world illumination complicates not only reflectance estimation but also a variety of other common visual recognition tasks. For example, shape-from-shading algorithms depend on the relationship between surface orientation and reflected surface radiance, which in turn depends on both illumination and reflectance. Motion estimation, object recognition, and scene recognition algorithms also depend on assumptions of illumination. Techniques that assume that all light radiates from a single point source may suffice for images acquired in a laboratory, but they often fail in the real world.

Despite its complexity, real-world illumination does possess some level of regularity. When looking at a chrome sphere, one expects to see some sharp edges, some bright light sources, some areas of slowly varying illumination, and so on. We recognize the sphere as chrome because we know what the world it reflects typically "looks" like. In other words, we have an intuitive feel for the spatial patterns of real-world illumination. We would like to convert these intuitions into a statistical description of real-world illumination that can be used to design better computer vision and graphics systems or to help understand the workings of biological visual systems.

■ 4.1 Measuring Illumination as an Image

One can measure the illumination incident from every direction at a particular point in the real world using a camera whose optical center is located at the point of interest. By combining photographs taken in different directions, one can compose a spherical map describing illumination at that point. Such spherical images are used as environment maps in computer graphics [23]. If all sources of direct and indirect illumination are relatively distant, the illumination map changes slowly as the hypothetical camera moves through space.

An illumination map is a type of image. However, accurate real-world illumination maps differ from typical photographs in two regards. First, illumination maps cover a much wider view angle, spanning the entire sphere instead of a narrow view angle near the horizontal. Second, accurate illumination maps must possess a much higher dynamic range than typical photographs to capture accurately the luminance of both the brightest and darkest areas. This is particularly true for illumination maps that contain localized primary light sources such as incandescent lights or the sun.

A number of researchers have devoted a great deal of effort to capturing statistics of typical photographs, or "natural image" statistics [13, 31, 47, 89, 102, 109]. They have found that normal photographs of indoor and outdoor scenes display a great deal of regularity, particularly in power spectra and distributions of wavelet coefficients. These statistics have led to effective image denoising and compression schemes [13, 82, 99] and helped explain the architecture of biological visual systems [31, 57, 96, 101].

We wish to determine whether illumination maps display statistical regularities of the same form. This chapter examines the statistics of illumination maps using distributions of illumination intensities (Section 4.4), spherical harmonic power spectra (Section 4.5), and distributions of wavelet coefficients (Section 4.6). Each section highlights both similarities and differences between traditional natural image statistics and the statistics of illumination maps.

■ 4.2 Data Sets

We worked with two different sets of illumination maps, each consisting of high dynamic range images that represent the radiance incident at a point in the real world. The first set consisted of 95 illumination maps based on imagery acquired by Teller *et al.* [108] in the environs of the MIT campus (*http://city.lcs.mit.edu/data*). The second set consisted of nine maps from Debevec's Light Probe Image Gallery (*http://www.debevec.org/Probes/*) [24]. Debevec's maps represent diverse lighting conditions from four indoor settings and five outdoor settings. Two examples from each data set are shown in Figure 4.1.

The images in both sets were acquired by combining photographs at multiple exposures to obtain pixel values that are linear in luminance, using the technique of Debevec and Malik [25]. We converted them all to gray-scale images with pixel values proportional to luminance. Debevec's illumination maps, which were computed from photographs of a chrome ball, cover the entire sphere. Teller's illumination maps were each mosaiced from multiple calibrated narrow-angle images. These mosaics cover the entire upper hemisphere as well as a band below the equator.

We compare our results to those of previously published studies of the statistics of traditional restricted-angle photographs. Huang and Mumford performed a number of statistical analyses on a particularly large set of images, consisting of 4000 photographs collected and calibrated by van Hateren and van der Schaaf [114]. These images were collected outdoors, but include photographs of buildings and roads as well as more "natural" scenes. Other image sets such as that of Tolhurst [109] include indoor images.

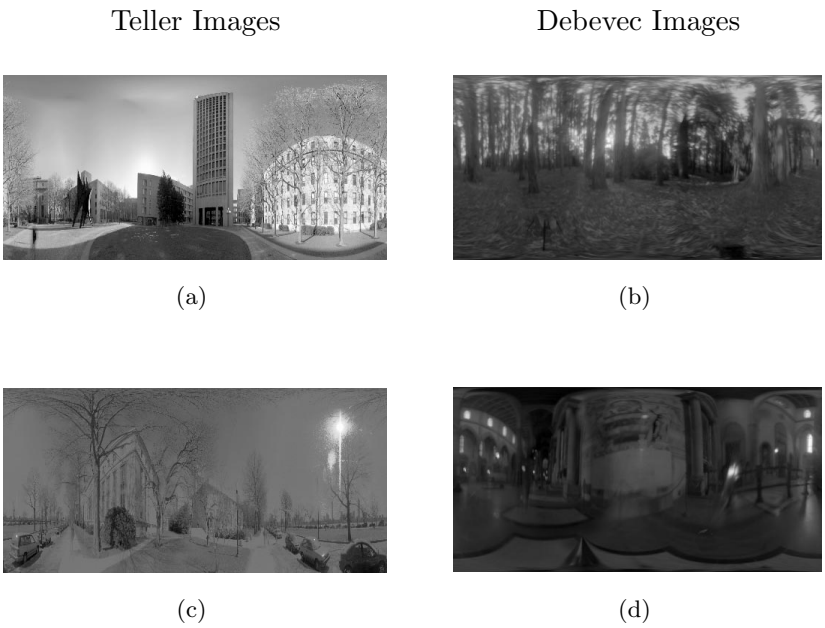


Figure 4.1. Examples of the illumination maps we used, shown in equal-area cylindrical projection (see Section 4.3). (a) and (c) are drawn from Teller's data set, while (b) and (d) are drawn from Debevec's. Dynamic range has been compressed for display purposes.

■ 4.3 Spherical Projection

Whereas image statistics have previously been analyzed on a planar domain, illumination maps are naturally defined on a sphere. We will describe our handling of this issue in each of the following sections. We found that storing illumination maps in equal-area cylindrical projection [16] facilitated certain computations. To construct this projection, one places the sphere at the center of a vertically oriented cylinder and projects each point on the spherical surface horizontally outward to the surface of the cylinder (Figure 4.2). One then unwraps the cylinder to obtain a rectangular map of finite extent. Regions of equal area on the sphere map to regions of equal area on the cylinder.[1] Figure 4.1 displays illumination maps in equal-area projection with $k = \frac{2}{\pi}$, where k is the ratio of the radius of the cylinder to the radius of the sphere.

[1] In particular, an infinitesimal patch on the sphere at latitude θ will find itself expanded by a factor of $k\frac{1}{\cos\theta}$ in the horizontal direction and reduced by a factor of $\cos\theta$ in the vertical direction. Because the product of these two factors is a constant k, this projection preserves areas, even though it heavily distorts angles near the poles.

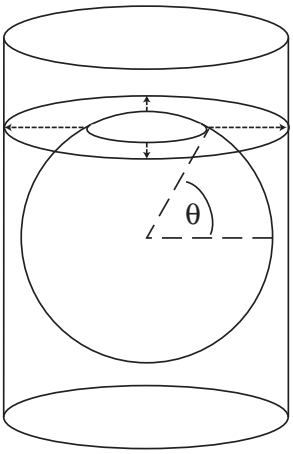


Figure 4.2. To produce the equal-area cylindrical projection of a spherical map, one projects each point on the surface of the sphere horizontally outward onto the cylinder, then unwraps the cylinder to obtain a rectangular map.

■ 4.4 Illumination Intensity Distribution

■ 4.4.1 Marginal Distribution of Intensity

Although light is typically incident on a real-world surface from every direction, the strongest illumination usually comes from primary light sources in a few directions. To quantify this intuition, we examined the marginal distribution of illumination intensity for our sets of illumination maps. This distribution is effectively just a histogram of pixel values. To compute it accurately, we must take into account the solid angle corresponding to each pixel of the illumination map. For an equal-area projection, this solid angle is constant, so we can compute the marginal distribution of illumination intensities with an unweighted pixel histogram.

Figure 4.3 shows total illumination intensity distributions for the 95 Teller images and for the 9 Debevec images. Panels (a) and (b) show the distribution of linear luminance values, while panels (c) and (d) show the distribution of log luminance values. The linear luminance distribution plots reveal the general trend we expect — a majority of pixels at low intensity, with a heavy positive tail corresponding to pixels of much higher intensities. A typical digital photograph stored in eight-bit format necessarily lacks this heavy positive tail due to limited dynamic range.

The log luminance histograms of Figure 4.3(c) and (d) show that a majority of pixels fall near the mean log luminance, with a smaller proportion of particularly dark or bright pixels. Huang and Mumford [47] attributed the asymmetry in the distribution of log luminance values for the 12-bit images they analyzed to the presence of sky in many of their images. Our distributions exhibit more striking asymmetries, partly because both the Teller and Debevec data sets contain not only sky but also more localized light sources. The distribution for the Teller set is particularly asymmetric

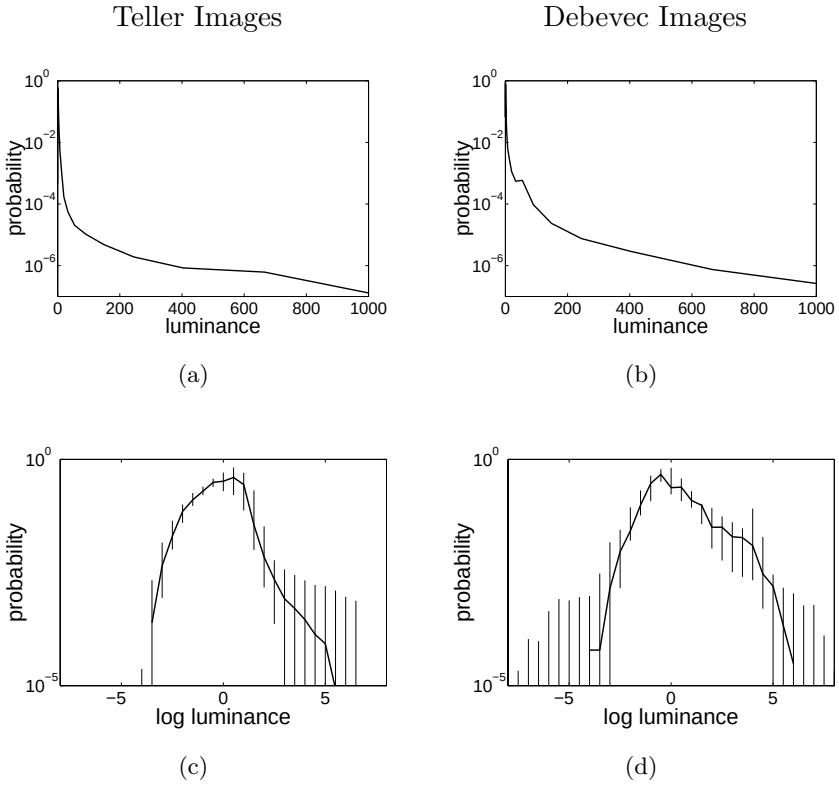


Figure 4.3. Illumination intensity distributions. (a) and (b) show mean histograms of linear luminance values for the 95 Teller images and 9 Debevec images, respectively. (c) and (d) show median histograms of natural log luminance values for the two image sets. The vertical bars extend from the 20th percentile to the 80th percentile of the distribution values over the image set. All image intensities were scaled linearly before analysis such that their mean log value was zero (*i.e.*, such that their geometric mean was one).

due to the presence of the sun in many images and to underexposure in the imaging system at very low light intensities.

The distribution of log luminance values for the Teller image set has standard deviation $\sigma = 1.04$, kurtosis $\kappa = 4.04$, and differential entropy $\mathcal{H} = 2.06$.[2] The Debevec image set has $\sigma = 1.32$, $\kappa = 12.49$, and $\mathcal{H} = 2.21$. Huang and Mumford found $\sigma = 0.79$, $\kappa = 4.56$, and $\mathcal{H} = 1.66$. The kurtosis values are influenced heavily by individual outliers. The variance and entropies of the distributions are higher for our data sets than for those of of traditional photographs, due to the higher dynamic range and the presence of concentrated illumination sources.

[2]The kurtosis of a random variable X with probability density $f(x)$ is defined as $\kappa = \frac{\int (x-\bar{x})^4 f(x)dx}{\left(\int (x-\bar{x})^2 f(x)dx\right)^2}$. The kurtosis of a Gaussian is 3, and distributions with kurtosis higher than 3 are often referred to as heavy-tailed. The differential entropy $\mathcal{H}$ of X is defined as $\mathcal{H}(X) = -\int f(x) \log f(x)dx$.

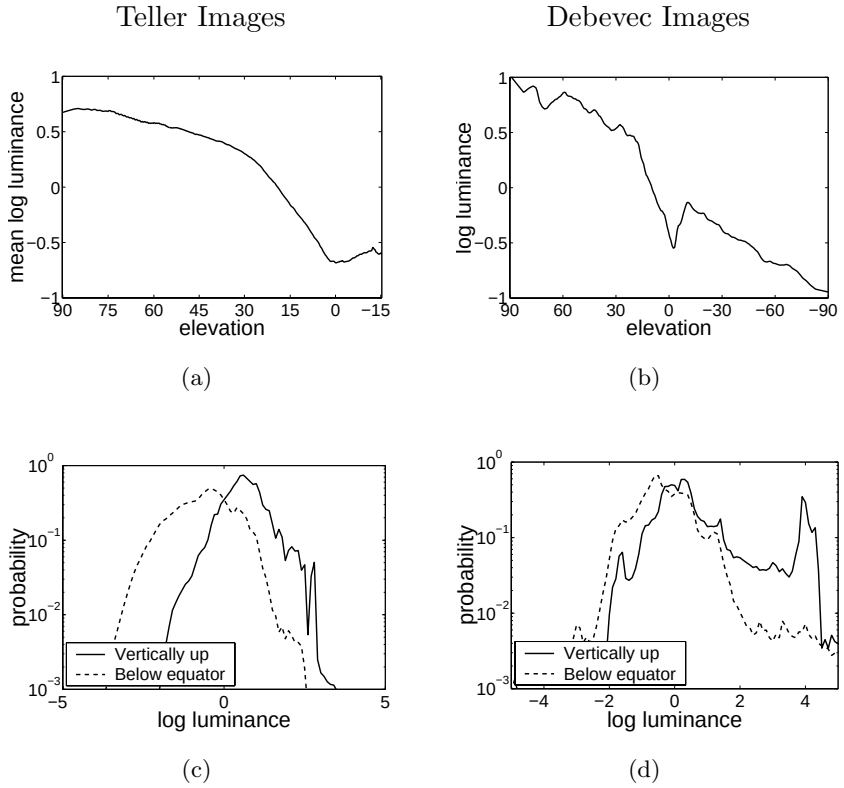


Figure 4.4. Dependence of illumination on elevation. (a) and (b) show mean log luminance as a function of elevation. (c) and (d) each show two histograms of illumination intensities, one for directions within 30° of the upward vertical and the other for directions from 0° to 15° below the equator.

Despite the aforementioned overall trends, marginal intensity distributions vary a great deal from one illumination to the next. The degree of variation in the distribution between images is summarized by the vertical lines in Figure 4.3(c) and (d), which extend from the 20th percentile to the 80th percentile of the distribution values over all the images.

■ 4.4.2 Non-Stationarity

Most researchers in image processing treat images as samples of a stationary statistical process. That is, they assume that all parts of the image possess identical statistical properties and they therefore treat each part of the image in the same way. Illumination maps clearly violate this stationarity assumption, if only because primary light sources such as the sun, sky, and indoor lights are more likely to appear in the upper hemisphere.

Figure 4.4(a) and (b) show mean luminance as a function of elevation for the two data sets. As expected, illumination generally increases with elevation. Interestingly, the mean intensity reaches a local minimum at the horizontal view direction. Both

Figure 4.5. This image, due to Torralba, represents the pixelwise mean of over 300 images of outdoor scenes containing a person whose head spans approximately two pixels. The images are aligned with respect to the person's head before averaging, so that a human-like shape is visible in the center. The remainder of the average image is of non-uniform intensity, with increased intensity near the top of the image and a noticeable dip in intensity near the horizon. Reprinted from [113] with author's permission.

data sets contain illumination maps in which the ground reflects a significant amount of light from above, while visible surfaces in the horizontal direction are shadowed (*e.g.*, Figure 4.1b). Torralba [112,113] observed that images of large-scale scenes viewed from a horizontal direction also have non-stationary means. He aligned large sets of images with respect to a feature of interest, such as a person, and averaged the images within each set pixelwise to obtain "average images" such as that shown in Figure 4.5. In most outdoor urban and natural settings, the average images exhibit a dip in intensity near the horizon [112], similar to the dip we observed for illumination maps in Figure 4.4(a) and (b).

Panels (c) and (d) of Figure 4.4 each show two illumination intensity histograms at different ranges of elevations. The marginal distributions for higher view directions have a larger mean as well as heavier positive tails, reflecting the larger probability of bright localized sources at higher elevations.

■ 4.4.3 Joint Distribution of Illumination from Adjacent Directions

To describe the spatial structure of real-world illumination maps, we must use statistics that depend on joint distributions of multiple pixels. The simplest way to do this is to examine the joint distributions of pairs of pixels with some specific spatial relationship. Figure 4.6 shows contour plots of the joint histograms of horizontally adjacent pixels

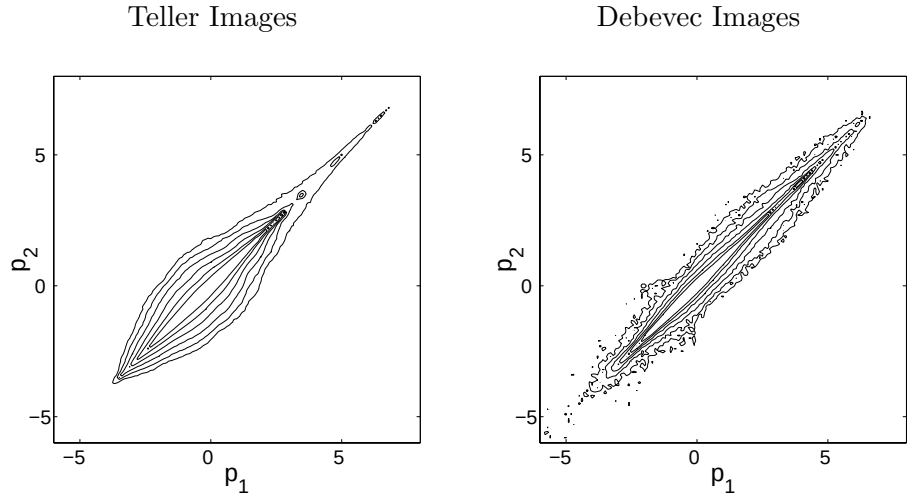


Figure 4.6. Joint histograms of log luminance at horizontally adjacent pixels p_1 and p_2 in the Teller images (left) and Debevec images (right).

from all of the Teller illumination maps and from all of the Debevec maps.[3]

Figure 4.6 shows that log luminance values at horizontally adjacent pixels p_1 and p_2 are highly correlated. Much of the mass of the joint histogram concentrates near the diagonal where $p_1 = p_2$. In agreement with Huang and Mumford, we found that $p_1 + p_2$ and $p_1 - p_2$ are more nearly independent than p_1 and p_2. In particular, the mutual information of p_1 and p_2 is 2.41 bits for the Teller images and 3.25 bits for the Debevec images, while that of $p_1 + p_2$ and $p_1 - p_2$ is only 0.10 bits for the Teller images and 0.07 bits for the Debevec images.[4] Hence, the percentage difference between the luminance incident from two horizontally adjacent spatial directions is roughly independent of the mean luminance from those two directions.

The variability of marginal pixel histograms from image to image leads to variability in the joint pixel histogram from image to image. The ensemble pixel histograms of Figure 4.6 also vary between the two data sets. In both panels of Figure 4.6, the increased extent of the joint distributions in the upper right quadrant compared to the lower left reflects the asymmetry of the marginal distribution illustrated in Figure 4.3.

The utility of joint pixel histograms for examining spatial illumination structure is limited by the difficulty of visualizing joint histograms of three or more pixels. In addition, the histograms vary from one illumination map to another. We wish to

[3] We define the horizontal direction in the global coordinate frame such that "horizontally adjacent" pixels lie along the same line of latitude. We divide each line of latitude into 512 "adjacent" pixels. Requiring that each pixel pair be separated by a fixed distance on the sphere results in virtually identical histograms.

[4] The mutual information of random variables X and Y is defined as $\mathcal{I}(X,Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X,Y)$, where $\mathcal{H}(X)$ and $\mathcal{H}(Y)$ are the entropies of X and Y, respectively, and $\mathcal{H}(X,Y)$ is the entropy of their joint density.

identify specifically the statistical regularities in illumination. We therefore turn to two image processing techniques that have formed the basis for statistical characterization of spatial properties of natural images — frequency domain analysis and wavelet analysis.

■ 4.5 Spherical Harmonic Power Spectra

Much early work on natural image statistics focused on the regularity of their power spectra. A number of authors [31, 89, 109] have observed that two-dimensional power spectra of natural images typically fall off as $1/f^{2+\eta}$, where f represents the modulus of the frequency and η is a small constant that varies from scene to scene.

The natural spherical equivalent of the planar Fourier transform is a spherical harmonic decomposition. The spherical harmonics form a countable orthonormal basis for square integrable functions on the sphere. Associated with each basis function is an order L, a nonnegative integer analogous to frequency. The $2L + 1$ spherical harmonics of order L span a space that is closed under rotation [49].

Just as planar white noise has a flat two-dimensional power spectrum, white noise on the sphere produces equal power in every spherical harmonic. Similarly, if the regularities observed in the natural image statistics literature carry over to spherical illumination maps, the average power of the spherical harmonics at order L will fall off as $1/L^{2+\eta}$.

We computed spherical harmonic coefficients for the illumination maps in both data sets using the formulas given by Inui [49]. We obtained average power at each order L as the mean of squares of the coefficients at that order. Teller's data lacks information about the lowest portion of the illumination hemisphere. We applied a smooth spatial window to these illumination maps before transforming them to the spherical harmonic domain.

Figure 4.7 shows the relationship between average power and harmonic order for the four illumination maps of Figure 4.1, when pixel value is proportional to log luminance. All four images have power spectra that lie close to a straight line of slope -2 on log-log axes, corresponding to a power spectrum of the form k/L^2. The great majority of images in both data sets exhibit similar behavior.

We obtain very different results for the same illuminations when we compute power spectra for illumination maps whose pixel values are linear in luminance. Illumination maps that lack concentrated primary light sources, such as those of Figure 4.1(a) and (b), have spherical harmonic spectra that are well approximated by $k/L^{2+\eta}$ with η small. On the other hand, illumination maps that contain intense, localized light sources have smooth power spectra that remain flat at low frequencies before falling off at higher frequencies. The illuminations of Figure 4.1(c) and (d) both display this behavior; the power spectrum of a linear luminance version of Figure 4.1(c) is shown in Figure 4.8. In these images, one or a few luminous sources, such as the sun or incandescent lights, dominate the power spectrum. Because these light sources approximate point sources, their spectrum is flat at low frequencies. If one clips the brightest pixel values in these

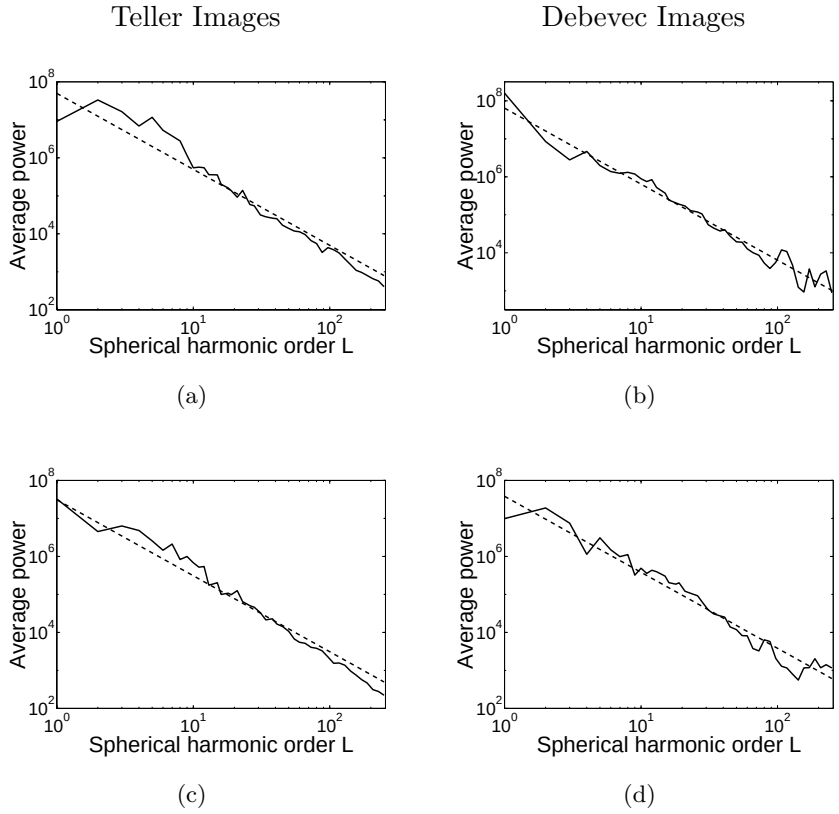


Figure 4.7. Spherical harmonic power spectra (solid lines) of illumination maps (a), (b), (c), and (d) in Figure 4.1, with pixel value proportional to log luminance. The dotted lines of slope -2 correspond to power spectra of the form k/L^2.

images, the power spectra return to the familiar $k/L^{2+\eta}$ form (Figure 4.8).

Figure 4.9 shows the mean spherical harmonic power spectrum of all the illuminations in the Teller data set, with vertical bars indicating the variability from one image to another. Panels (a) and (b) represent the spectra of linear luminance images, while (c) represents the spectra of log luminance images, and (d) represents the spectra of images where the brightest pixel values have been clipped. In panel (a), the images were normalized to have identical mean luminance values before computation of the power spectra. The power spectra exhibit a great deal of variability, but this results predominantly from differences in the total variance (power) of the different images. If the images are normalized for total variance instead, the variability of the power spectra decreases. The error bars are still quite large at low frequencies, however, because images dominated by one or a few point sources have flat power spectra at low frequencies. Clipping the brightest luminances or log transforming the image leads to more

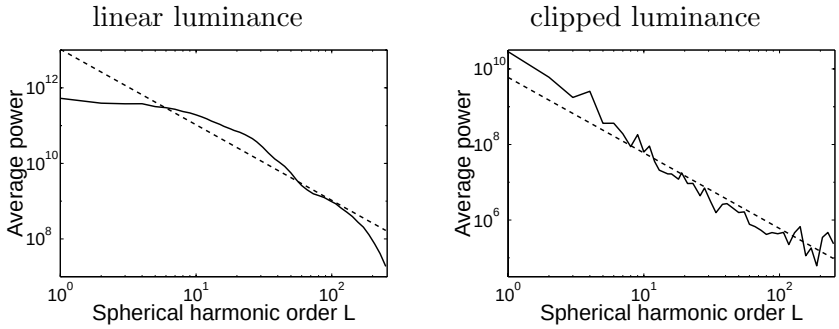


Figure 4.8. Left, the spherical harmonic power spectrum of the illumination map in Figure 4.1(c), with pixel values linear in luminance. Right, the corresponding spectrum after the pixel values corresponding to the sun have been clipped to a luminance value only slightly greater than that of the sky. Clipping these extremely bright pixels reduces power at all frequencies and produces a more linear power spectrum. The dotted lines of slope -2 correspond to power spectra of the form k/L^2.

regularly shaped power spectra, as indicated by the smaller error bars of (c) and (d).[5]

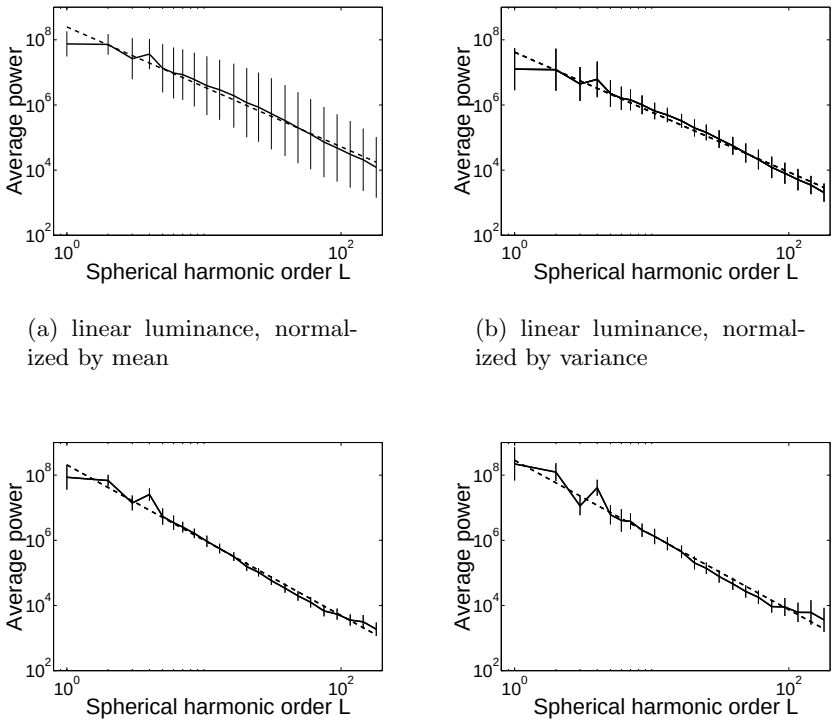
Previous work on natural images has reported $1/f^{2+\eta}$ power spectra whether pixel values are linear or logarithmic in luminance [89]. These results on linear luminance images differ from ours because most previous researchers have avoided photographs of point-like luminous sources and have used cameras of limited dynamic range, such that a few maximum intensity pixels could not dominate the image power spectra. A natural illumination map, on the other hand, may be dominated by light sources occupying a small spatial area. Once the relative strength of such sources is reduced through clipping or a logarithmic transformation, illumination maps have power spectra similar to those of typical photographs.

■ 4.6 Wavelet Statistics

The fact that a single bright source can dominate the power spectrum of an illumination map represents a shortcoming of frequency domain analysis. Wavelets[6] allow a more localized analysis; a single point-like source will affect only a few wavelet coefficients. Indeed, wavelet-domain analysis forms the basis for most recent work in the natural image statistics literature [89,101,117]. The distributions of wavelet coefficients at various scales and orientations capture not only power spectral properties, but also the non-Gaussian nature of real-world images. These distributions tend to be highly kurtotic, with many small coefficients and a few larger ones, indicating that wavelets provide a sparse representation of natural images. The scale-invariant properties of

[5]Log transforming an image tends to amplify camera noise and discretization noise at small pixel values, introducing variability in the power spectrum at high frequencies. This variability is notable in the rightmost error bars of Figure 4.9(d).

[6]We use the term "wavelet" to refer generally to image decompositions based on multiscale bandpass filter pyramids, including quadrature mirror filter pyramids and overcomplete steerable pyramids.



(a) linear luminance, normalized by mean

(b) linear luminance, normalized by variance

(c) clipped luminance, normalized by variance

(d) log luminance, normalized by variance

Figure 4.9. Mean power spectra of the 95 Teller images. Heavy solid lines indicate the mean of the individual power spectra at each spherical harmonic order, while each vertical bar extends both above and below this line by one standard deviation. The power spectra of (a) and (b) were computed on images whose pixel values were linear in luminance. In (a), images were scaled to have the same mean, while in (b), images were scaled to have the same pixelwise variance (*i.e.*, the same total non-DC power). In (c), power spectra were computed for "clipped" images, which were linear in luminance up to a ceiling value slightly brighter than the typical brightness of the sky. The power spectra of (d) were computed for log luminance images. The images of (c) and (d) were scaled to have the same variance. The dotted lines are best-fit lines corresponding to power spectra of the form $k/L^{2+\eta}$, where η is -0.16 in (a) and (b), 0.31 in (c) and 0.29 in (d). Each point on the heavy solid curve represents the average power of an interval containing one or more discrete frequencies. Note that the vertical lines are not traditional error bars, because they represent standard deviation rather than standard error of the mean. These standard deviations were computed on log power values.

natural images translate into predictable relationships between wavelet coefficient distributions at different scales. The regular nature of these distributions facilitates image denoising [82, 99], image compression [13], and texture characterization [43, 81], and has also proven useful in understanding neural representations in biological visual systems [96, 101].

Previous analysis of natural images and textures has assumed that the data is defined

on a planar domain. Because illumination maps are defined as functions of orientation, they are most naturally analyzed in a spherical domain. To this end, we utilized the spherical wavelet framework introduced by Schröder and Sweldens [95]. These wavelet transforms operate on data defined on a subdivided icosahedron whose vertices are quasi-regular on the surface of the sphere. We used a transform based on second-generation wavelets with vanishing zero- and first-order moments, constructed from simple hat functions using a linear lifting scheme.[7] We computed wavelet coefficients using a fast lifted wavelet transform implementation by Amaratunga and Castrillon-Candas [3].

Figure 4.10 shows marginal distributions of spherical wavelet coefficients at three successive scales for the 95 Teller images. The distributions are highly kurtotic, with the great majority of coefficients near zero and a few much larger coefficients. Figure 4.11 summarizes the variation from image to image for the distribution at one scale, for both linear luminance and log luminance images. The distributions are remarkably similar from one image to another, although the distributions associated with the linear luminance images exhibit variations in the overall scale of the wavelet coefficient distribution. The sun and other bright localized sources that dominate the entire power spectra of some of the illumination maps (Section 4.5) have a less noticeable effect on the distributions of wavelet coefficients because they influence only a handful of wavelet coefficients.[8] Several authors have observed that generalized Laplacian distributions of the form $\mathcal{P}(x) \propto \exp(-|x/s|^{\alpha})$ accurately model the wavelet coefficient distributions of typical photographs and of ensembles of photographs [13, 47]. Panels (a) and (b) of Figure 4.10 show maximum likelihood fits of this form to the ensemble histogram of wavelet coefficients from the Teller images. The fits are reasonably accurate, although they tend to underestimate the actual distribution for high wavelet coefficient magnitudes. We observed similar behavior for fits to empirical wavelet coefficient distributions for individual illumination maps. This discrepancy from results reported in the natural image statistics literature may be due to the higher dynamic range of the illumination maps we analyzed.

The wavelet coefficient distributions of Figure 4.10 also exhibit evidence of scale invariance in illumination maps. Distributions of coefficients at different scales are similar apart from an overall normalization constant. Scale invariance requires that all statistics computed on an ensemble of images $I(\mathbf{x})$ be identical to those computed on normalized, rescaled versions of the images $\beta^{\nu} I(\beta\mathbf{x})$, where the exponent ν is independent of the scale β [89]. An exponent $\nu = 0$ leads to two-dimensional power spectra of the form spectra $1/f^2$, where f is the modulus of frequency. More generally, a nonzero

[7] These wavelets are known as second-generation because they are not exact translates and dilates of a single function. Their first-order moments vanish only approximately, because the arrangement of vertices on the sphere is not completely regular.

[8] The variance of wavelet coefficients at a particular scale and orientation provides a measure of spectral power in some frequency band. A single localized light source can greatly influence this variance by contributing a few large outlying wavelet coefficients. However, it will have a relatively small effect on the shape of the histogram.

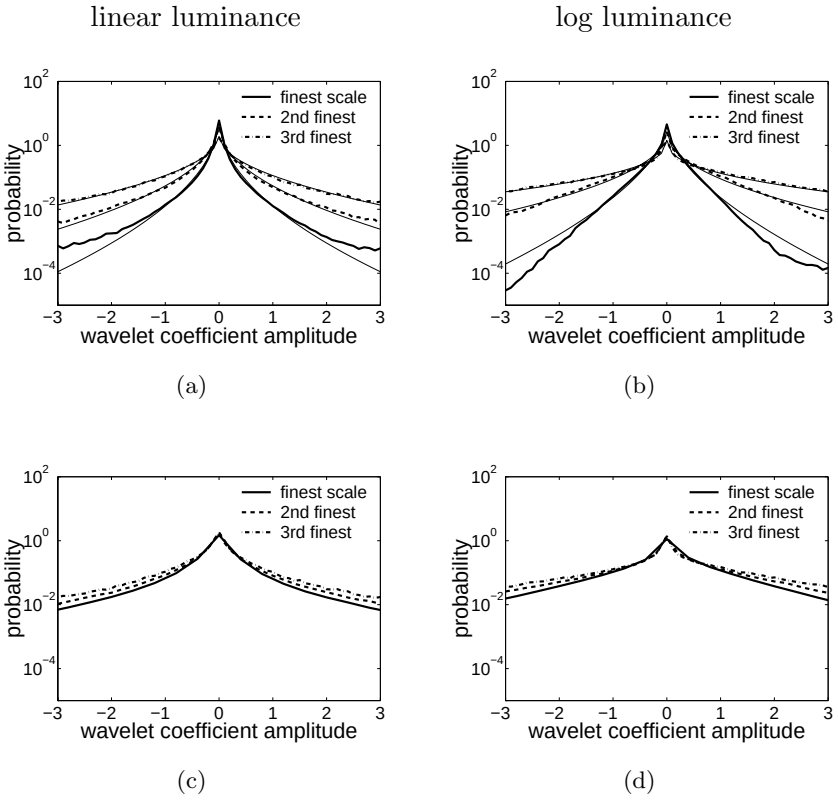


Figure 4.10. Distributions of spherical wavelet coefficients at successive scales (thick lines), along with generalized Laplacian fits (thin lines in (a) and (b)), for the 95 Teller images. In (a) and (b), as elsewhere in this chapter, the spherical wavelet basis functions are normalized to have identical power at every scale. In (c) and (d), their amplitudes are divided by 4 at the finest scale and by 2 at the next finest scale. (a) and (c) were computed on images whose pixel values were linear in luminance, while (b) and (d) were computed on log luminance images. The α parameters of the generalized Laplacian fits ranged from 0.50 to 0.52 for the linear luminance images, and from 0.41 to 0.59 for the log luminance images. We windowed the illumination maps as described in Section 4.5 before computing the wavelet transform, and discarded wavelet coefficients corresponding to the absent portions of the illumination map. We divided each linear luminance image by its mean before computing wavelet coefficients.

exponent ν leads to power spectra of the form $1/f^{2-\nu}$. For a scale-invariant image ensemble, the variance of wavelet coefficient distributions will follow a geometric sequence at successively coarser scales. If the wavelet basis is normalized such that wavelets at different scales have constant power, as measured by the L^2 norm, then the variance will increase by a factor of $2^{2+\nu}$ at successively coarser scales. If we increase the amplitude of these wavelets by a factor of 2 at each coarser scale, then the variance of the coefficients will increase by a factor of only 2^{ν} at successively coarser scales. Panels (c) and (d) of Figure 4.10 illustrate the results of such rescaling. Because ν is small, the distributions change little from one scale to the next. Note that linear luminance

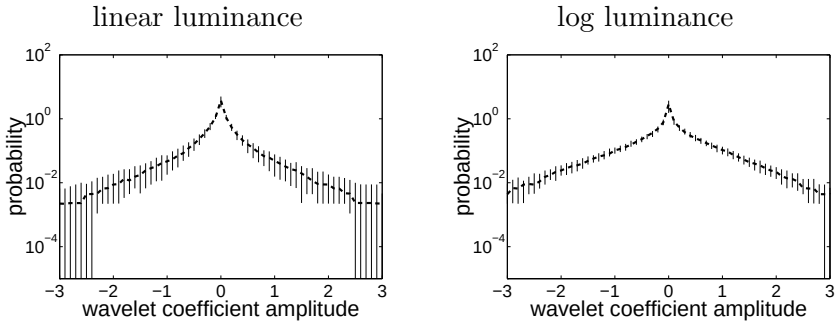


Figure 4.11. Variation in marginal distributions of wavelet coefficients from one image to another, for the second-finest-scale band of Figure 4.10. The heavy dashed lines indicate the median of the histogram values across the 95 images. The vertical bars extend from the 20th percentile to the 80th percentile of the distribution values across images. We divided each linear luminance image by its mean before computing wavelet coefficients but did not normalize either linear or log luminance images for variance.

illumination maps are not strictly scale invariant, as evidenced by the fact that their power spectra often deviate significantly from the $1/f^{2-\nu}$ form. The distributions of wavelet coefficients at successive scales suggest, however, that illumination maps do possess scale-invariant properties apart from the contributions of bright localized light sources.

The spherical wavelet basis used to generate Figures 4.10 and 4.11 consists of wavelet functions with approximate radial symmetry. Oriented wavelet pyramids have proven indispensable in characterizing natural image statistics and in other areas of image processing. A potentially productive line of future research involves construction of oriented pyramids, including steerable pyramids, for a spherical data domain. Because such transforms are not readily available, however, we applied planar wavelet analysis to equal-area cylindrical projections of the Teller and Debevec illumination maps. Although this projection introduces spatially-varying distortion that may affect the image statistics, it allows direct comparison of our results to the existing literature on natural image statistics and texture analysis. Horizontal lines in the projected images correspond to lines of latitude on the sphere, while vertical lines correspond to lines of longitude.

Figure 4.12 shows marginal distributions of wavelet coefficients at several successive scales for quadrature mirror filter pyramids constructed on cylindrical projections of the Teller illumination maps. The coefficient distributions for both vertically- and horizontally-oriented wavelets are similar to those observed for spherical wavelets in Figure 4.10.

Authors in the natural image statistics literature have noted that even though wavelet coefficients are approximately decorrelated, coefficients that are near one another in position, scale, or orientation exhibit codependencies that are remarkably reproducible for different images and different wavelet transforms [13,47,102]. Figure 4.13

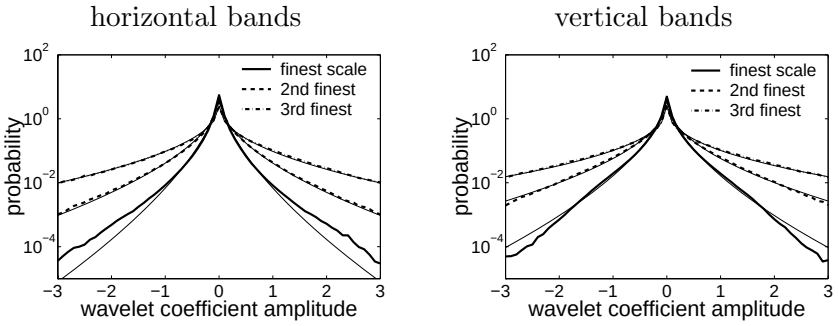


Figure 4.12. Distributions of planar quadrature mirror filter (QMF) wavelet coefficients at successive scales (thick lines), along with generalized Laplacian fits (thin lines), for the 95 Teller images. Left, horizontal bands. Right, vertical bands. These distributions were computed for log luminance images and are therefore best compared to the spherical wavelet coefficient distributions shown in Figure 4.10(c). The finest scale in this decomposition is slightly finer than that of the finest scale in Figure 4.10(c), such that the variances of the distributions in this figure are somewhat lower. The α parameters of the generalized Laplacian fits ranged from 0.48 to 0.63 for the horizontal bands and from .42 to .58 for the vertical bands. The QMF pyramid was computed on an equal-area projection of the spherical illumination map (as defined in Section 4.3), with $k = \frac{2}{\pi}$. We used an eight-tap quadrature mirror filter pyramid described by Johnston [53] and implemented by Simoncelli [98].

shows the conditional distributions of the horizontal wavelet coefficents of the Teller illumination maps given the the values of several nearby coefficients. These distributions are shown as images, with each column representing the distribution of the horizontal coefficient given a particular value of a related coefficient. Brighter pixels represent higher probabilities, with the probabilities in each column summing to one.

All four of the joint distributions exhibit a "bowtie" shape characteristic of natural images [13, 102]. The variance of a wavelet coefficient increases with the magnitude of neighboring coefficients at the same scale and orientation, and also with the magnitude of coefficients of other scales and orientations at the same spatial location. Intuitively, edges and bright sources tend to produce large wavelet coefficients at multiple scales and orientations and at nearby positions. Figure 4.13(d) shows that two horizontally adjacent, horizontally-oriented wavelet coefficients at the same scale also exhibit significant correlation. This correlation reflects the tendency of edges in an image or illumination to continue in the same direction; horizontally-oriented wavelets respond strongly to horizontal edges.

■ 4.7 Summary and Discussion

Despite their complexity and variability, real-world illuminations display a great deal of statistical structure. Many of the regularities observed through earlier studies of low-dynamic-range, restricted field-of-view photographs carry over to real-world illumination maps. The pixel histograms, power spectra, and wavelet coefficient distributions of spherical illumination maps share some of the characteristics previously noted

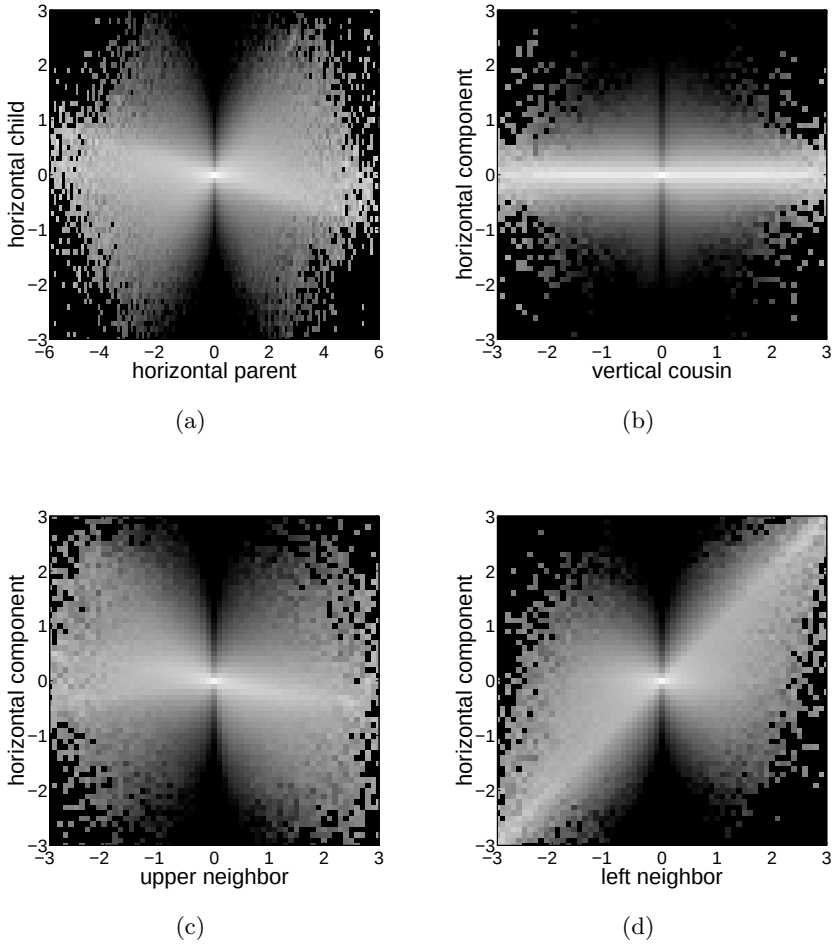
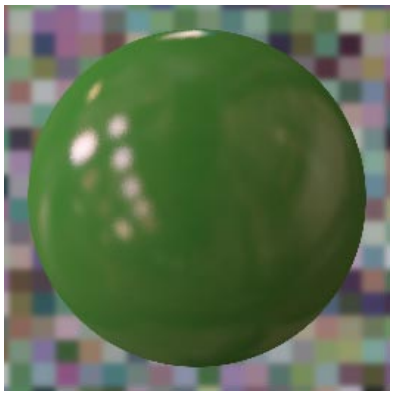
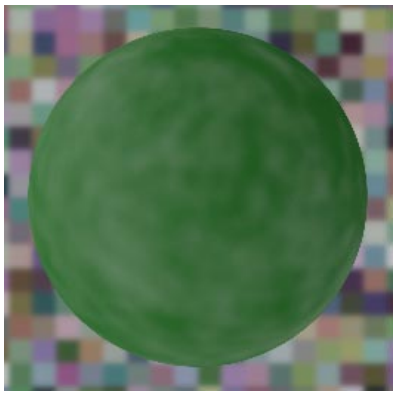


Figure 4.13. Conditional histograms for a horizontal wavelet coefficient given the values of its neighbors. The brightness of each pixel indicates a probability; the probabilities in each column sum to unity. The vertical axis is a fine-scale horizontal coefficient of an 8-tap quadrature mirror filter decomposition [53]. The horizontal axis represents (a) the horizontal coefficient at the same position but at the next coarser scale, (b) the vertical coefficient at the same scale and position, (c) a vertically adjacent horizontal coefficient at the same scale, and (d) a horizontally adjacent horizontal coefficient at the same scale. The conditional histograms represent average distributions over the 95 Teller log luminance images. Note that the horizontal axis of (a) is compressed by a factor of two relative to the other axes.

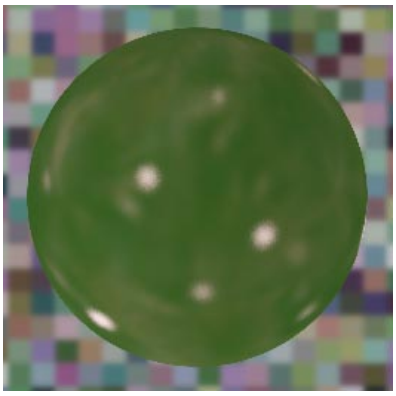
in the image statistics literature. Illumination maps, however, display non-stationary statistical properties, such as the predominance of illumination from upward directions. In addition, the fact that one or a few primary light sources may contain a majority of the power in an illumination map sometimes leads to power spectra that deviate significantly from the $k/f^{2+\eta}$ model, violating scale invariance. Wavelet coefficient distributions are fairly regular from image to image, but fits to generalized



(a) Original



(b) $1/f^2$ power spectrum



(c) Heeger and Bergen texture

(d) Portilla and Simoncelli texture

Figure 4.14. Spheres of identical reflectance properties rendered under a photographically-acquired illumination map (a) and three synthetic illumination maps (b-d). The illumination in (b) is Gaussian noise with a $1/f^2$ power spectrum. The illumination in (c) was synthesized with the procedure of Heeger and Bergen [43] to match the pixel histogram and marginal wavelet histograms of the illumination in (a). The illumination in (d) was synthesized using the technique of Portilla and Simoncelli, which also enforces conditions on the joint wavelet histograms. The illumination map of (a) is due to Debevec [24].

Laplacian distributions are less tight than those previously observed for more typical photographs [13, 47].

The domains in which we have characterized natural illumination statistics — distributions of intensities, power spectra, and distributions of wavelet coefficients — are also used to characterize texture [43, 82]. This suggests that we might think of illumination patterns as types of textures. We can test the extent to which a set of statistics captures the perceptually essential characteristics of real-world illumination by apply-

ing texture synthesis algorithms to generate novel illuminations whose statistics match those of real illuminations. Panel (a) of Figure 4.14 shows a sphere rendered under the photographically-acquired illumination map of Figure 4.1(d). Panels (b), (c), and (d) show identical spheres rendered under synthetic illumination maps. The illumination map of (b) consists of Gaussian noise with a $1/f^2$ power spectrum; although the power spectrum resembles that of natural illumination, the resulting sphere does not look realistic at all.[9] The illumination map of (c) was synthesized to have a pixel intensity distribution and marginal wavelet coefficient distributions identical to those of (a) using the texture synthesis technique of Heeger and Bergen [43]. This sphere looks much more realistic, and human observers are able to recognize that its reflectance properties are similar to those of the sphere in (a) [34]. Finally, the illumination map of (d) was created using the texture synthesis technique of Portilla and Simoncelli [82], which ensures that not only its pixel intensity distribution and marginal wavelet coefficient distributions, but also certain properties of its joint wavelet coefficient distributions, match those of (a). This synthetic illumination map captures the presence of edges in the real illumination map, leading to a sphere whose reflectance properties are even more similar to that of (a). This suggests that the statistical properties of natural illumination discussed in this chapter play an important role in reflectance estimation by the human visual system [34].

In Chapters 5 and 6, we consider the implications of illumination statistics in reflectance estimation. These statistics have other potential applications that suggest additional lines of future research. A statistical characterization of illumination could enable shape-from-shading under unknown illumination, or even joint estimation of shape and reflectance. In computer graphics applications, one often wishes to recover the illumination of an object or scene for re-rendering purposes. A statistical model of real-world illumination may help reconstruct a particular illumination from sparse or incomplete measurements.

One could extend our treatment of real-world illumination by considering the statistics of the five-dimensional plenoptic function that describes all the rays of light passing through every point in a three-dimensional volume [2]. The plenoptic function can be described as the set of two-dimensional spherical illumination maps at every point in a three-dimensional volume. Understanding its statistics requires understanding how an illumination map tends to change as the camera recording it moves in space. Because

[9] The illumination map of Figure 4.14b was synthesized in the spherical harmonic domain. The maps of (c) and (d) were synthesized in a rectangular domain corresponding to an equal-area cylindrical projection of the sphere. In (c) and (d), we performed principle component analysis in color space to produce three decorrelated color channels, each of which is a linear combination of the red, green, and blue channels. We then synthesized textures independently in each channel of this remapped color space, as suggested by Heeger and Bergen [43]. Unfortunately, the dependencies between the decorrelated color channels are much more severe for high-dynamic-range illumination maps than for the 8-bit images common in the texture analysis literature. To reduce artifacts associated with these dependencies, we passed the original illumination maps through a compressive nonlinearity on luminance before wavelet analysis, and then applied the inverse nonlinearity to the synthesized illumination maps. The compressive nonlinearity leads to a less heavy-tailed distribution of pixel intensities.

image-based rendering involves resampling the plenoptic function [67], statistical priors on this function could facilitate image-based rendering with sparse data.

Chapter 5

From Illumination Statistics to Image Statistics

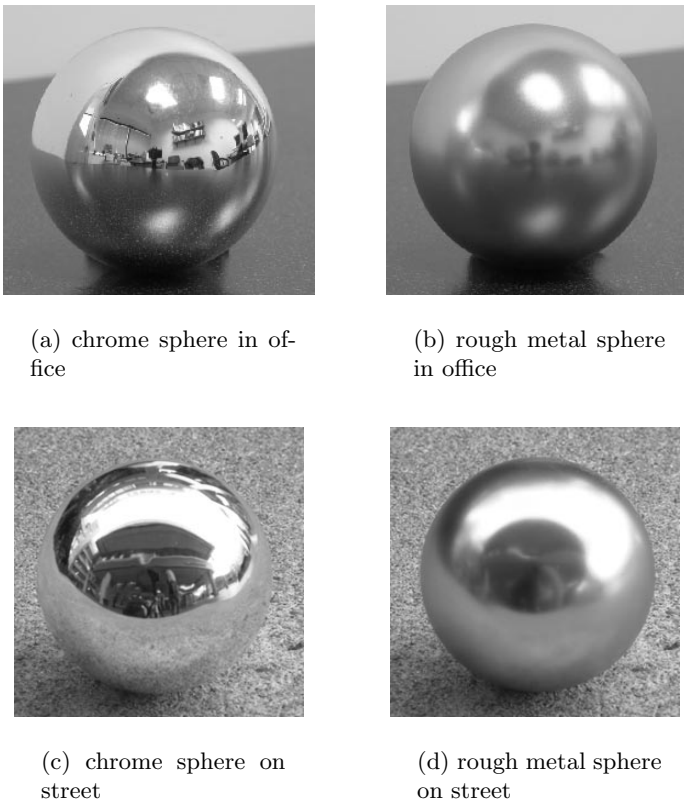
The statistical regularity of real-world illumination leads to statistical regularity in images of reflective surfaces. Certain image statistics vary little from one real-world illumination to another but vary substantially from one reflectance to another. We can exploit these regularities to estimate surface reflectance properties, disambiguating an otherwise underconstrained recognition problem.

As an example, Figure 5.1 shows images of two spheres, each photographed in two locations. The two illuminations differ substantially, but both contain sharp edges. Both images of the chrome sphere therefore contain sharp edges, while both images of the rough metal sphere contain blurred edges. The presence of such features depends not only on the physical laws governing reflection, but also on the empirical properties of real-world illumination.

One can quantify regularities in surface images with statistical measures similar to those used in Chapter 4 to describe regularities in illumination. In Section 5.1, we use a parameterized reflectance model to examine the dependence of individual image statistics on reflectance. In Section 5.2, we propose a reflectance classifier based on these relationships. By utilizing a machine learning approach, we are able to build classifiers for arbitrary reflectances, not restricted to reflectances described by the parameterized model. Chapter 6 considers these machine learning techniques in more detail, and extends our classification scheme to surfaces of arbitrary known geometry.

■ 5.1 Dependence of Image Statistics on Reflectance

To gain insight into the relationship between illumination statistics, surface reflectance, and surface image statistics, we consider surface reflectances defined by the isotropic Ward model. This model, defined by Equation (2.4), represents reflectance as a sum of specular and diffuse components. The monochromatic version of the Ward model contains only three parameters. The diffuse reflectance ρ_d specifies the energy in the diffuse, Lambertian component. The specular reflectance ρ_s specifies the energy in the specular component, while the roughness parameter α specifies the blur in the specular component. Conservation of energy implies that $\rho_d + \rho_s \leq 1$. For a highly



(a) chrome sphere in office

(b) rough metal sphere in office

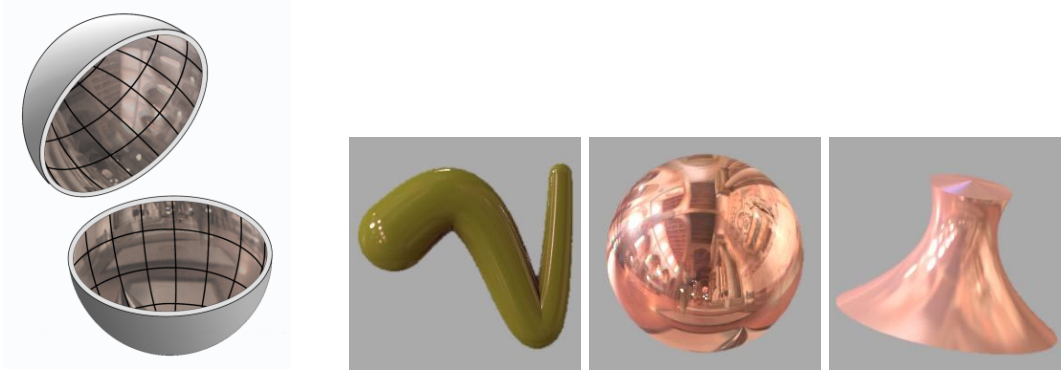
(c) chrome sphere on street

(d) rough metal sphere on street

Figure 5.1. A chrome sphere and a rough metal sphere, each photographed under two different real-world illuminations.

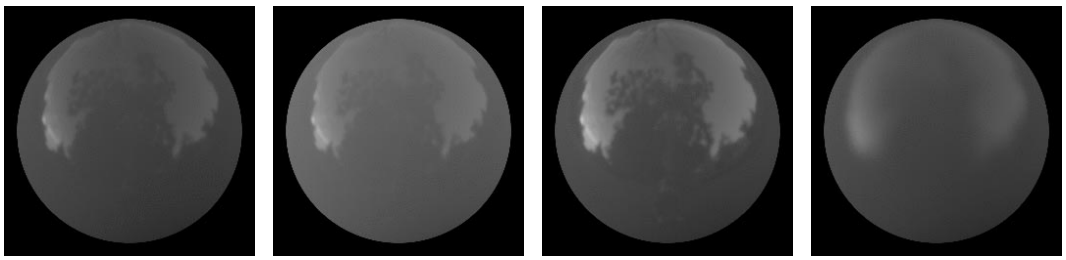
reflective surface, $\rho_d + \rho_s$ will be close to 1; for a dark surface that absorbs most incident illumination, $\rho_d + \rho_s$ will be closer to 0.

Given the geometry of a surface, its BRDF, and a description of illumination from every direction at every point on the surface, we can use Equation (3.1) to render the surface as it would appear to a viewer at any position. Assuming that sources of direct and indirect illumination are distant relative to the size of the surface, we can represent illumination at all points on the surface with a single spherical illumination map. Evaluating Equation (3.1) at a single point on the surface under arbitrary illumination requires integration over a hemisphere of the illumination map and therefore represents significant computational expense. Traditional computer graphics systems often assume lighting by a collection of point sources to simplify this computation, but such renderings fail to capture the appearance of surfaces under complex real-world illumination (Figure 2.9). We rendered surfaces of various geometries and reflectance under photographically-acquired illumination maps using Ward's *Radiance* rendering package [56]. *Radiance* employs stochastic sampling algorithms that accelerate the ren-



(a) (b) (c) (d)

Figure 5.2. (a) A photographically-acquired illumination map, illustrated on the inside of a spherical shell. The illumination map is identical to that of Figure 4.1d. (b-d) Three surfaces of different geometry and reflectance rendered under this illumination map using the methods of Appendix B.

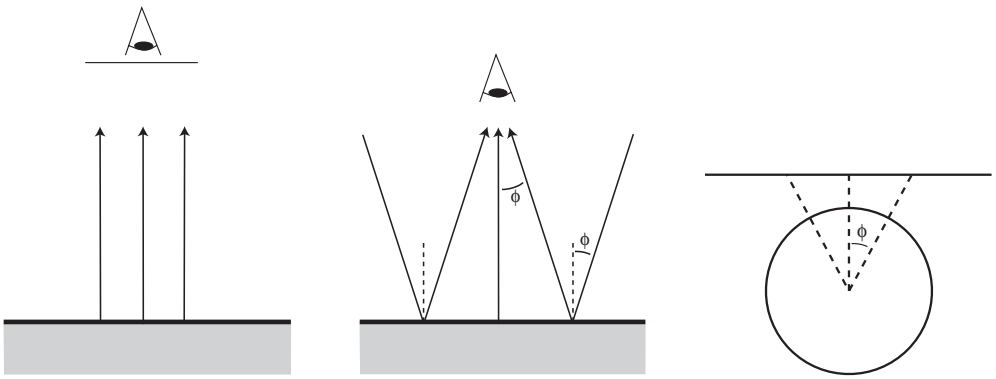


(a) Original reflectance (b) Increased ρ_d (c) Increased ρ_s (d) Increased α

Figure 5.3. Effect of changes in Ward model reflectance parameters on images of a sphere, rendered under a photographically-acquired illumination map due to Debevec [24].

dering process but also introduce image artifacts. We developed techniques to reduce these artifacts such that they are barely detectable visually, for images of individual surfaces under real-world illumination maps. Appendix B describes these rendering techniques. Figure 5.2 shows surfaces of several geometries and reflectance properties rendered under one photographically-acquired illumination map.

Figure 5.3 illustrates the effect of varying each Ward model parameter on a sphere rendered under fixed illumination. The diffuse component of these images, considered as a function of surface normal direction, represents a convolution of the illumination map with a clamped cosine kernel (Section 3.1.3). This component therefore varies slowly over the image. The specular component of the image can also be expressed as a



(a) Distant viewer (b) Nearby viewer (c) Projection of illumination

Figure 5.4. Specular reflection from a planar surface. (a) If the viewer is infinitely distant from the surface, then each point of the image represents a reflection of light from the same global direction. (b) If the viewer is close to the surface, then each point of the image represents a reflection of light from different directions. The diagram shows the direction of reflected light for a perfect chrome reflector. We will assume that the surface is perpendicular to the view direction. We define the distance from the viewer to the surface as 1, and measure distance on the surface accordingly. (c) An image of a chrome planar surface represents a gnomonic projection of the spherical illumination map onto the surface. If the surface is horizontal and the view direction is vertically downward, a surface point at angle ϕ from the viewer reflects light incident at an angle ϕ from the upward vertical.

convolution of the illumination map, using a different rotation operator (Appendix A). The smaller α is, the smaller the associated kernel, and the more quickly the specular component varies across the image.

To further simplify the analysis, we consider a particularly simple surface geometry, namely a flat tile. We assume distant illumination, such that the radiance incident on the surface from a particular direction does not vary from one point on the surface to another. To a distant viewer, the surface will appear to have constant brightness (Figure 5.4(a)). For a viewer close to the surface, on the other hand, the view angle changes from one point on the surface to another (Figure 5.4(b)). The radiance of a Lambertian reflector is independent of view direction, so the diffuse component of the observed image is still constant across the surface. The specular component, however, depends on view direction and therefore varies across the surface. In particular, one can approximate the specular reflectance over a narrow field of view as a blurred projection of the illumination map onto the plane (Figure 5.4(c)). For the view geometry shown in Figure 5.4(b), the blur is a Gaussian filter of variance $2\alpha^2$. We can therefore write the observed image as

$$I(x, y) = \rho_s \tilde{L}(x, y) * G_\alpha(x, y) + \rho_d D \tag{5.1}$$

where $I(x, y)$ is the projected illumination, G_α is a radially symmetric Gaussian ker-

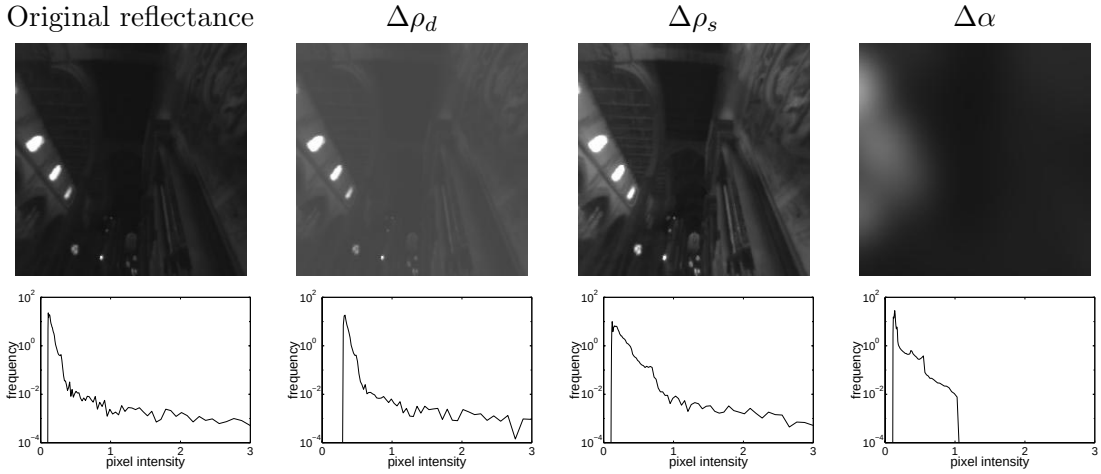


Figure 5.5. Effect of changes in Ward model reflectance parameters on an image and its pixel intensity histogram. The top row shows images of a flat surface under the illumination map of Figure 5.2(a). The bottom row shows corresponding histograms for each image, with logarithmic vertical axes. The leftmost image has reflectance parameter values $\rho_d = .1$, $\rho_s = .1$, and $\alpha = 0$. The remaining images have the same parameter values as the first except that ρ_d increases to .3 in the second column, ρ_s increases to .3 in the third column, and α increases to .1 in the fourth column. The surface normal direction is 45° from the vertical in each case, the view direction is perpendicular to the surface, and the images have a 90° field of view.

nel $G_\alpha(x, y) = \frac{1}{4\pi\alpha^2} \exp(-\frac{x^2+y^2}{4\alpha^2})$, and D is a measure of overall illumination intensity that is constant across the surface. Over a field of view of 90° or less, this simple approximation produces renderings that are visually almost indistinguishable from renderings computed directly from Equation (3.1). The top row of Figure 5.5 shows the effect of varying each reflectance parameter on an image of a tile rendered under one photographically-acquired illumination map.

Note that D represents the radiance of a perfectly reflective Lambertian surface, positioned perpendicular to the viewer. To resolve the ambiguity between overall scaling of illumination and reflectance, we will assume that an accurate estimate of D is available for each illumination (see Section 3.4). We therefore normalize the rendered image by D:

$$\frac{I(x, y)}{D} = \rho_s \frac{\tilde{L}(x, y)}{D} * G_\alpha(x, y) + \rho_d. \tag{5.2}$$

■ 5.1.1 Statistics of Intensity Distribution

Figure 5.5 shows the pixel histogram of the image observed under each reflectance parameter setting. Increasing the diffuse reflectance ρ_d adds a constant to the entire image, shifting the pixel histogram uniformly to the right. Increasing the specular reflectance ρ_s multiplies the specular component of the image by a constant, stretching the histogram to the right. Increasing the roughness α blurs the image, making the

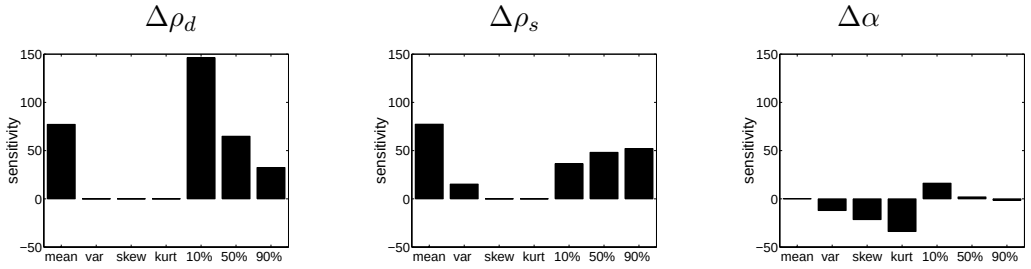


Figure 5.6. Sensitivity of pixel histogram statistics to ρ_d, ρ_s, and α. The bars in each plot represent sensitivity of the mean, variance, skewness, kurtosis, 10th percentile, median, and 90th percentile of the pixel intensity distribution. Sensitivity values were computed at parameter values $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .01$, using the Teller illumination set.

pixel histogram more symmetric and reducing its heavy positive tail.

Although the observed image of a surface will vary under different illuminations, the pixel histograms of different real-world illumination maps tend to have similar shapes (Section 4.4.1). This suggests that we may be able to find summary statistics, based on the distribution of pixel intensities, that depend on reflectance but vary little from one real-world illumination to another. We define the "sensitivity" of an image statistic to a reflectance parameter as the ratio between the derivative of the statistic with respect to the parameter and the standard deviation of the statistic across illuminations.[1] This sensitivity measure depends on the values of the reflectance parameters at which the derivatives are evaluated. Nevertheless, sensitivity values indicate which image statistics provide reliable information about each reflectance parameter. Large sensitivity values, either positive or negative, indicate that variations in a particular statistic due to change in a parameter are large relative to variations of that statistic between illuminations. To compute sensitivity values, we need a database of real-world illuminations. We computed sensitivity values separately for the Teller and Debevec illumination sets (described in Chapter 4), and obtained similar results. We present values computed for the Teller illumination set in this section.

Figure 5.6 shows sensitivities of several image statistics to each of the Ward model reflectance parameters. The image statistics in this figure are simply moments and percentiles of the pixel intensity distribution. Figure 5.6(a) suggests that the 10th percentile of pixel intensity may be particularly informative in estimating ρ_d. Most illumination maps contain regions of low illumination, where the specular component contributes little to observed radiance. The darkest areas of the observed image, as measured by the 10th percentile of pixel intensities, therefore prove indicative of the

[1] Let x_i denote the values of a particular statistic at a fixed parameter setting for illuminations $i = 1, 2, \ldots, N$. Let $\tilde{x}_i$ be the corresponding value of the statistic for each illumination when a particular parameter increases by a small quantity Δp. Let μ and σ denote the mean and standard deviation of x_i, while $\tilde{\mu}$ and $\tilde{\sigma}$ denote the corresponding quantities for $\tilde{x}_i$. We measure the local sensitivity of the statistic to the reflectance parameter as $\frac{\tilde{\mu}-\mu}{\Delta p\sqrt{(\sigma^2+\tilde{\sigma}^2)/2}}$.

diffuse reflectance of the surface.[2]

The pixel intensity distribution also provides some information about specularity. Increasing the strength of specular reflectance ρ_s scales the specular component of the reflectance, increasing the mean and variance of the pixel distributions. Increasing the surface roughness α tends to eliminate the heavy positive tail of the pixel distribution, reducing its variance, kurtosis, and skewness.

■ 5.1.2 Statistics of Marginal Distributions of Wavelet Coefficients

We can capture more information about ρ_s and α by examining statistics that capture spatial structure of the image. Motivated by the results of Chapter 4, we use wavelet domain statistics to describe this structure.[3] The variance of wavelet coefficients at a particular scale and orientation provides a measure of total spectral energy in a certain frequency band. As discussed in Chapter 4, distributions of wavelet coefficients also capture additional information about natural images and illuminations not specified by their second order statistics.

We found that the precise choice of wavelet basis is not critical in gathering image statistics. We chose to use a quadrature mirror filter (QMF) pyramid based on nine-tap even-symmetric filters [98].[4] We summarized the distributions of pyramid coefficients at each scale and orientation by computing moments and percentiles, as we did for the distributions of pixel coefficients in Section 5.1.1. Figure 5.7 shows the sensitivities of these statistics to the α and ρ_s reflectance parameters. The statistics are invariant with respect to the ρ_d parameter.[5]

Varying ρ_s scales the specular component of the image. For this simplified geometry, where the diffuse component is constant across the image, varying ρ_s also scales all the wavelet coefficients, regardless of orientation or scale, by a constant factor. Increasing ρ_s scales the coefficient distribution in each subband, without changing its shape. The 10th and 90th percentiles of the distribution scale with ρ_s, and the variance of the distribution scales with ρ_s^2, but the skewness and kurtosis do not change. The presence of the sun in some of the Teller illuminations leads to significant deviations in the

[2]For this simple geometry, where the diffuse component is constant across the image, using a minimum of the observed luminance gives an even higher sensitivity. However, the 5th or 10th percentile is more robust to image noise and proves a more reliable indicator of diffuse reflectance for curved surfaces, for which the diffuse image component varies spatially.

[3]Once again, we use the term "wavelet" loosely to refer to oriented bandpass filters computed at multiple scales.

[4]The use of even-symmetric filters in this chapter is significant only in our examination of skew in wavelet coefficient distributions, at the end of this section; we do not use the skewness in the reflectance classification algorithms of Chapter 6. In our examination of illumination statistics, described in Chapter 4, we originally chose odd-symmetric (8-tap) filters in order to compare our results to those of Huang and Mumford [47]. Both pyramids were computed using Eero Simoncelli's matlabPyrTools toolbox, available at *http://www.cns.nyu.edu/∼eero/software.html*. We modified the buildWpyr function to avoid subsampling the pyramid at each scale.

[5]Because the basis functions of the pyramid are not exactly zero-mean, we subtract the image mean before computing the pyramid coefficients.

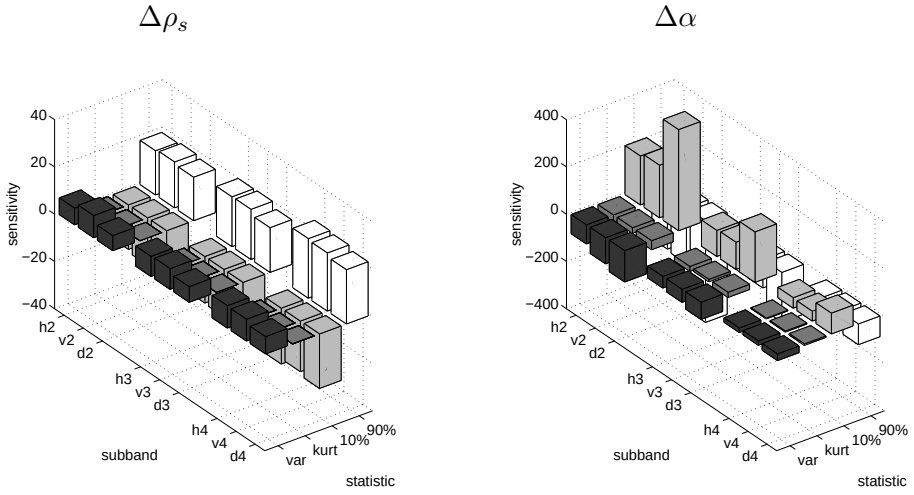


Figure 5.7. Sensitivities of wavelet coefficient statistics to the ρ_s and α parameters. Statistics shown are the variance, kurtosis, 10th percentile and 90th percentile of the horizontal, vertical, and diagonal wavelet coefficients at the second, third, and fourth finest scales. As in Figure 5.6, derivatives are computed at parameter values $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .01$.

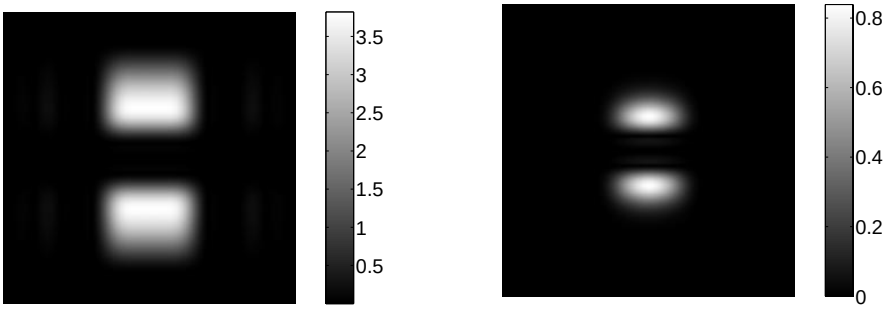
variance of coefficients from image to image, so the 10th and 90th percentiles of the subband coefficients give higher sensitivity measures than the variance (Figure 5.7).

The effects of changing α are more subtle. Blurring the specular component of the image lowers the energy in each high-frequency subband, but how much will it lower the energy, and how will it affect the shape of the coefficient distribution? To facilitate this analysis, note that applying a Gaussian blur to the image and then computing wavelet coefficients is equivalent to applying the same blur to the wavelet basis functions before using them to compute coefficients for the original image. Convolution of the basis functions with the Gaussian kernel G_α amounts to multiplication of their frequency response[6] by $e^{-\omega^2\alpha^2}$. This multiplication decreases the response of the filter at every nonzero frequency, particularly at high frequencies. When applied to a basis function whose frequency response is flat over some passband and zero elsewhere, it tends to produce a filter with a narrower radial passband. Figure 5.8 shows the frequency response of one basis function of the QMF pyramid before and after Gaussian blur.[7]

Field has observed that the kurtosis of filter outputs for natural images tends to peak for a radial filter bandwidth around one octave [32]. The QMF has a bandwidth of one octave, and we indeed find that narrowing that bandwidth through Gaussian filtering

[6] We use both f and ω to denote the modulus of frequency. The units of f are cycles per unit distance while those of ω are radians per unit distance, so $\omega = 2\pi f$.

[7] The effects of a Gaussian blur on the passband of the basis functions depend on the wavelet basis. In particular, the steerable pyramid filters proposed by Simoncelli *et al.* [100] fall off slowly as radial frequency decreases, so multiplication by a Gaussian may actually increase their effective bandwidth. The nine-tap QMFs used in this section also pass some energy at very low frequencies. Under sufficient blur, this energy begins to dominate the frequency response.



(a) Frequency response of original basis function

(b) Frequency response of blurred basis function

Figure 5.8. (a) Frequency response of the second-finest horizontal basis function of a QMF pyramid. The center of the image represents zero frequency. (b) Frequency response of the same basis function after convolution with a Gaussian filter.

tends to lower the kurtosis. However, this effect is minor (Figure 5.7); bandpass filters of various bandwidths all produce highly kurtotic distributions when applied to natural images or illumination. The primary effect of image blur on the distributions of wavelet coefficients is simply to reduce their scale. Thus the kurtosis of wavelet coefficients does not provide a reliable indicator of image blur.

Increasing α and decreasing ρ_s have similar effects on the distribution of coefficients in an individual wavelet subband. While decreasing ρ_s affects each subband similarly, however, increasing α affects them differently. Blurring reduces power more for higher frequency subbands. The Gaussian filter has frequency response $F(\omega) = e^{-\omega^2\alpha^2}$, so its derivative with respect to α is

$$\frac{dF}{d\alpha} = -\alpha\omega^2 e^{-\omega^2\alpha^2}. \tag{5.3}$$

This derivative attains a maximum with respect to frequency when $\omega = 1/\alpha$. The output variance for a filter whose passband is centered around frequency ω_o will therefore be most sensitive to blur parameters near $1/\omega_o$. At the value of α used to compute the sensitivities in Figure 5.7, distributions of coefficients at the second-finest scale show the greatest sensitivity to α.

The fact that a change in the roughness α affects different subbands differently suggests a means to distinguish an increase in α from a decrease in ρ_s. For a given illumination map, the ratio of coefficient variances for a high frequency subband and a low-frequency subband remains constant as ρ_s changes. On the other hand, this ratio decreases as α increases. Despite the variability of power spectra of real-world illuminations noted in Chapter 4, power falls off with some regularity at high frequencies (Figure 4.9b). Ratios of variances of subbands at different scales therefore serve as reli-

able indicators of α.[8] For the illumination set and reflectance parameters of Figure 5.7, for example, the variances of horizontal coefficients at the second- and third-finest scales have sensitivities with respect to α of -81 and -49, while the sensitivity of their ratio is -623.[9] Unlike the variance or percentiles of the individual bands, this ratio statistic is invariant to changes in ρ_s.

We chose to use even-symmetric filters in the pyramidal decomposition because we hypothesized that the skewness of each wavelet coefficient distribution might provide information about reflectance. In real-world illumination maps, intense light sources surrounded by dark areas are more common than very dark areas surrounded by intense light. This often leads to skewed distributions of wavelet coefficients, but we were not able to identify relatonships between reflectance properties and wavelet coefficient skewness that were sufficiently consistent to be useful in the reflectance classification tasks of Chapter 6.

■ 5.1.3 Statistics of Joint Distributions of Wavelet Coefficients

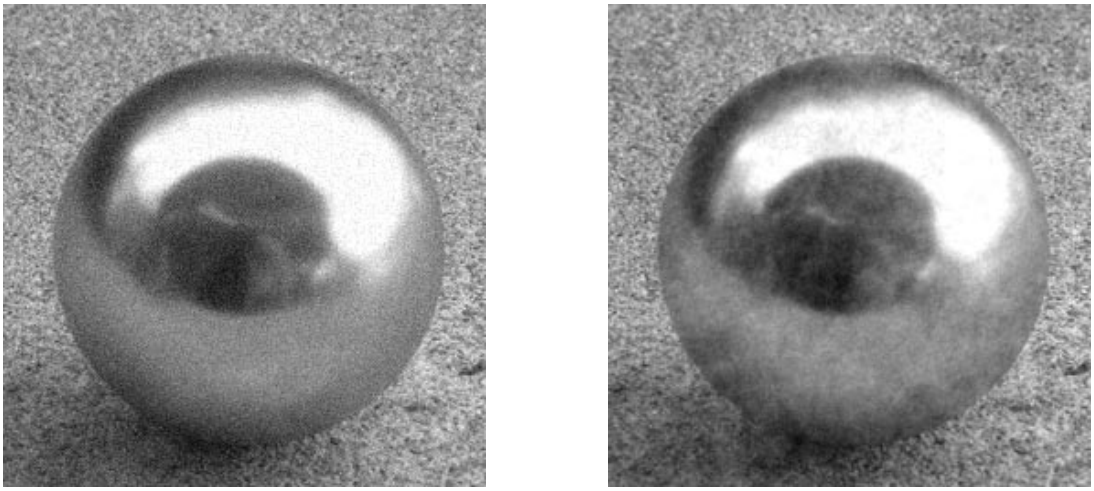
Statistics that depend only on marginal distributions of wavelet coefficients, including ratios of such statistics across scale, are sensitive to image noise. Figure 5.9 shows the rough metal sphere image of Figure 5.1(d) corrupted by white and pink Gaussian noise. The noise introduces power at high spatial frequencies, altering the variances and ratios of variances of the fine-scale wavelet coefficients. The image of Figure 5.9(a) actually has more power at high frequencies than a noise-free image of a chrome sphere under the same illumination (Figure 5.1(c)). Yet, neither of the images in Figure 5.9 appear any more chrome-like than the image of Figure 5.1(d). A human observer easily distinguishes the effect of introducing high-frequency noise from that of sharpening the specular reflectance.

A computational algorithm can distinguish between high-frequency image content due to noise and that due to specular reflection by taking advantage of the cross-scale dependencies of wavelet coefficients. Real-world illumination maps tend to contain edge-like structures that introduce large wavelet coefficients in corresponding positions at different scales (Section 4.6). The magnitudes of wavelet coefficients at a particular image location therefore tend to be strongly correlated across scales. Noise such as that of Figure 5.9 lacks this property.

A measure of specular energy or sharpness that is robust to image noise should

[8] If one assumes that real-world illuminations have power spectra of the form $\frac{k}{f^{2+\eta}}$, then blur by a Gaussian results in a spectrum $\frac{k}{f^{2+\eta}} \exp\left(-(2\pi f)^2 \alpha^2\right)$. Fitting such a model to the high-frequency portion of the spectrum of images rendered using Equation (5.1) gives an accurate estimate for α, but the accuracy of such an estimate degrades in the presence of noise or edge effects. We found that, in practice, the reflectance classifiers discussed in Chapter 6 performed equally well with simple ratios of variances.

[9] Note that large negative sensitivity values, like large positive values, indicate informative statistics. These sensitivity values are negative because the variances of wavelet coefficients at a particular scale decrease as α increases, with the variance of fine-scale coefficients decreasing faster than that of coarse-scale coefficients.



(a) White noise (b) Pink noise

Figure 5.9. The image of Figure 5.1(d), corrupted by (a) Gaussian noise with a flat power spectrum and (b) Gaussian noise with a power spectrum that falls of with the square of frequency. In both cases, the standard deviation of the noise is 20% that of the standard deviation of the noise-free image.

respond preferentially to large fine-scale wavelet coefficients that coincide spatially with large coefficients at coarser scales. One such measure is the covariance of the magnitude of coefficients $W_j(x, y)$ at scale j with the magnitude of coefficients $W_k(x, y)$ at a coarser scale k but at the same orientation:

$$c_{jk} = \lambda_{|W_j||W_k|} = E\big[(|W_j| - E[|W_j|])(|W_k| - E[|W_k|])\big]. \tag{5.4}$$

This statistic depends on the absolute scale of the wavelet coefficients in both bands. Like the variance of coefficients in a single band, it is sensitive to changes in both ρ_s and α. If we divide c_{jk} by the variance of $|W_k|$, we obtain a statistic that is sensitive to changes in α but invariant to scaling of the specular component by ρ_s:

$$r_{jk} = \frac{\lambda_{|W_j||W_k|}}{\sigma^2_{|W_k|}} = \frac{E\big[(|W_j| - E[|W_j|])(|W_k| - E[|W_k|])\big]}{E\big[(|W_k| - E[|W_k|])^2\big]}. \tag{5.5}$$

The statistic r_{jk} typically has a sensitivity to α slightly lower than the sensitivity of the variance ratio $\frac{\sigma_j^2}{\sigma_k^2}$ of the corresponding bands, but proves significantly more robust to noise. For the reflectance parameters and illuminations of Figure 5.7, for example, the sensitivity of r_{23} is -623 and that of $\frac{\sigma_2^2}{\sigma_3^2}$ is -498. Decreasing α from .01 to .005 in the absence of noise increases the mean of $\frac{\sigma_2^2}{\sigma_3^2}$ over all illuminations approximately

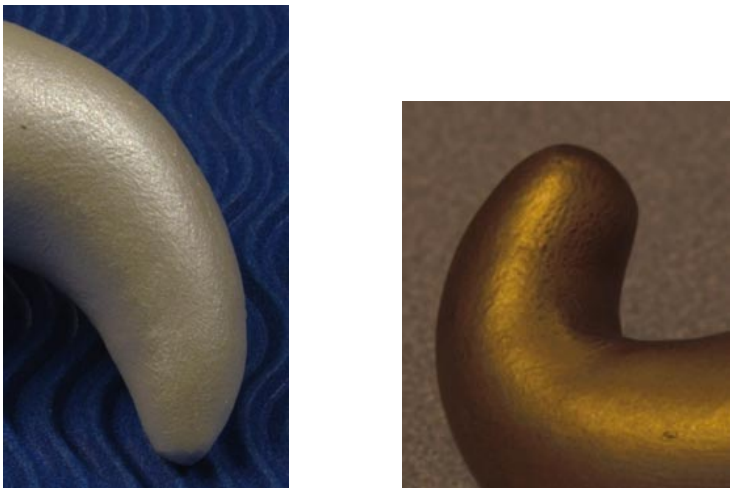


Figure 5.10. Two painted clay surfaces. The visible bumpiness of these surfaces introduces high spatial frequencies in the image.

four-fold and the mean of r_{23} two-fold.[10]. However, adding a small amount of white noise increases the mean of $\frac{\sigma_2^2}{\sigma_3^2}$ by a factor of 2.6 while changing that of r_{23} by only 21%. The same noise increases the mean of $\frac{\sigma_1^2}{\sigma_2^2}$ fifty-fold while that of r_{12} increases by less than 10%.

Statistics that exploit the dependencies of wavelet coefficients across scale are also useful in distinguishing reflected illumination patterns from high-frequency intensity variation due to three-dimensional texture, dirt, or surface imperfections. For example, Figure 5.10 shows images of two painted clay surfaces with visible fine-scale variations in surface normal. Although this variation introduces high frequencies in the image, the spatial structure is different from that of a reflected illumination map. We found that the statistics c_{jk} and r_{jk} improved the accuracy with which we could recognize such surfaces from photographs (Chapter 6). We discuss three-dimensional surface texture further in Section 7.2.5.

The relationships between reflectance parameter values and the observed image become more complicated as we relax assumptions on the geometry of the surface and the position of the viewer. For curved geometries such as a sphere, for example, the

[10]For reflectance parameters $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .01$, the mean values of $\frac{\sigma_2^2}{\sigma_3^2}$ and $\frac{\sigma_1^2}{\sigma_2^2}$ in the absence of noise were .013 and .0088, respectively. Adding white Gaussian noise with standard deviation approximately 4% the mean value of the specular image component increased the mean values of $\frac{\sigma_2^2}{\sigma_3^2}$ and $\frac{\sigma_1^2}{\sigma_2^2}$ to .035 and .44, respectively. For the same reflectance parameters, the respective means of r_{23} and r_{12} were .076 and .047 in the noise-free case and .061 and .066 after addition of noise. For reflectance parameters $\rho_d = .25$, $\rho_s = .05$, and $\alpha = .005$, the mean values of $\frac{\sigma_2^2}{\sigma_3^2}$ and r_{23} in the absence of noise were .055 and .15, respectively.

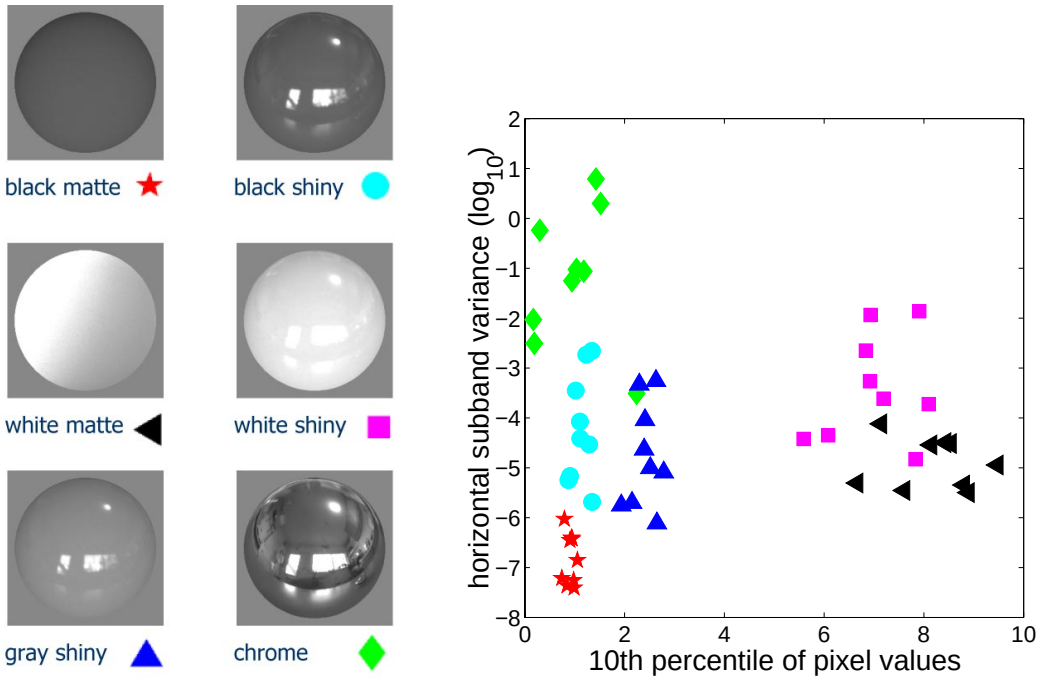


Figure 5.11. At left, synthetic spheres of 6 different reflectances, each rendered under one of Debevec's illumination maps. Ward model parameters are as follows: black matte, $\rho_d = .1$, $\rho_s = 0$; black shiny, $\rho_d = .1$, $\rho_s = .1$, $\alpha = .01$; white matte, $\rho_d = .9$, $\rho_s = 0$; white shiny, $\rho_d = .7$, $\rho_s = .25$, $\alpha = .01$; chrome, $\rho_d = 0$, $\rho_s = .75$, $\alpha = 0$; gray shiny, $\rho_d = .25$, $\rho_s = .05$, $\alpha = .01$. We rendered each sphere under the nine photographically-acquired illuminations depicted in Figure 2.7 and plotted a symbol corresponding to each in the two-dimensional feature space at right. The horizontal axis represents the 10th percentile of pixel intensity, while the vertical axis is the log variance of horizontally-oriented QMF wavelet coefficients at the second-finest scale, computed after geometrically distorting the original image as described in Section 6.1.2.

radiance of the diffuse component varies across the surface. This slow variation tends to decrease the kurtosis of the wavelet coefficient distributions because it increases their variance without contributing extreme values. This decrease in kurtosis becomes more pronounced when the variance of wavelet coefficients due to the specular component alone decreases. Kurtoses of the wavelet coefficient distributions do therefore carry information about Ward model reflectance parameters for curved surfaces. As α increases, for example, the contribution of the specular component to the fine-scale wavelet coefficients tends to drop relative to the contribution of the diffuse component, so the kurtosis decreases.

Figure 5.11 shows how differently illuminated spheres of a particular reflectance tend to cluster in a space defined by appropriately chosen image statistics. We rendered six spheres of different reflectances under nine photographically-acquired illumination maps, and plotted the position of each image in a two-dimensional feature space. The horizontal axis represents the 10th percentile of pixel intensities, while the vertical

axis represents the variance of wavelet coefficients at one scale. The former correlates strongly with diffuse reflectance, while the latter depends on the strength and sharpness of the specularity.

■ 5.2 Reflectance Classification Based on Image Features

One can exploit the relationships between image statistics and reflectance either to classify reflectance (Section 3.3.2) or to estimate reflectance parameters (Section 3.3.1). We focus on the classification approach, because it allows us to handle reflectances not described by a simple parameterized model and even to build a classifier using photographs of real surfaces whose BRDFs are unknown. We attempt to select from a finite set of reflectances the one that most closely represents the reflectance of an observed surface, as illustrated in Figure 3.7. The candidate reflectances may be defined by different parameter settings of a reflectance model such as the Ward model, but they could also be arbitrary BRDFs specified as lookup tables or by a set of photographs of a real surface. Given images of several surfaces under various illuminations, we wish to classify photographs under unknown, novel illumination according to surface reflectance.

Although the analysis of Section 5.1 guides our choice of image statistics, we determine the relationships between reflectance classes and image statistics from empirical data. This transforms the problem to a machine learning problem. We choose a set of image features, and then divide the feature space into regions corresponding to different reflectances based on a set of example images of each. Figure 5.12 shows the regions of a feature space assigned to different reflectance classes by a support vector machine learning algorithm described in Section 6.1.1, using the examples of Figure 5.11 as training data.

For purposes of illustration, the classifier in Figure 5.12 uses only two image statistics. It incorrectly classifies three of the training images; it will misclassify images under novel illumination at least as often. By adding additional image statistics as features, however, one can improve classifier performance. In Chapter 6, we show that a classifier based on six image statistics achieves nearly perfect performance in cross-validation tests for images of these six reflectances under real-world illumination conditions.

An alternative approach to reflectance classification involves explicit derivation of the distributions of image statistics for each reflectance, under an assumed probabilistic model for illumination. Such an approach requires the determination of the joint probability distribution of a set of image statistics given the surface reflectance. Unfortunately, an accurate generative model for illumination is not readily available, although several authors have proposed generative models that capture some of the statistics of natural images [40, 58, 90]. The learning approach also offers important practical advantages over analytic derivation:

- One can train a learning-based classifier with photographs of surfaces of known material types whose BRDF has not been measured explicitly. Accurate measure-

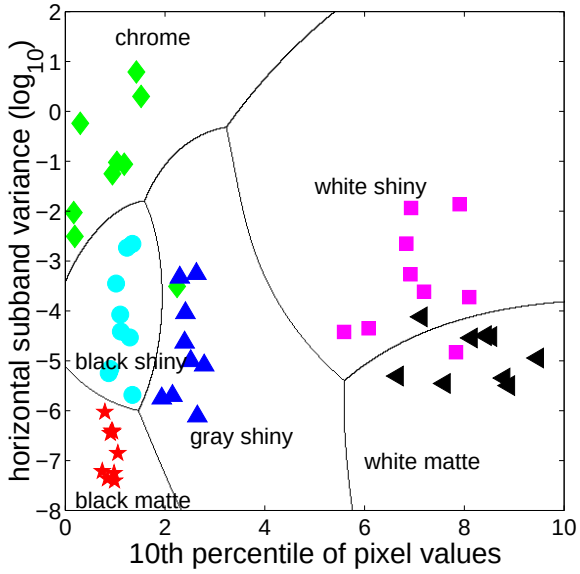


Figure 5.12. The curved lines separate regions assigned to different reflectances by a simple classifier based on two image features. The training examples are the images described in Figure 5.11. The classifier is a one-versus-all support vector machine, described in Section 6.1.1. Using additional image features improves classifier performance.

ment of a BRDF is a laborious process because of the high dimensionality of the data involved.

- A learning-based classifier has greater flexibility, as it can be trained to compensate for variations in imaging geometry, surface texture, or camera nonlinearities. We return to these points in Chapter 6.

The machine learning approach suffers from several disadvantages relative to an approach based on explicit estimator derivation:

- In practice, collection of training data is time-consuming because it requires photography of surfaces. One can also render training images for surfaces with known BRDFs, but this requires real-world illumination maps, which in turn must be collected photographically.
- Without assumptions on the ensemble of images to be classified, one cannot guarantee the performance of such a classifier theoretically. We therefore resort to empirical testing.

Whether we rely on machine learning methods or explicit derivation from a generative illumination model, the performance of a classification scheme depends on the statistical regularity of real-world illumination. A classifier trained on images under

such illumination will often misclassify surfaces rendered under synthetic illuminations with significantly different statistics. Human performance in reflectance recognition tasks also degrades under synthetic illumination (Section 2.3).

■ 5.3 Summary and Discussion

Certain statistics of an image of a surface act as useful indicators of reflectance because their dependence on reflectance properties is significant relative to their variation from one real-world illumination to another. We have identified several such statistics, based on distributions of pixel intensities, distributions of wavelet coefficients at various scales, and joint distributions of co-located wavelet coefficients at different scales.

One can classify surface reflectance on the basis of these statistics by using machine learning techniques to partition a feature space into regions corresponding to the various reflectance classes. One can train such a classifier using either photographs of surfaces in each reflectance class or synthetic images of surfaces rendered under photographically-acquired real-world illumination. Chapter 6 addresses issues involved in the design of an effective reflectance classifier. These issues include the choice of an appropriate machine learning technique, the choice of a specific set of image statistics as features for a particular classification problem, and compensation for differences in image statistics due to differences in surface geometry. The resulting reflectance classifiers rival human performance when surface geometry is known in advance and reflectance is homogeneous across the surface.

Chapter 6

Designing a Reflectance Classifier

The previous chapter showed that the statistical regularity of real-world illumination leads to predictable relationships between the reflectance of a surface and certain statistics of its image. That chapter also pointed out that one can use these statistics to classify surface reflectance on the basis of an image under unknown, real-world illumination. The present chapter discusses issues involved in building a practical reflectance classifier — in particular, the use of machine learning techniques for training a classifier on the basis of sample images (Section 6.1), and classification for different surface geometries (Section 6.2). Each section includes examples of classification from photographs as well as from synthetic images rendered under photographically-acquired illumination maps.

■ 6.1 Machine Learning Techniques for Reflectance Classification

To build a reflectance classifier, we select a set of image statistics as features and then partition the resulting feature space into regions corresponding to each reflectance class. We first consider the problem of classifier construction given a feature set, and then discuss the selection of a sensible feature set.

■ 6.1.1 Classification Techniques

Suppose that we wish to classify each image into one of m classes $r_1, \ldots, r_m$ on the basis of the statistical image features[1] $\mathbf{x} = \mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$ by selecting the most likely reflectance class given the observed feature values. A direct approach to constructing a maximum likelihood classifier involves estimating the probability distribution $p(\mathbf{x}|r_k)$ for each reflectance class from sample images of surfaces in that class. One then classifies a new sample by computing the probability of each reflectance class given the observed feature values:[2]

$$p(r_k|\mathbf{x}) = \frac{p(\mathbf{x}|r_k)p(r_k)}{p(\mathbf{x})} \tag{6.1}$$

[1] We use $\mathbf{x} = \mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$ to denote a random vector of features, and $\mathbf{x} = x_1, x_2, \ldots, x_n$ to denote a sample value of that random vector.

[2] Because k takes on discrete values, $p(r_k)$ and $p(r_k|\mathbf{x})$ represent discrete probability distributions, while $p(\mathbf{x}|r_k)$ represents a continuous distribution.

where $p(r_k)$ is the probability of observing reflectance class r_k. Equation (6.1) follows the Bayesian approach of Section 3.2.2, but assumes that the statistical features $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$ capture all information about the observed image that is relevant to classifying reflectance. The classifier selects the reflectance class r_k that maximizes $p(r_k|\mathbf{x})$.

Given an unlimited number of samples of each reflectance class, one could estimate the densities $p(\mathbf{x}|r_k)$ to arbitrary accuracy using non-parametric methods.[3] In practice, the number of available training samples is typically small relative to the number of features and reflectance classes, because either the training images themselves or the illuminations used to render them must be acquired photographically. In Section 6.1.2, for example, we discuss classifiers that distinguish between 9 reflectance classes in a 36-dimensional feature space, given only 6 sample images of each class. This amount of training data does not suffice to build an accurate non-parametric density model. One alternative is the use of a parametric density model. Unfortunately, we lack a convenient parametric form for these densities. In particular, a Gaussian distribution with a diagonal covariance matrix will typically provide a poor fit to the data, because many of the statistical image features discussed in Chapter 5 are highly correlated with one another. We generally lack sufficient data to fit Gaussian distributions with arbitrary covariance matrices. A classifier based on 36 features, for example, would require a covariance matrix with 666 independent elements to describe the distribution of each class in the feature space.

We turn instead to discriminative learning methods that separate classes without estimating the entire distribution $p(\mathbf{x}|r_k)$. One such method relies on support vector machines (SVMs). We provide a brief introduction to SVMs in the context of the present work, but refer the reader to several reviews [15, 88] for more thorough and general discussion.

The basic support vector machine operates as a binary classifier that separates two classes in an n-dimensional feature space using a linear boundary. If the class labels are denoted by $+1$ and -1, the classifier takes the form

$$f(\mathbf{x}) = \text{sign}(\mathbf{x} \cdot \mathbf{w} + b). \tag{6.2}$$

The parameters $\mathbf{w}$ and b that specify the separating hyperplane depend on a set of training samples[4] $\mathbf{x}_i$ and their corresponding labels y_i. One determines $\mathbf{w}$ and b by minimizing $\frac{1}{2}||\mathbf{w}||^2$ subject to the constraints

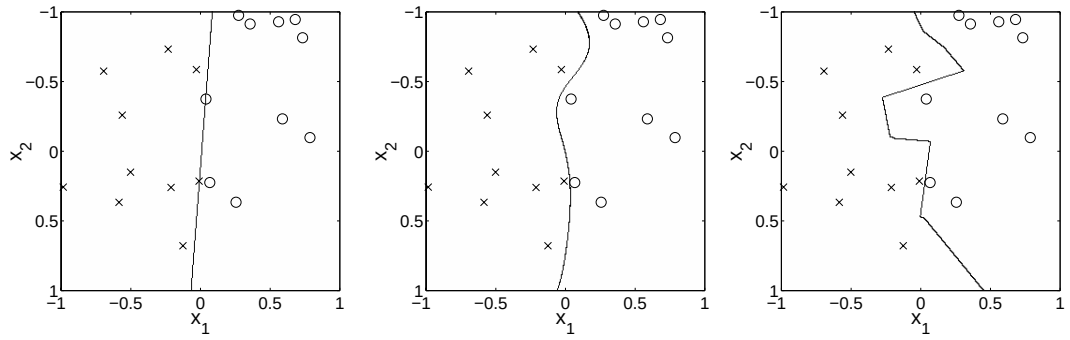
$$\mathbf{x}_i \cdot \mathbf{w} + b \;\geq\; +1 \quad \text{for} \quad y_i = +1 \tag{6.3}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \;\leq\; -1 \quad \text{for} \quad y_i = -1. \tag{6.4}$$

A solution satisfying these constraints exists only when the training samples are linearly separable — that is, when there exists a hyperplane in the feature space such that all

[3] For example, one might use a Parzen density estimator with arbitrarily small windows [7].

[4] Note that $\mathbf{x}_i$ denotes the feature vector corresponding to the ith training sample, while x_i denotes the ith element of feature vector $\mathbf{x}$.



(a) (b) (c)

Figure 6.1. Class boundaries assigned by three different discriminative classifiers on the basis of the same training data points, for a simple toy problem. Circles indicate data points of one class, while crosses indicate data points in the other. (a) Linear SVM. (b) Gaussian SVM with $\sigma^2 = 1$. (c) Nearest neighbor classifier.

the samples in one class lie on one side of the hyperplane, and all the samples in the other class lie on the other side. In this case, the SVM selects as a boundary the separating hyperplane that maximizes the classifier's "margin," defined as the minimum distance from the boundary to any point in the training set. Figure 6.1(a) shows the class boundary assigned by a linear SVM for a simple toy problem, where 10 training samples are available from each of two data sets. The values of x_1 were sampled from a uniform distribution between 0 and 0.5 for one class, and between 0.5 and 1 for the other class. The values of x_2 were sampled from a uniform distribution between 0 and 1 for both classes. Thus, the "ideal" boundary in this case is a line at $x_1 = 0.5$.

One can apply the SVM framework to non-separable data by adding "slack variables" ξ_i that specify classification error for each data point. In particular, one minimizes $\frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i$ subject to the constraints

$$\begin{aligned} \mathbf{x}_i \cdot \mathbf{w} + b &\geq +1 - \xi_i \quad \text{for} \quad y_i = +1 \\ \mathbf{x}_i \cdot \mathbf{w} + b &\leq -1 + \xi_i \quad \text{for} \quad y_i = -1 \\ \xi_i &\geq 0 \quad \forall i. \end{aligned} \tag{6.5}$$

The parameter C determines the relative importance of achieving a large margin and of minimizing error on the training set.

The optimization problem of Equation (6.5) can be solved using quadratic programming techniques. Analysis of the dual problem leads to a solution of the form

$$\mathbf{w} = \sum_{i=1}^{N_s} \alpha_i y_i \mathbf{s}_i \tag{6.6}$$

where the "support vectors" $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N$ are a subset of the training points $\mathbf{x}_i$ and the α_i are positive real weights. The support vectors typically lie close to the class boundary. The test function of Equation (6.2) can be written in terms of the support vectors as

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N_s} \alpha_i y_i \mathbf{s}_i \cdot \mathbf{x} + b\right). \tag{6.7}$$

One can generalize the SVM framework to allow non-linear boundaries between classes by mapping the data points to a higher-dimensional space and then performing linear classification in that space. The test function of Equation (6.7) depends on the test sample $\mathbf{x}$ only in terms of the dot product of $\mathbf{x}$ with training data points. Likewise, one can rewrite the optimization problems defined above such that they depend only on dot products $\mathbf{x}_i \cdot \mathbf{x}_j$ between training data points. One may avoid explicitly mapping the data points to a higher dimensional feature space by defining a kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, where Φ is the mapping from the original feature space to the higher-dimensional space. Equation (6.7) becomes

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b\right). \tag{6.8}$$

This "kernel trick" provides an efficient method to implement nonlinear SVMs. The mapping Φ need not even be defined explicitly, as long as a kernel is available. Training and testing a nonlinear kernel-based classifier is little more expensive than training the linear classifier. A particularly popular choice for the kernel function is the Gaussian kernel, defined by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right). \tag{6.9}$$

The kernel variance σ^2 determines the degree of regularization on the class boundary. Smaller values of σ allow the classifier to achieve better performance on the training data set, but may lead to overfitting and therefore to poorer performance on novel data. Figure 6.1(b) shows the boundary assigned by a Gaussian SVM with $\sigma^2 = 1$, given the same training data points as in Figure 6.1(a).

Researchers have proposed several methods for extending SVMs to multiclass problems in which one wishes to associate each possible feature vector with one of m classes [88]. The simplest of these is known as one-versus-all (OVA) classification. One trains m separate binary classifiers, each of which distinguishes a single reflectance class from all the other classes. To classify a novel feature vector $\mathbf{x}$, one evaluates $\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b$ for each of the m classifiers (see Equation (6.8)) and chooses the class corresponding to the largest value of the sum. Rifkin argues that the OVA scheme typically performs as well as or better than the more complicated alternatives [88].

Support vector machines are an attractive machine learning method for reflectance classification for several reasons:

- Statistical learning theory suggests that SVMs should generalize well given a limited amount of training data. The performance of a classifier on novel test data depends both on its ability to accurately fit the training data by describing a sufficiently complex boundary between classes, and on its ability to regularize that boundary in order to avoid overfitting the training data. SVMs are based on a theoretical foundation for balancing these two conflicting requirements [30,115].

- SVMs have been found empirically to outperform most other classification techniques, particularly in problems where the number of available training samples is small relative to the number of features and the number of classes [15,88]. For example, Ramaswamy *et al.* achieved high (78%) accuracy in assigning cancers to one of 14 tumor types on the basis of over 16,000 gene expression features, using SVMs constructed with a total of 143 training samples [87].

- The application of an SVM to novel data via Equation (6.8) is computationally efficient. Recent work has also led to more efficient methods for training SVMs [52,80,88]. Because we typically have a limited number of training images, training efficiency tends not to be a major issue for reflectance classification. Our experience is that the primary computational demand of SVM-based reflectance classification lies in feature computation, both in the testing and training phases. Some alternative classification methods, such as the nearest-neighbor scheme discussed below, lead to more computationally expensive classifiers.

- If the prior probability $p(r_k)$ of observing reflectance class k is known, SVMs provide a straightforward method to incorporate it into classifier design. The previously described training framework assumes that $p(r_k) = q(r_k)$, where $q(r_k)$ is the fraction of the training samples corresponding to class k. In the more general case where $p(r_k)$ and $q(r_k)$ may differ, one trains the SVM corresponding to the kth class in the OVA scheme by minimizing

$$\frac{1}{2}||\mathbf{w}||^2 + C_+ \sum_{\{i|y_i=k\}} \xi_i + C_- \sum_{\{i|y_i\neq k\}} \xi_i, \tag{6.10}$$

  subject to the constraints of Equation (6.5), with $C_+ = C\frac{p(r_k)}{q(r_k)}$ and $C_- = C\frac{1-p(r_k)}{1-q(r_k)}$. This is equivalent to replicating the training samples for each class such that the total number of training samples for a class is proportional to the probability of observing that class.

We built one-versus-all SVM classifiers using Version 1 of Collobert's freely available SVMTorch[5] software [18]. The design of a such a classifier involves several heuristic choices. Most importantly, one must select a kernel. Most authors either select a kernel arbitrarily or use some form of cross-validation, although some have observed

[5]The Torch library is available at *http://www.torch.ch*.

that Gaussian kernels with appropriately chosen variances tend to perform as well as or better than other common kernels on a variety of data sets [88]. We selected Gaussian kernels because we found that, on several data sets of real and synthetic images, they outperformed linear and quadratic kernels in cross-validation tests. We selected the kernel variance σ^2 and the training error weight C using a similar cross-validation approach. Rifkin points out that C should scale roughly inversely with the number of training samples, while σ^2 should increase with the number of features and decrease with the number of training samples [88]. He also notes that the optima for C and σ^2 tend to be broad; we made similar observations on the data sets discussed in Section 6.1.2. The SVM performance results reported in this thesis all use[6] $C = 100$ and $\sigma^2 = 12.5$. One could obtain slightly better performance results by adjusting these parameters according to the training data set.

The Gaussian kernel depends on the distance between data points $\|\mathbf{x}_i - \mathbf{x}_j\|$. Scaling the units of a feature therefore changes its relative importance in the classification scheme. For each reflectance classification problem discussed in the following sections, we scaled the features to have identical mean magnitudes across training illuminations for one particular "standard" reflectance. For each classification problem, we choose a reflectance with large specular and diffuse components as the standard. The classifier therefore considers proportional changes in feature values relative to their typical values for the standard reflectance. Changing the standard reflectance typically has only a minor effect on classifier performance.

While we have obtained better reflectance classification results using support vector machines than other methods, the use of SVMs is not critical to the construction of an effective reflectance classifier. Other multiclass discriminative classification methods with regularization on the class boundaries produce similar results. In particular, we compared the performance of SVM classifiers to nearest-neighbor classifiers, which assign a test feature vector to the class associated with the closest training point in feature space. Figure 6.1 illustrates the boundary assigned by this classifier in a simple problem.

The nearest-neighbor classifier will correctly classify the training data, no matter how complex a boundary is required. It therefore suffers from a tendency to overfit the training data, particularly when abundant training data is available. One can regularize the classifier by using a k-nearest neighbor (k-NN) scheme, which assigns a test feature vector to the class best represented among the k closest training samples. The parameter $k \geq 1$ may be set proportional to the density of the training points in the feature space. In Section 6.1.2, we compare the performance of SVM, NN, and k-NN classifiers for reflectance classification problems.

[6] The SVMTorch software, like much of the SVM literature, defines the Gaussian kernel as $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{\sigma^2}\right)$. In this notation, σ^2 is double the variance, so we set $\sigma = 5$.

■ 6.1.2 Feature Selection and Classifier Performance

The analysis of Chapter 5 suggests use of the following image statistics as features for reflectance classification:

- The mean and 10th percentile of pixel values
- The variance of the coefficients in each wavelet subband
- The kurtosis of the coefficients in each wavelet subband
- The ratios of variances of coefficients in wavelet subbands of identical orientation at different scales
- The cross-scale statistics r_{jk} and c_{jk} defined by Equations (5.4) and (5.5), for wavelet coefficients of identical orientation at different scales

The classifiers discussed in this section use wavelet coefficient statistics based on the quadrature mirror filter pyramid of Chapter 5. Our experience is that the precise choice of pyramidal decomposition is not critical in gathering image statistics. The basis functions of the QMF pyramid have power spectra that fall off quickly outside their single-octave bandwidth (Figure 5.8). Ratios of variances at successive scales are most sensitive to changes in specular blur when the power spectra of basis functions at successive scales do not overlap, so the QMF decomposition leads to slightly better classifier performance than the difference of Gaussian filters of Section 6.2.

The set of statistics listed above is redundant in the sense that some of the statistics can be expressed in terms of others — for example, ratios of variances of wavelet coefficients are deterministic functions of individual variances. We found that including these redundant statistics improves the separability of the classes in the feature space. Likewise, we found that taking logarithms of the statistics computed in the wavelet domain improves classifier performance by giving each class a more symmetric distribution in the feature space.

The examples of this section involve photographs and rendered images of spheres. Section 6.2 considers classification for surfaces of other geometries and shows that such problems can be expressed in terms of classification of spheres. However, the geometry of a spherical surface leads to variations in image statistics between different regions of the image. The image of a chrome sphere, for example, is simply a distorted version of an illumination map, but each region of the illumination map is distorted differently. We have found that classifier performance typically improves if one "unwraps" the observed image as shown in Figure 6.2 before computing image statistics. A vertical cross-section of the unwrapped image corresponds to a radial cross-section of the original image. Appendix C justifies this unwrapping procedure in terms of the effect of surface geometry on local image statistics.

Table 6.1 summarizes the performance of reflectance classifiers based on various image statistics for four different image sets. Three sets consist of synthetic images rendered under photographically-acquired illumination maps, while the fourth set consists

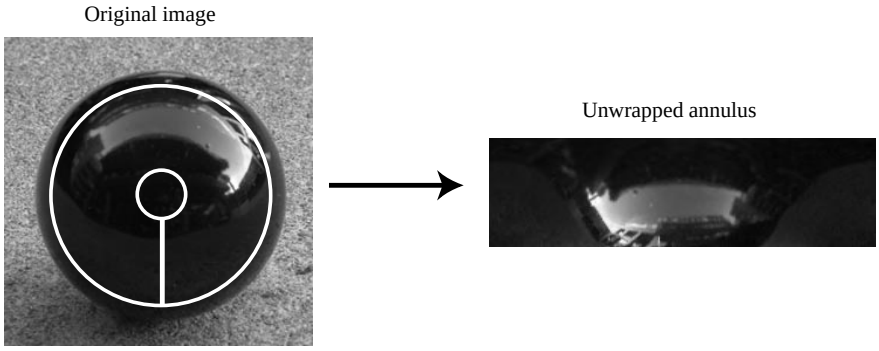


Figure 6.2. Before computing statistics on the image of a sphere, we extract an annulus of the image and unwrap it using a polar-to-rectangular coordinate transformation. In the examples of this section, the annulus omits the inner 20% and the outer 10% of each radial cross-section, because image statistics of the unwrapped image would vary most significantly in these regions (Appendix C).

	Image set			
Feature set	(1) 6 classes, 9 Debevec illums (rendered)	(2) 6 classes, 95 Teller illums (rendered)	(3) 11 classes, 95 Teller illums (rendered)	(4) 9 classes, 7 illums (photos)
(a) Chance	16.7	16.7	9.1	11.1
(b) Image mean	51.9	72.3	26.6	42.9
(c) Full set (4 levels)	81.5	98.8	98.9	81.0
(d) Full set (3 levels)	81.5	99.3	99.1	79.4
(e) No r_{jk}, c_{jk} (3 levels)	87.0	99.5	98.9	81.0
(f) 6 selected statistics	100.0	99.8	98.0	93.7
(g) Not unwrapped (no r_{jk}, c_{jk}; 3 levels)	90.7	98.2	96.8	74.6

Table 6.1. Cross-validation performance (in % accuracy) for SVM classifiers using different features on different image sets.

of photographs of surfaces. The reflectances of the surfaces in the rendered images were specified by the Ward model, described in Sections 2.1.2 and 5.1. The first two image sets included spheres of the 6 reflectances shown in Figure 5.11. The third set included spheres of 11 reflectances, shown in Figure 6.3. These 11 reflectances were chosen to present a more difficult classification problem in which several spheres share each total reflectance ($\rho_s + \rho_d$). Each sphere in the first image set was rendered under each of the nine spherical illumination maps due to Debevec *et al.* (Section 4.2). Each sphere

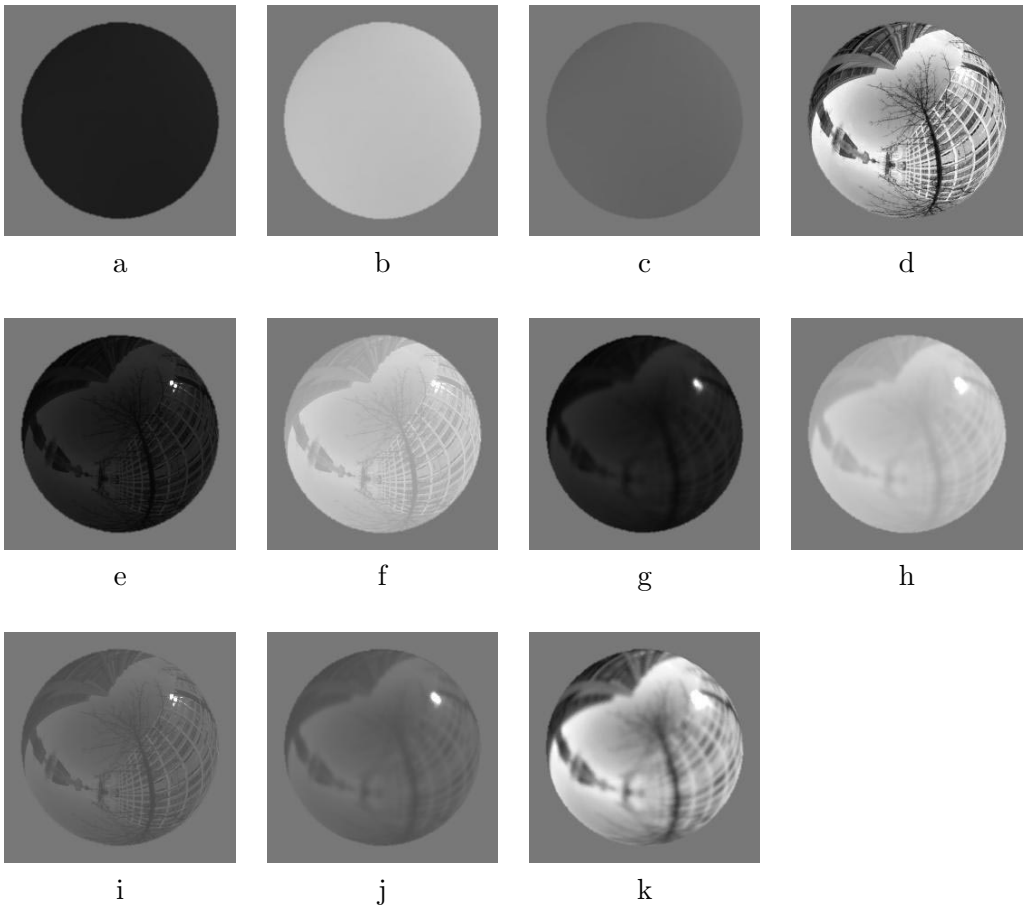


Figure 6.3. Synthetic spheres representing 11 reflectances, each rendered under one of Teller's illumination maps. Ward model parameters are as follows: (a) $\rho_d = .1$, $\rho_s = 0$, (b) $\rho_d = .9$, $\rho_s = 0$, (c) $\rho_d = .35$, $\rho_s = 0$, (d) $\rho_d = 0$, $\rho_s = .75$, $\alpha = 0$, (e) $\rho_d = .05$, $\rho_s = .05$, $\alpha = 0$, (f) $\rho_d = .7$, $\rho_s = .2$, $\alpha = 0$, (g) $\rho_d = .05$, $\rho_s = .05$, $\alpha = .02$, (h) $\rho_d = .7$, $\rho_s = .2$, $\alpha = .02$, (i) $\rho_d = .25$, $\rho_s = .1$, $\alpha = 0$, (j) $\rho_d = .25$, $\rho_s = .1$, $\alpha = .02$, (k) $\rho_d = 0$, $\rho_s = .75$, $\alpha = .02$. The total reflectance $\rho_s + \rho_d$ is .1 for (a), (e), and (g); .35 for (c), (i), and (j); .75 for (d) and (k); and .9 for (b), (f), and (h).

in the second and third image sets was rendered under the 95 illumination maps due to Teller *et al.* (Section 4.2). Because Teller's illumination maps do not capture the entire lower hemisphere of the environment, we mirrored the upper hemisphere into the lower hemisphere before rendering surfaces under these illuminations. Each sphere was rendered from a view angle midway between a top view and a horizontal view. The pixel values of the images used to train and test the classifier were linear in luminance, although we applied a compressive nonlinearity to produce the images shown in figures.

We acquired the fourth image set by photographing nine spheres made of different materials under seven diverse illumination conditions, including both indoor and outdoor settings. Figure 6.4 shows photographs of all nine spheres under one illumination,

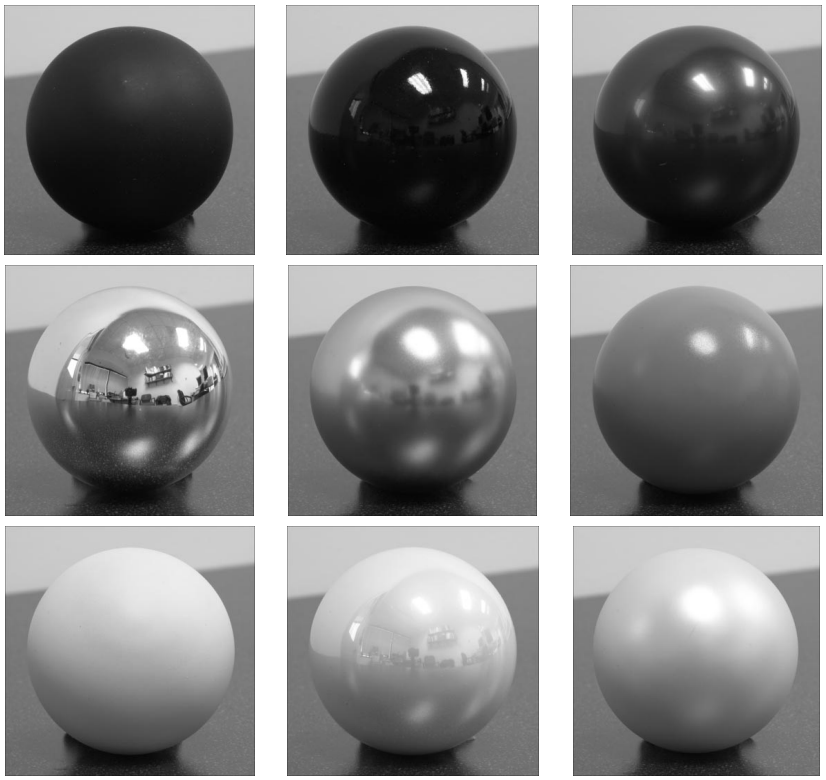


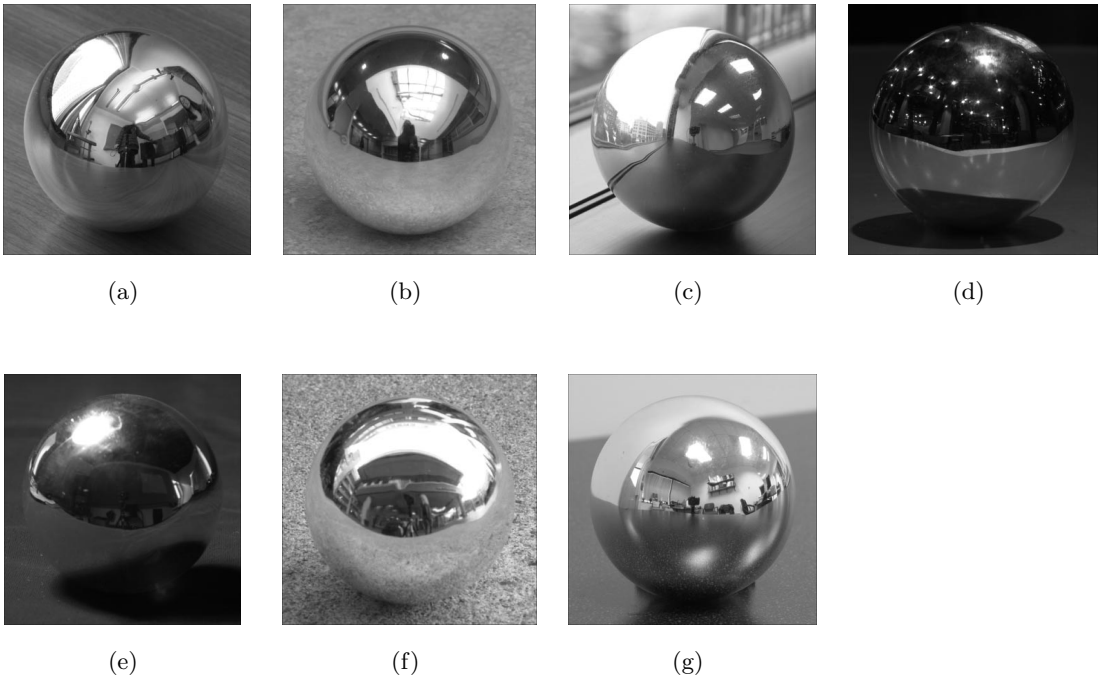
Figure 6.4. Photographs of nine spheres, all under the same illumination. Our classifiers use image data only from the surface itself, not from the surrounding background.

while Figure 6.5 shows photographs of one sphere under all seven illuminations. Figure 3.7 provides additional examples of these photographs. Several of the spheres were somewhat translucent. Images were acquired in 24-bit RGB format using a Nikon D1 digital camera and then converted to gray-scale for further processing. Using standard calibration methods, we found that the pixel values are a sigmoidal function of luminance.[7] We worked with the original pixel values rather than the luminance values in order to show that our techniques apply to uncalibrated photographs.[8]

We normalized for overall strength of illumination as discussed in Section 3.4. For rendered images, we scaled the intensities in each illumination map such that a matte white surface perpendicular to the view direction had a fixed luminance. For photographs, we scaled intensities of images under each illumination such that the mean value of the white matte surface was constant across illuminations.

[7] In other words, pixel value x can be expressed in terms of luminance y as $x = d\frac{y^g}{y^g+k}$, where d, g, and k are constants.

[8] Calibration of the photometric response of the camera may be necessary in order to train a classifier using photographs from one camera and apply it to photographs from a second camera with a significantly different response.



(a) (b) (c) (d)

(e) (f) (g)

Figure 6.5. Chrome spheres photographed under each of seven illuminations.

Table 6.1 lists the accuracy of classifiers based on different feature sets, as measured by a cross-validation test for each image set. We used the images under a particular illumination to test a classifier trained on images under other illuminations. Each image therefore serves as a test image once. The numbers in the table represent the percentage of images classified correctly. For images sets (1) and (4), with a total of 9 and 7 illuminations respectively, we trained each classifier using the images under all but one illumination. For image sets (2) and (3), we split the 95 illuminations into 5 groups of 19 and used the images in each group to test a classifier trained on the other 4 groups.

Each row of the table lists the performance of SVM classifiers based on different image statistics. As a baseline, row (a) indicates the expected performance of a random classifier, and row (b) indicates the performance of a classifier based only on the image mean. Row (c) lists the performance of classifiers based on the following "full" set of 36 image statistics, computed using the four finest levels of the QMF pyramid for a 256x128 unwrapped image:[9]

- The mean and 10th percentile of pixel values (2 statistics)

[9]Including statistics based on the diagonally-oriented coefficients has little impact on classifier performance.

- The variance of horizontally- and vertically-oriented wavelet coefficients at each of four scales (8 statistics)
- The kurtosis of horizontally- and vertically-oriented wavelet coefficients at each of four scales (8 statistics)
- Ratios of variances at successive scales (σ_1^2/σ_2^2,σ_2^2/σ_3^2, and σ_3^2/σ_4^2) for horizontally- and vertically-oriented wavelet coefficients (6 statistics)
- The statistics r_{12}, r_{23}, r_{34}, c_{12}, c_{23}, and c_{34} for horizontally- and vertically-oriented wavelet coefficients (12 statistics)

The classifiers for image sets (2) and (3), which are based on a large number of training images, achieve high classification performance (around 99%) using this large set of features. For image sets (1) and (4), however, only 8 and 6 examples respectively of each class are used to train each classifier. In these cases, classifier performance is significantly lower.

In principle, providing a classifier with fewer features limits the amount of information available for classification. Given only a small number of training examples, however, classification accuracy typically increases with the elimination of uninformative features. Figure 6.6 illustrates this phenomenon with a toy example. Figure 6.6(a) shows the training data points of Figure 6.1. In Figure 6.6(b), the uninformative feature x_2 has been eliminated. The classification problem is now significantly easier, because one need only choose a boundary point on the line rather than a boundary curve in the plane. Figure 6.6(c) illustrates a classification problem based on two identical features. This classification problem is no more difficult than that of Figure 6.6(b). Elimination of redundant features does not necessarily improve classifier performance, but elimination of uninformative features with high variance tends to improve performance.

Rows (d), (e), and (f) of Table 6.1 show examples of reduced feature sets that lead to improved classifier performance for these particular sets of reflectances. In row (d), we have eliminated the ten statistics based on wavelet coefficients at the fourth-finest scale. These statistics are useful for differentiating between highly blurred specular components, but most of the specular reflectances in these four image sets are relatively sharp. In row (e), we have further eliminated the cross-scale statistics r_{jk} and c_{jk}. These statistics are useful for classification of images with high frequency components due either to noise or three-dimensional surface texture. The image sets of Table 6.1, however, exhibit little noise or surface texture. The classifiers in line (e) therefore perform as well as or better than those in line (c). The painted clay surfaces discussed in Section 6.2.2, on the other hand, have blurrier specular components and more evident surface texture. For these images, eliminating the cross-scale statistics or the statistics based on coarser-scale wavelet coefficients leads to a reduction in classifier performance.

Row (f) shows classifier performance for a set of six hand-selected features. This set consists of the mean and tenth percentile of the pixel intensities, the variance of wavelet coefficients in the finest and second-finest radially-oriented subbands, the ratio

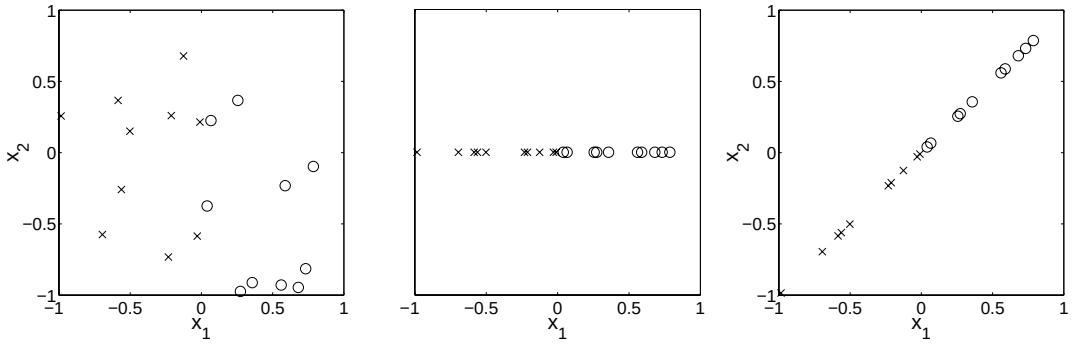


Figure 6.6. Elimination of uninformative features improves classifier performance. (a) The data points of Figure 6.1. (b) The same data set, with the uninformative feature x_2 removed. (c) The data set of (b), with $x_2 = x_1$. The classification problems of (b) and (c) are easier than that of (a).

of variances of the second- and third-finest radially-oriented subbands, and the kurtosis of the second-finest radially-oriented subband. For image sets (1) and (4), where the classifiers are trained on very few samples, this reduction in features leads to a significant improvement in performance. Classifiers based on this set of statistics made no cross-validation errors for the set of 6 reflectances under 9 illuminations, and misclassified only 4 of the 63 photographs.

Row (g) of Table 6.1 represents the performance of a classifier based on the same statistics as row (e), but computed on the observed image without the "unwrapping" transformation of Figure 6.2. For three of the four data sets, the unwrapping transformation improves classifier performance.

Table 6.2 compares the performance of support vector machine, nearest neighbor, and 5-nearest neighbor classifiers for classification of the 11 synthetic reflectances of Figure 6.3. The table shows cross-validation results for classifiers trained using different numbers of sample images and different numbers of features. The SVM matches or exceeds the performance of the alternative classifiers in each case, but in most cases, the NN and k-NN classifiers are also quite accurate.

■ 6.1.3 Misclassifications

One can gain further insight into the performance of these reflectance classifiers by examining common misclassifications. Not surprisingly, certain reflectances are confused more often than others. For example, a classifier based on the full set of 36 statistics (row (c) of Table 6.1) incorrectly classifies 11 of 1045 images in image set (3). Eight of these misclassifications involve confusion between reflectances (d), (f), and (i) of Figure 6.3. These three reflectances all share a sharp specular component ($\alpha = 0$).

Classification technique	6 features		36 features	
	10 illums	95 illums	10 illums	95 illums
SVM	92.7	98.0	90.0	98.9
nearest neighbor	92.7	96.0	86.4	95.3
5-nearest neighbor	90.0	96.2	81.8	93.9

Table 6.2. Comparison of performance of support vector machine, nearest neighbor, and k-nearest neighbor (k = 5) classifiers for the 11 reflectances of Figure 6.3. The first two columns shows cross-validation performance using the set of 6 features from row (f) of Table 6.1. The latter two columns show cross-validation performance for the set of 36 features from row (c) of Table 6.1. For each feature set, we performed cross-validation using images rendered under all 95 Teller illuminations, with each classifier trained on 76 illuminations and tested on the remaining 19. We also performed cross-validation using images rendered under 10 randomly selected Teller illuminations, with each classifier trained on 9 illuminations and tested on the remaining one.

They differ in the relative energies of the specular and diffuse components, as well as in total reflectance. Some of the illumination maps contain substantial amounts of light incident from almost all directions; under these conditions, the relative strengths of the specular and diffuse reflectance are difficult to distinguish. In the remaining three misclassifications, reflectance (k) is misclassified as (j) or (h). These three reflectances are identical to (d), (f), and (i), except for a difference in the specular blur α.

A small fraction of illuminations often accounts for a majority of the misclassifications. These illuminations are generally "atypical" in that their statistics differ significantly from the majority of the training set. This is precisely the behavior one would expect from a classifier based on training examples, given that the problem of reflectance estimation under arbitrary illumination is underconstrained. For example, two of the four misclassifications for the set of photographs in row (f) of Table 6.1 corresponded to the one illumination created by the photographer specifically for the purpose of collecting these photographs (Figure 6.5(e)). These images were photographed directly under a desk lamp in an otherwise dark room. The first row of Figure 6.7 shows the more egregious of the misclassifications, where a white matte surface was misclassified as metallic because its appearance was so different from that of the white matte surface under other illuminations. The upper portion of the misclassified image is saturated to the maximum pixel value, leading to a distinct edge at the boundary of the region directly illuminated by the lamp. The second row of the figure shows an example of a misclassification under a different illumination.

■ 6.2 Geometry

The previous section considered reflectance classification for spherical surfaces. This section discusses methods for applying a reflectance classifier to surfaces of arbitrary known geometry. Figure 6.8 illustrates an example of a classification problem where the classifier is trained on surfaces of one geometry, and then applied to surfaces of several geometries. We also consider the effect of surface geometry on the statistics of the

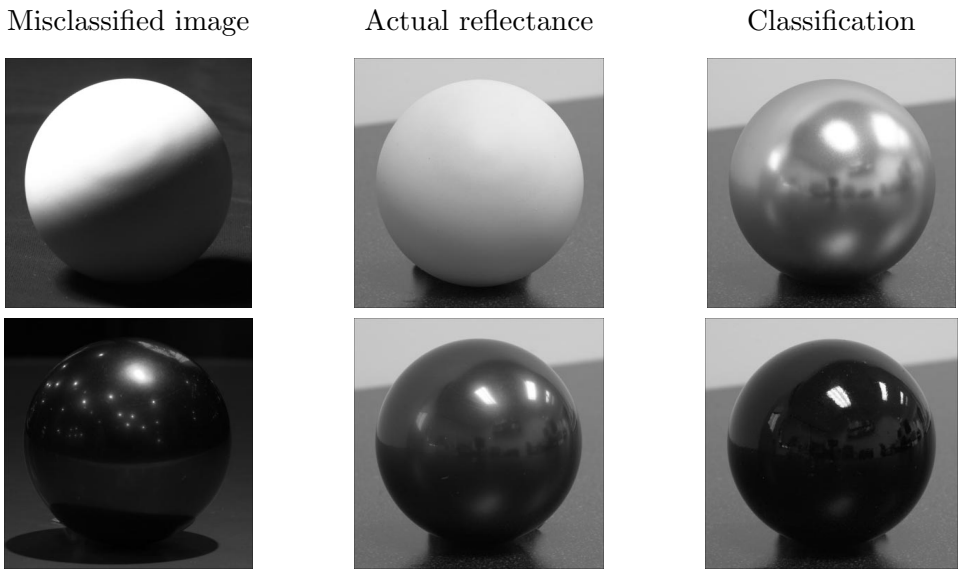


Figure 6.7. Examples of classification errors for the set of photographs. The left-hand column shows the misclassified image. The middle column shows a sphere with the same reflectance as the sphere at left, while the right-hand column shows a sphere with the reflectance assigned by the classifier to the image at left. The images in the middle and right-hand columns are under illumination (g) of Figure 6.5. The misclassified image in the first row is under illumination (e), while that in the second row is under illumination (d).

observed image of a surface. If the image statistics proposed for reflectance estimation in Chapter 5 were invariant to changes in surface geometry, one could effectively ignore surface geometry in classifying reflectance. Unfortunately, this is not generally the case. Before proceeding with further analysis, we provide two simple examples to illustrate this point.

Figure 6.9 shows three spheres rendered under a single photographically-acquired illumination map. Spheres (a) and (b) share identical reflectance properties, but sphere (a) is three times the size of sphere (b). The image of sphere (a) is therefore identical to that of sphere (b), but three times larger. Image (b) possesses spectral energy at higher spatial frequencies than image (a). This leads to differences in the variances of wavelet coefficients, as well as other image statistics, at particular spatial scales. In fact, the power spectral characteristics of sphere (a) are similar to those of sphere (c), a smaller sphere whose reflectance has a larger specular blur parameter.

Given images of differently sized spheres, one can account for the effect of geometry on image statistics by simply scaling the images appropriately. Eliminating differences in image statistics due to more general differences in geometry is less trivial. Figure 6.10 shows two surfaces of different shapes, made of the same material and photographed under the same illumination. Although both images exhibit specular reflections of the same bright light sources, the reflections are distorted in image (b) relative to

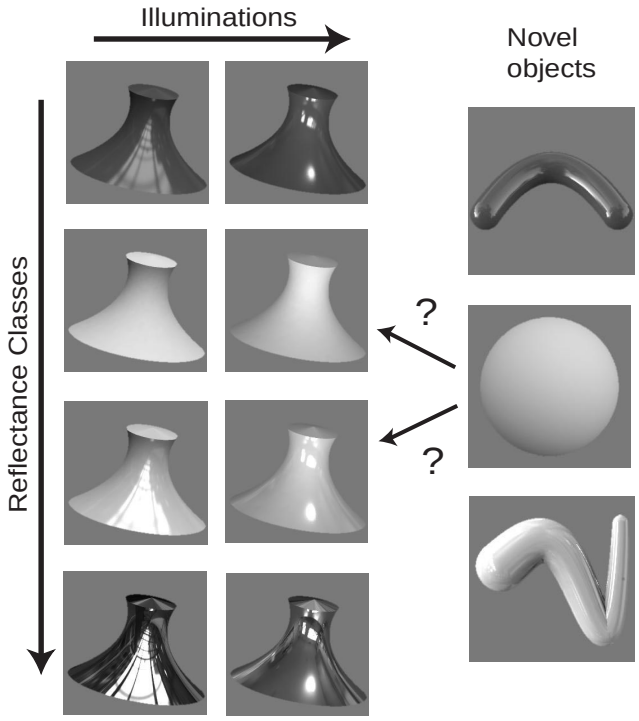


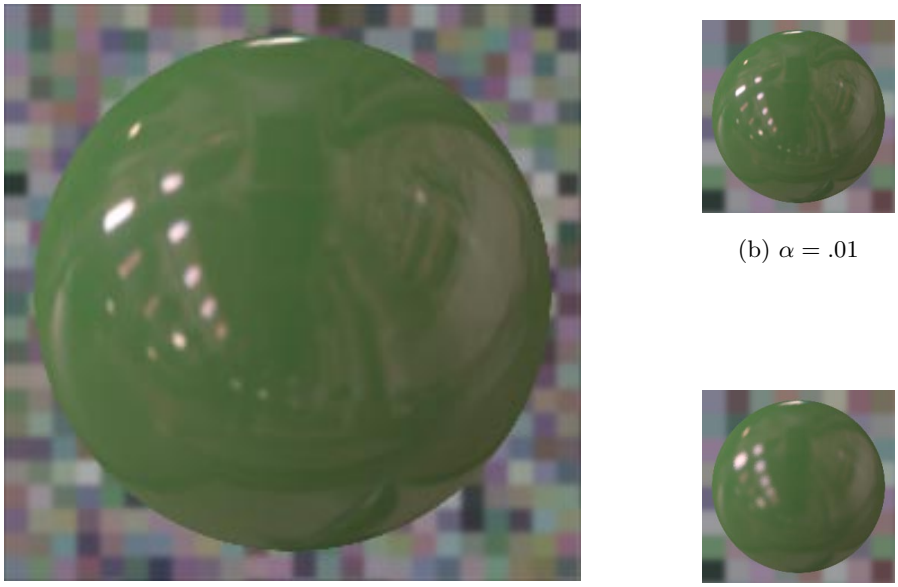
Figure 6.8. Classification for surfaces of arbitrary known geometry. We wish to train a reflectance classifier using images of surfaces under several real-world illuminations, and then apply that classifier to images of surfaces of different known geometries under novel, unknown illuminations.

their appearance in image (a). In image (b), distinctive image structures tend to be elongated in a direction parallel to the contours separating the metallic surface from the background. The regions highlighted by rectangles, for example, all correspond to reflections of the same fluorescent lamp. These distortions affect the power spectrum and wavelet coefficient statistics in each region of the image.

Section 6.2.1 proposes a method for reflectance classification that applies to surfaces of arbitrary known geometry. Section 6.2.2 considers the robustness of this method to differences between assumed and actual geometry, including an analysis of the effects of surface geometry on image statistics.

■ 6.2.1 Handling Known Differences in Geometry: Remapping to the Equivalent Sphere

Our approach to handling differences in surface geometry depends on the observation that, for a surface of given reflectance under given illumination, the intensity of an image point depends primarily on the orientation of the corresponding surface patch. More precisely, when a distant viewer observes a convex surface under distant illumination, the radiance reflected toward the viewer depends on the orientation of a surface



(a) $\alpha = .01$

(b) $\alpha = .01$

(c) $\alpha = .03$

Figure 6.9. Three spheres rendered under the same illumination. Sphere (a) is three times the size of spheres (b) and (c). Spheres (a) and (b) have identical reflectances, described by the Ward model with $\alpha = .01$. The reflectance of sphere (c) has a blurrier specular component, with $\alpha = .03$. The edges of the specularities in image (a) are blurrier than those in image (b) but similar to those in image (c).

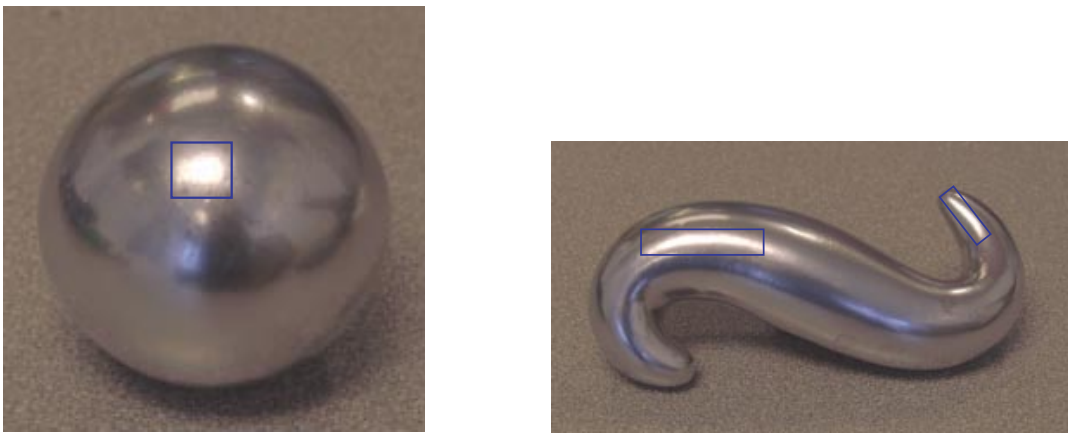


Figure 6.10. Two surfaces with the same reflectance photographed under the same illumination. The rectangles indicate specular reflections of the same lamp. The different appearances of these reflections illustrate the fact that the image structure depends on surface geometry.

patch, not on its position (Section 3.1.3). Under these conditions, one can associate each possible surface orientation with a unique image intensity. This mapping from orientation to intensity underlies shape-from-shading techniques [45].

One can specify the relationship between surface orientation and image intensity for a surface of a particular reflectance under a particular illumination using a hypothetical sphere with the same reflectance under the same illumination. We will refer to this sphere as the *Gaussian sphere*.[10] This representation is identical to the "Lit Sphere" proposed by Sloan *et al.* [103] for use in artistic rendering.

Given an image of a surface with arbitrary known geometry, one can partially reconstruct the Gaussian sphere for that combination of illumination and reflectance. One maps each point on the observed image to the point on the sphere with the same surface normal. Figure 6.11 shows several examples of such reconstructions. For nonconvex surfaces such as those shown in the figure, multiple image points may map to the same point on the sphere. The intensities of these image points may differ due to variations in illumination. In particular, a non-convex surface may give rise to interreflections or cast shadows on itself. In the examples of Figure 6.11, these phenomena have only a minor effect on the reconstructions.

Because the Gaussian sphere is a geometry-invariant image representation, we can build reflectance classifiers for surfaces of arbitrary known geometry by using reconstructed spheres in both the training and test phases. This approach presents two complications. First, the data is naturally defined on a sphere rather than on a plane, because image intensity depends on surface orientation. Planar wavelet transforms are not directly applicable in this domain. One might employ the spherical wavelet framework introduced in Section 4.6. However, this transform is based on radially symmetric basis functions that are less sensitive to the oriented structures in natural images and illumination maps than the oriented filters of Chapter 5. We elected instead to simply project the sphere onto a plane. An image of the sphere as observed by a distant viewer is simply an orthographic projection of the sphere. We "unwrap" this image about its center, as discussed in Section 6.1.2, before computing the wavelet transform.[11] The unwrapped orthographic projection is equivalent to a cylindrical projection of the Gaussian sphere. Appendix C further discusses the distortions introduced by projection of the sphere onto the plane. Section 7.2.2 proposes an alternative approach to dealing with differences in geometry that avoids such distortion.

The second complication stems from the sampling of surface normals in the original image. Regions of high curvature in the observed surface lead to sparse sampling of surface normals. Moreover, the observed surface may simply lack certain normal

[10] Although we use the term *Gaussian sphere*, we concern ourselves only with the hemisphere visible from the view direction.

[11] When "unwrapping" the images of spheres in Section 6.1, we used an annulus that omitted regions of the image near the center and near the boundaries. In the examples of the present section, we do not omit these regions. If the observed surface lacks densely sampled surface normals in many directions, ignoring these regions may lead to a significant reduction in the amount of image data remaining for reflectance estimation, causing degradation of classifier performance.

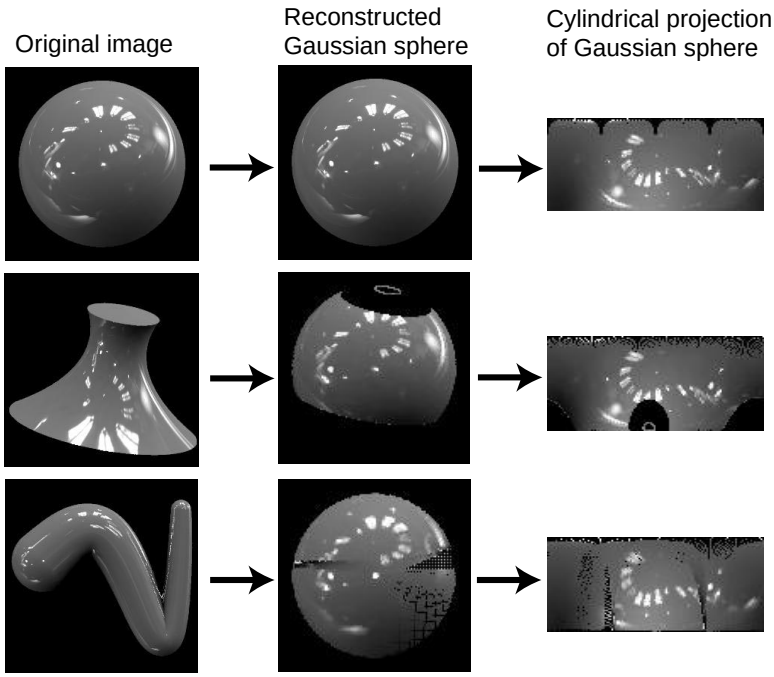


Figure 6.11. Images of surfaces of three geometries, each mapped to a Gaussian sphere of the same reflectance and illumination, then "unwrapped" to a cylindrical projection of the sphere. All three original surfaces have the same reflectance and were rendered under the same illumination. Black regions within the Gaussian spheres and cylindrical projections indicate missing data, due to sparse sampling of surface normals in the observed geometries.

directions.[12] The reconstructions of Figure 6.11, which were constructed by forward mapping from the observed image to the Gaussian sphere and its cylindrical projection, use black pixels to indicate normal directions that are not sampled in the original image. One might fill in these regions by using backward sampling or by performing interpolation on the reconstructed image, but one cannot accurately estimate fine-scale wavelet coefficients in areas of the reconstructed sphere where the normal directions are sampled too sparsely.

We estimate wavelet coefficients from the incompletely sampled cylindrical projection using an instantiation of normalized differential convolution [121]. For computational convenience, we approximate the outputs of derivative of Gaussian filters rather than the quadrature mirror filters used in Section 6.1.2. At each point in

[12]Strictly speaking, any surface that is topologically equivalent to a sphere has all possible surface normals. For non-differentiable surfaces such as the pedestal-shaped surface of Figure 6.11, an entire region of the Gaussian sphere may map to a curve or point in the image, so that hardly any information about the intensity associated with surface normals of that region is available. One may also wish to estimate reflectance from an image of a portion of a surface, either because an image of the rest of the surface is not available or because the rest of the surface has different reflectance properties. In this case, information about certain regions of the Gaussian sphere may be entirely absent.

		Test geometry		
	Training geometry	Spheres	Worms	Pedestals
(a)	spheres	99.5	98.2	93.5
(b)	worms	97.5	98.6	90.7
(c)	pedestals	98.6	96.5	96.7
(d)	spheres (geometry ignored)	98.9	74.0	80.5
(e)	spheres (with fine-scale statistics)	98.9	80.2	87.2

Table 6.3. Cross-validation performance (in % accuracy) for SVM classifiers trained and tested on different geometries.

the reconstructed image, we find the linear combination of a constant function, a horizontally-oriented ramp, and a vertically-oriented ramp that best approximate, in a weighted mean-squared sense, the observed samples over a Gaussian window surrounding that point. The coefficients of the two ramp functions represent our estimate of the derivative-of-Gaussian filter outputs at that point.[13] The size of the Gaussian window varies by a factor of two from scale to scale. Following Westelius [120], we compute a confidence measure for the filter output estimates based on the amount of data present within a particular Gaussian window. When computing statistics based on filter outputs, we use only those outputs whose associated confidence exceeds a threshold value.

Table 6.3 shows the performance of reflectance classifiers trained and tested on images of surfaces of three different geometries. We rendered images of each surface with each of six reflectances under each of the 95 Teller illumination maps. Figure 6.12 shows images of each surface geometry and reflectance under the same illumination.

Row (a) of Table 6.3 list the performance of a classifier trained on spheres and tested on all three geometries. This classifier uses the features described in Section 6.1.2, but computed using difference of Gaussian filters at two scales, corresponding roughly to the second- and third- finest scales of the QMF pyramid.[14] As in Section 6.1.2, we measured performance using a cross-validation procedure. We split the illuminations into five groups, and used images rendered under each group of illuminations to test a classifier trained on the other four groups. The second and third rows represent cross-validation performance values for classifiers using the same feature set, trained on the worm-shaped and pedestal-shaped surfaces, respectively.

Each of the classifiers based on reconstructed Gaussian spheres achieves over 90%

[13] A Matlab implementation of such a decomposition, by Gunnar Farnebäck, is available at *http://www.isy.liu.se/∼gf/software/*. The results presented in this thesis relied on our own implementation of these algorithms.

[14] More specifically, the feature set included the mean and 10th percentile of pixel values, the variance and kurtosis of horizontally- and vertically-oriented wavelet coefficients at the second- and third-finest scales, and ratios of variances at the second- and third-finest scales for horizontally- and vertically-oriented wavelet coefficients.

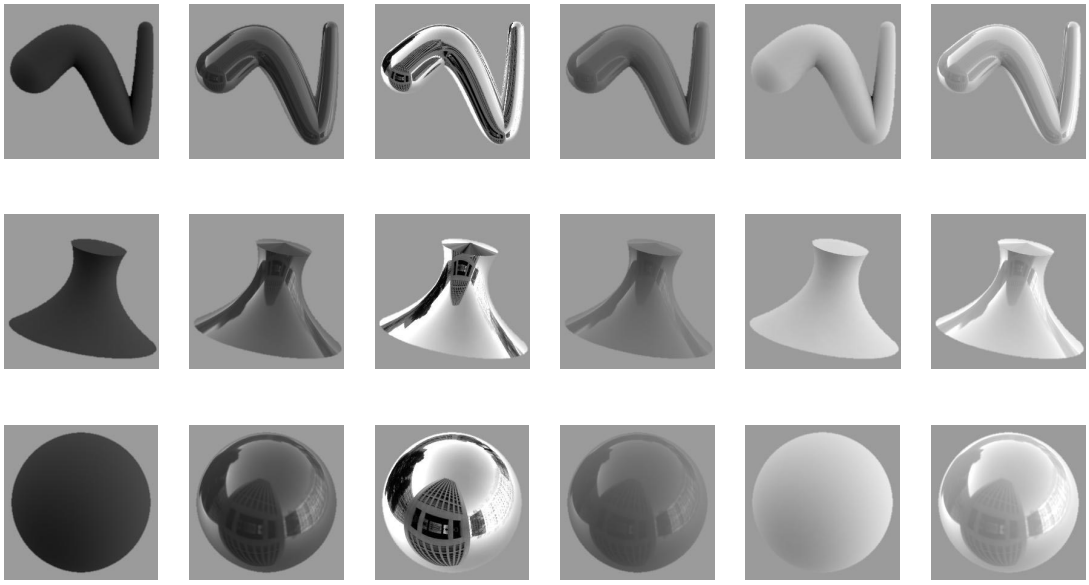


Figure 6.12. Synthetic images of surfaces of three different geometries, each rendered with 6 reflectances under one of Teller's illumination maps. The set of images of spheres is identical to image set (2) of Table 6.1. Ward model parameters corresponding to each reflectance are specified in Figure 5.11.

accuracy on each of the test sets, despite the lack of convexity of the worm and pedestal shapes. By contrast, row (d) of Table 6.3 shows the performance of a classifier trained and tested using statistics computed directly on the observed image. In computing these statistics, we ignored surface geometry, apart from the specification of the region of the image occupied by the surface. This classifier performs well when trained and tested on surfaces of the same geometry, but poorly when tested on surfaces of other geometries.

Although the Gaussian sphere itself is independent of geometry, each of the classifiers based on reconstructed Gaussian spheres achieves highest performance when the test set geometry is identical to the training set geometry. This stems primarily from the fact that one can compute statistics of the Gaussian sphere with different degrees of confidence for different surface geometries. For example, a highly curved surface leads to sparse sampling of surface normals on the sphere, such that fine-scale wavelet coefficients cannot be estimated accurately. A classifier trained on a particular geometry tends to weight more heavily the features that prove reliable for that geometry.

Row (e) of Table 6.3 shows the performance of a classifier similar to that of row (a), but trained using an expanded set of features that includes image statistics computed on wavelet coefficients at a finer scale. Given the resolution of the observed images, these wavelet coefficients can be estimated reliably from images of spheres, but not from images of geometries such as the worm-shaped surfaces. The classifier therefore proves

significantly less accurate when tested on a surface of this geometry than a classifier that does not use the fine-scale statistics at all (row (a)). The design of a reflectance classifier that takes the confidence of feature estimates into account represents an area for future research (Section 7.2.1).

■ 6.2.2 Robustness to Incorrect Assumed Geometry

Most previous approaches to reflectance estimation have relied on some form of inverse rendering [9,73,86,123,124], where one optimizes reflectance parameters by minimizing the difference between a synthesized image and the observed image. Such an approach requires an accurate description of surface geometry. Incorrect assumed geometry will prevent the synthesized and reflected images from matching, even when the reflectance is correct.

Reflectance estimation based on image statistics, on the other hand, is robust to some errors in the assumed geometry. Figure 6.13 shows an example where a sphere and an elongated ellipsoid of identical reflectance have been rendered under identical illumination. Although both surfaces have identical circular contours when viewed from the front, the two images differ significantly at the pixel level. In particular, the specular highlights appear at different locations. We trained a classifier using images of spheres of six reflectances (Figure 5.11). We then tested this classifier on images of ellipsoids such as that shown in Figure 6.13. When we correctly specified the geometry of the ellipsoid, the classifier correctly labelled 53 of 54 test examples. When we incorrectly specified the geometry as spherical, the classifier still labelled 51 of 54 images correctly. This robustness to a change in geometry is hardly surprising. The image statistics we use for reflectance classification are similar for the two images of Figure 6.13. In fact, a human observer recognizes the similarity in reflectance of the two objects more easily than the difference in geometry.

Robustness to differences between actual and assumed geometry is desirable in a reflectance classifier because it lightens the burden of geometry recovery and represents a step toward joint estimation of geometry and reflectance. We therefore wish to understand how unknown deviations in geometry will affect the performance of our classification techniques. To this end, we first address the more basic question of how geometry affects the image of a reflective surface and the statistics of that image.

Effect of geometry on local image statistics

The image of a convex surface observed by a distant viewer is a distorted version of the Gaussian sphere. In the neighborhood surrounding a differentiable point on the surface, the image is simply a warped version of a corresponding neighborhood on the Gaussian sphere. The map from the Gaussian sphere to the image depends on the local geometry of the surface — in particular, on its orientation and curvature. The distortions introduced by this transformation affect wavelet coefficients computed at that point in the image.

To quantify these relationships, we first consider a two-dimensional world, in which

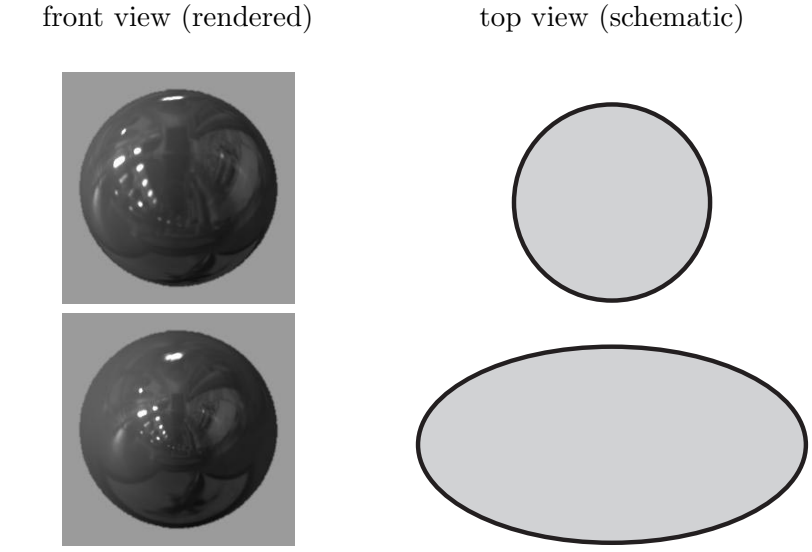


Figure 6.13. Images of a sphere and an ellipsoid with identical reflectance, rendered under the same illumination. The major axis of the ellipsoid, directed toward the viewer, is twice the length of its other two axes. Schematic top views of these geometries are shown at right.

surfaces are simply curves and images are one-dimensional (Figure 6.14). We introduce a coordinate system (x, z), where the z-axis is in the direction of a distant viewer and the x-axis is orthogonal to the viewer. One can parameterize the observed image by x. The Gaussian sphere in this two-dimensional world is a function of surface normal angle γ. Suppose that at a point (x_0, z_0) on the surface, the surface normal angle is γ_0. The image of the surface near x_0 is a contracted or dilated version of the Gaussian sphere in the neighborhood of γ_0. The amount of contraction depends on the rate at which the surface normal angle changes as a function of image position x. Mathematically, this rate is given by $\gamma'(x_0) = \frac{d\gamma}{dx}\big|_{x=x_0}$. If $\gamma'(x_0) = 1$, then the image near x_0 matches an orthogonal projection of the Gaussian sphere onto its local tangent at γ_0. If $\gamma'(x_0) > 1$, the image near x_0 is a contracted version of the corresponding region on the Gaussian sphere; if $\gamma'(x_0) < 1$, then the image near x_0 is a dilated version of that region of the Gaussian sphere.

The rate $\frac{d\gamma}{dx}$ depends on the local orientation and curvature of the surface. One can express the surface normal angle as $\gamma = \frac{\pi}{2} - \arctan\left(\frac{dz}{dx}\right)$, so

$$\frac{d\gamma}{dx} = \frac{\frac{d^2z}{dx^2}}{1 + \left(\frac{dz}{dx}\right)^2}. \tag{6.11}$$

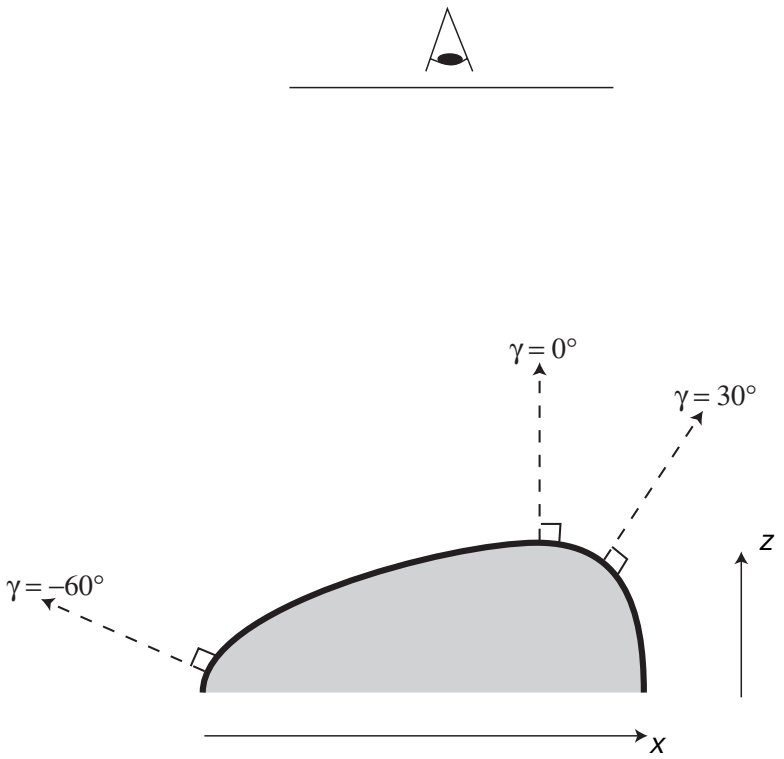


Figure 6.14. Sample surface geometry in a two-dimensional world. The viewer is distant and the view direction is given by $\theta_r = 0$. We parameterize the image by x and the Gaussian sphere by γ.

The curvature of the surface [105], as a function of image position x, is given by

$$\kappa(x) = \frac{\left|\frac{d^2z}{dx^2}\right|}{\left|1 + \left(\frac{dz}{dx}\right)^2\right|^{3/2}}. \tag{6.12}$$

We can therefore rewrite Equation (6.11) as

$$\left|\frac{d\gamma}{dx}\right| = \kappa(x)\sqrt{1 + \left(\frac{dz}{dx}\right)^2}. \tag{6.13}$$

The amount by which the Gaussian sphere is contracted in the mapping to the observed image increases as surface curvature increases or as the angle between the view direction and the surface normal direction increases.

Appendix C describes similar analysis for a two-dimensional image of a surface in a three-dimensional world. Although the mathematics are more involved, the mapping from the Gaussian sphere to the image can still be described locally as a linear warping that depends on the first and second derivatives, or orientation and curvature, of the

surface. This analysis describes, for example, the distortions noticeable in the images of Figures 6.9 and 6.10. The large sphere of Figure 6.9(a) has lower curvature than the smaller sphere of Figure 6.9(b), so each region of the image of the large sphere is dilated relative to the corresponding region in the image of the small sphere. In regions near the contours separating the surfaces of Figure 6.10 from the background, surface orientation changes more quickly as one moves toward the contour in the image than as one moves parallel to it. The reflection of the Gaussian sphere is therefore elongated along the direction of the contour and compressed in the perpendicular direction.

If the assumed surface geometry is incorrect, the image will be warped incorrectly when mapped to the Gaussian sphere. Locally, the difference in the mapping depends on the change in local surface orientation and curvature. Applying a local linear distortion to part of the Gaussian sphere affects the local wavelet coefficients. Each wavelet coefficient represents the output of a bandpass filter. Distorting the image locally also distorts its local frequency spectrum. Small rotations and scaling transformations will typically have only a limited effect on the distributions of filter outputs. As long as the differences in local orientation and curvature between the assumed and actual surface geometry are small, one would expect a classifier based on these distributions to perform accurately. Such is the case for the sphere and the ellipsoid of Figure 6.13. On the other hand, the wavelet coefficient distributions of the reconstructed Gaussian sphere will change substantially if one uses an assumed surface geometry that differs substantially in local orientation and curvature from the actual geometry. This will be the case, for example, if the actual surface geometry contains sharp corners while the assumed geometry is smooth.

Throughout this chapter, we have assumed a distant viewer. If the viewer is not distant relative to the extent of the surface, the view direction will vary significantly from one point on the surface to another. Under these conditions, the radiance reflected from a surface patch toward the viewer depends on the position of that surface patch as well as on its orientation.

If the surface geometry is known and convex, one could take this effect into account by mapping each point on the observed image to a point on the sphere that would reflect light from the same direction if both surfaces were mirrors. For a flat surface observed by a nearby viewer with a relatively narrow view angle, the image information will map to a small region of the Gaussian sphere that can be approximated by a plane. When examining flat surfaces under such a view geometry in Chapter 5, we therefore assumed that image statistics were stationary across the image. For a nearby viewer, the image of the reconstructed sphere will not be identical to the image of an actual sphere of the same reflectance under the same illumination. The diffuse reflection of a homogeneous material, for example, is constant across a flat surface under distant illumination. Under the same illumination and view conditions, the diffuse reflection varies across the surface of a sphere. However, we have found that the statistics of spheres reconstructed from convex surfaces are not particularly sensitive to view distance. When trained on images of spheres rendered from a distant viewpoint, for example, classifiers such as those in

Section 6.1.2 accurately classify spheres rendered from a nearby viewpoint.

■ 6.2.3 Misalignment Artifacts for Non-Convex Surfaces

For a non-convex surface, even minor deviations from the assumed surface geometry, the distant viewer assumption, or the distant illumination assumption can introduce severe artifacts to the Gaussian sphere reconstruction. Several points on a non-convex surface may share the same surface orientation and therefore map to the same point on the Gaussian sphere. Misalignment of the image information from different parts of the image on the Gaussian sphere can lead to severe artifacts in the reconstruction.

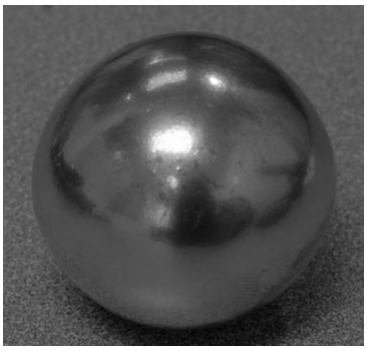
Figure 6.15 illustrates this point using an example. We attempted to reconstruct a Gaussian sphere based on the image of the snake-shaped object in Figure 6.15(a). A photograph of a sphere made of the same material under the same illumination conditions, shown in Figure 6.15(b), provides a "ground truth" for this reconstruction.

As in many vision applications, we lack an accurate geometric model for the snake of Figure 6.15(a). However, we can obtain a reasonable estimate of the geometry using a technique known as *inflation* [48,122]. This technique involves segmenting the snake-shaped object from the background and then treating the contour separating the object and background as the boundary of an elastic balloon. We performed the segmentation by hand and then applied an inflation algorithm due to Tappen [106], which assumes that all surface points on the boundary contour lie in a plane perpendicular to the viewer. Panels (c) and (d) of Figure 6.15 illustrate the estimated geometry in terms of its surface normals. The image of the snake in panel (c) is shaded such that the gray level of a pixel is proportional to the component of the unit normal vector in the direction of the observer. The image in panel (d) is shaded according to the component of the unit normal vector in the direction toward the top of the page. Although not exact, the estimated geometry is reasonable.

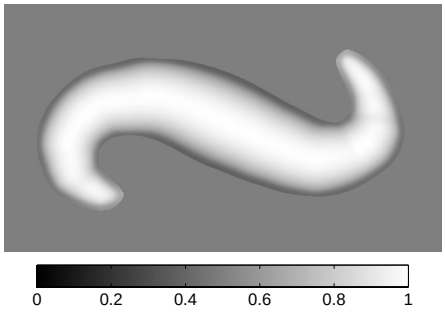
We reconstructed a Gaussian sphere from the image of Figure 6.15(a) using this estimated geometry. Figure 6.15(e) shows the result of forward mapping pixels of the observed image onto pixels of the reconstructed sphere. In panel (f), the missing pixels of panel (e) have been filled in by interpolation on the image. The reconstruction is reasonable in the sense that many of the reflected structures visible in panel (b) are also discernible in (f). The reconstruction contains edge-like artifacts absent from the photographed sphere, however, particularly in the upper half of the image. These coincide with boundaries between pixels mapped from different parts of the original image. The errors in estimated geometry, combined with variations in illumination from one portion of the surface to another, lead to spurious edges in the reconstructed image. Unfortunately, such artifacts lead to large spurious wavelet coefficients and may heavily influence the statistics that we use for reflectance classification.

Even if the assumed surface geometry is correct, illumination from nearby surfaces may lead to artifacts in the reconstruction. The surfaces in Figure 6.15, for example, were photographed on a carpet. The reflections of this carpet vary between different points on the snake-shaped surface with the same orientation.

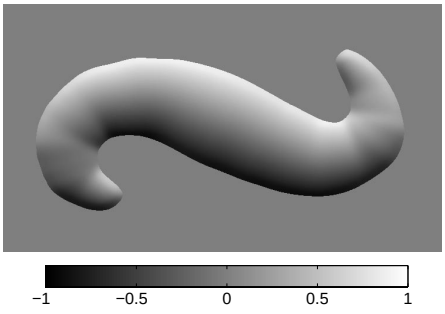
(a) Original photograph



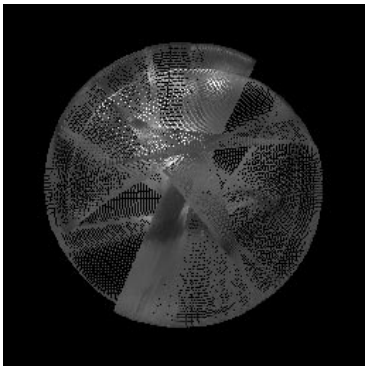
(b) A sphere of the same reflectance, photographed under the same illumination



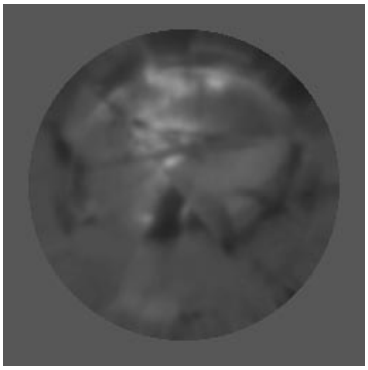
(c) Component of estimated surface normal toward viewer



(d) Component of estimated surface normal toward top of page



(e) Pixels of snake photograph, mapped to Gaussian sphere



(f) Interpolated Gaussian sphere

Figure 6.15. Reconstruction of a Gaussian sphere for a non-convex shape, using approximate geometry. See text for details.

In an experiment using real photographs and approximate geometries, we found a large difference in classifier performance for convex and non-convex geometries. We used objects fashioned of clay and painted to produce nine different reflectances. For each reflectance, we used one convex, approximately spherical object and one non-convex, snake- or crescent-shaped object. We photographed the objects using a Nikon D1 camera. In this case, pixel values were linear in luminance.[15] The convex objects were photographed under eight illuminations, while the non-convex objects were photographed under four of these illuminations. Figure 6.16 shows convex objects of all nine reflectances under one illumination. Figure 6.17 shows the non-convex shapes with the corresponding reflectances under the same illumination. Several of the reflectances are quite similar, leading to a difficult classification problem.

We reconstructed Gaussian spheres from each image, using geometries approximated from hand-segmented contours. We used circular contours for the convex objects, even though the shapes were crafted by hand and therefore were not perfectly spherical. We then trained reflectance classifiers using a full set of image statistics, as defined in Section 6.1.2, computed using difference of Gaussian filters at four scales. In a cross-validation test, the resulting classifiers correctly classified 76% of the images of the convex surfaces, including 79% of the images photographed under the four illuminations used for the non-convex surfaces. When applied to images of the non-convex surfaces, the same classifiers achieved only 43% accuracy. A classifier that ignores geometry entirely, apart from from the specification of the region of the image occupied by the surface, achieved higher performance (49%) when trained on the convex shapes and tested on the non-convex shapes.

Section 7.2.2 proposes an alternative approach to dealing with surfaces of different geometries that avoids explicit reconstruction of the Gaussian sphere. By computing wavelet coefficients on appropriately distorted local neighborhoods of the observed image, this alternative approach circumvents the artifacts discussed in this section for non-convex objects with inaccurate geometry estimates.

■ 6.3 Summary and Discussion

By taking advantage of the consistent relationships between image statistics and surface reflectance properties, one can reliably classify the reflectance of surfaces viewed under unknown real-world illumination. In this chapter, we performed such classification using the image statistics discussed in Chapter 5 — statistics summarizing the distribution of pixel intensities, the distributions of wavelet coefficients at various scales and orientations, and the joint distributions of co-located wavelet coefficients at different scales. We found that one-versus-all support vector machine classifiers lead to high accuracy in reflectance classification, even given only a few examples of each reflectance in a multi-way classification task. We also demonstrated the successful application of

[15] We used the camera's native "raw" format, and converted the images to 48-bit TIFFs. We verified that pixel values were linear in luminance using standard calibration techniques.

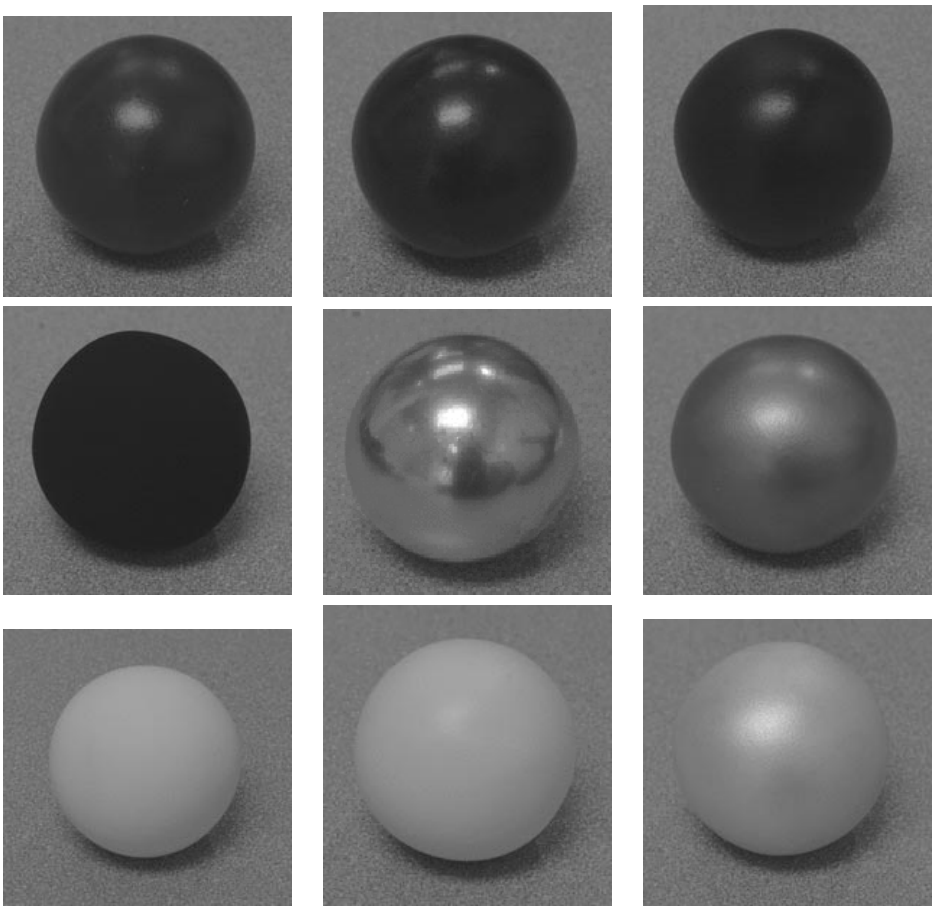


Figure 6.16. Painted clay shapes with approximately spherical geometries, all photographed under the same illumination. Each is painted to have a different reflectance.

classifiers constructed using these techniques to surfaces of various known geometries.

In an informal test, the author of this thesis attempted to classify the reflectances of the convex surfaces from the two sets of monochrome photographs discussed in this chapter (Figures 6.4 and 6.16 include some of the more easily classified images in these two sets). In both cases, the author was unable to match the performance of the reflectance classifiers reported in this chapter.

While our classification techniques are robust to certain errors in assumed geometry for convex surfaces, their performance degrades substantially with deviations from assumed geometry for non-convex surfaces. The human visual system, on the other hand, recognizes reflectance robustly without assuming that geometry is known in advance or that reflectance is homogeneous across the surface. The concluding chapter of this thesis suggests possible approaches to overcoming the current limitations of our classification techniques.

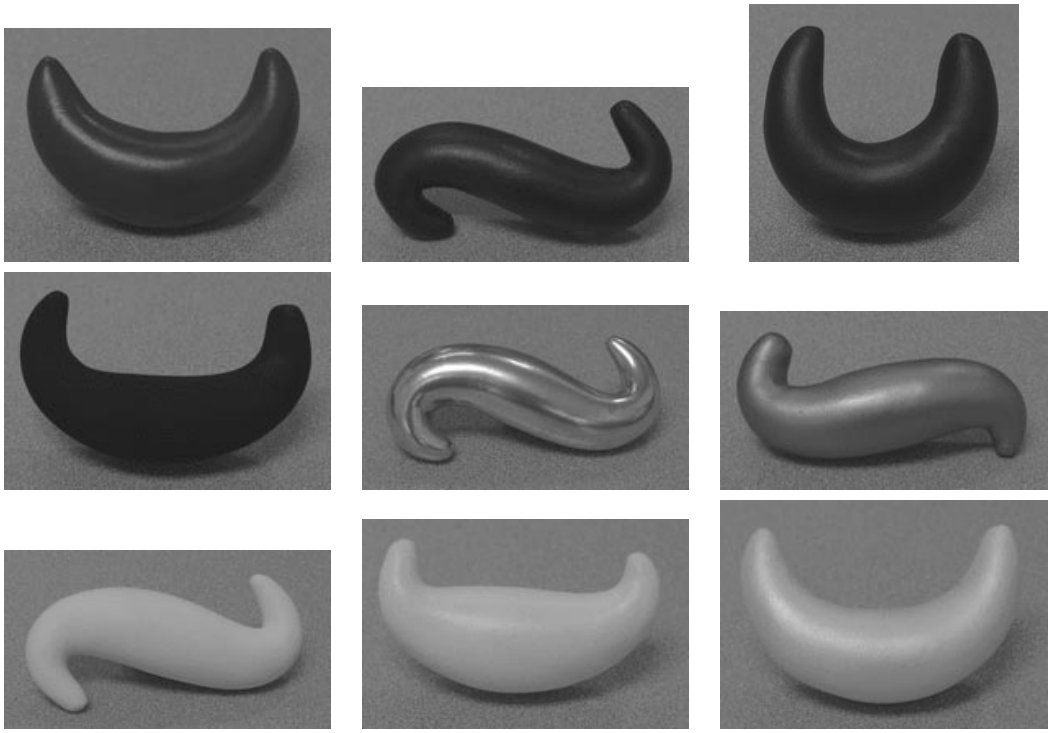


Figure 6.17. Painted clay shapes with non-convex geometries. The reflectances correspond to those of Figure 6.16.

Chapter 7

Conclusions and Suggestions for Future Research

This thesis represents progress toward the goal of allowing a computer vision system to recognize material properties in the real world. We have shown that the regular statistical properties of illumination lead to predictable relationships between reflectance and certain image statistics. One can use these statistics to classify reflectance properties reliably, even though the problem of reflectance recognition under arbitrary illumination is underconstrained. Real-world illumination statistics are relevant to other problems in vision and graphics, including recovery of illumination and geometry. Our computational work also provides a foundation for investigation of the mechanisms underlying reflectance recognition in the human visual system. This computational work has motivated experimental work by ourselves and others [33, 34, 41].

Although the algorithms described in this thesis are applicable in a variety of vision and graphics systems requiring the ability to recognize materials or recover reflectance properties, effective application in most cases requires further research. In particular, one might generalize the techniques developed in this thesis to handle shadows, unknown geometry, and spatially varying reflectance. One could also improve performance by exploiting additional image cues such as motion, color, and context. This chapter summarizes the contributions of the thesis before discussing these future research directions. We also discuss the relationship between reflectance and texture in the broader context of material recognition.

■ 7.1 Thesis Contributions

The most substantial contributions of this thesis may be in the basic problem formulation. We suggest that reflectance recognition be approached as a statistical problem, and that statistical properties of illumination be used to regularize underconstrained visual recognition problems.[1] This section recapitulates the specific results of Chapters 4, 5, and 6.

[1] Both of these ideas also arise in contemporary work by Weiss [119], although in a different form and for different applications.

7.1.1 Illumination Statistics

One can characterize illumination at a single point using a spherical illumination map that indicates the amount of light incident at that point from every direction. We analyzed two collections of illumination maps acquired photographically at points in the real world, and found that certain statistics are relatively invariant from one illumination map to another. For example, the pixel intensity distributions peak at low intensities, with fewer pixels of much higher intensity. The frequency spectra, computed using spherical harmonics, fall off at a predictable rate at high frequencies. Wavelet coefficients at each scale and orientation have highly kurtotic distributions of a predictable shape. Wavelet coefficients adjacent in scale, orientation, and position exhibit strong statistical dependencies. Although the coefficients themselves are roughly uncorrelated, their magnitudes are heavily correlated. These predictable statistics correspond to intuitive notions such as the presence of sharp edges at different scales in real-world illumination patterns.

A large existing literature on natural image statistics has emphasized the fact that real-world images display predictable statistical structure. Computer graphics researchers have pointed out that one can represent illumination at a point with a spherical map, and that such a representation proves useful in rendering surfaces under distant illumination. To the best of our knowledge, however, the combination of these two concepts is novel.

Illumination maps are "natural images," in the sense of images acquired in the real world. Unlike the photographs analyzed in most of the natural image statistics literature, however, the illumination maps we analyzed have a very wide field of view and contain primary light sources represented with high dynamic range. The statistics we report for real-world illumination are similar to those reported in the natural image statistics literature. One significant difference is the variability of illumination power spectra at low frequencies, caused by the presence of strong point-like light sources in some scenes. Illumination maps lack statistical stationarity, as evidenced by different distributions of illumination intensity at different elevations. Typical photographs also likely lack stationarity, but their non-stationary properties have received little attention in the literature.

7.1.2 Relationship Between Surface Reflectance and Image Statistics

The regularity of real-world illumination leads to informative relationships between the reflectance of a surface and certain statistics of an image of that surface. We studied these relationships empirically and analytically using a parametrized reflectance model from computer graphics, the Ward model. Although this model does not capture the full range of real-world reflectances, it allows us to examine the effects of common variations in surface reflectance on statistics of the observed image. We identified several statistics that provide information about surface reflectance:

- The brightness of the darkest portions of the image, as quantified for example by

the tenth percentile of pixel intensity, is indicative of the diffuse reflectance.

- Mean pixel intensity increases with overall surface reflectance, both specular and diffuse.
- The heavy positive tail of the distribution of image pixel intensities, as measured by its kurtosis or 90th percentile, decreases for surfaces with blurrier specular components.
- The variance of wavelet coefficients at each scale and orientation provides a measure of total spectral energy in an oriented frequency band. This variance decreases as the specular component becomes less intense or less sharp.
- Ratios of variances of wavelet coefficients at different scales are insensitive to changes in intensity of the specular component, but sensitive to changes in sharpness.
- Blurring a real-world illumination pattern does not have a major effect on the kurtosis of its wavelet coefficient distributions. However, when the diffuse component varies slowly across the surface, as is the case for a curved surface, the kurtosis of wavelet coefficient distributions tends to decrease with blur in the specular component.
- By exploiting cross-scale dependencies of wavelet coefficients, one can construct more robust indicators of specularity. These statistics distinguish reflections of structures such as edges from high frequency content due to noise.

■ 7.1.3 Classification of Reflectance Based on Image Statistics

One can take advantage of the relationship between image statistics and reflectance to classify surfaces according to their reflectance properties, given an image under unknown real-world illumination. Such a classifier involves the partitioning of a multi-dimensional feature space into regions corresponding to different reflectance classes. We use the image statistics proposed in Chapter 5 and recapitulated in Section 7.1.2 as features. We assign regions of the feature space to reflectance classes on the basis of empirical examples, using machine learning techniques. One can train such a classifier using either photographs of real surfaces, or renderings of surfaces under photographically-acquired illumination.

Because the number of training samples available is typically limited, the performance of such a classifier depends upon an appropriate choice of machine learning techniques. We obtained high accuracy in cross-validation tasks using classifiers based on support vector machines.

With an appropriately chosen feature set, the performance of such classifiers for surfaces with homogeneous reflectance and known geometry rivals or exceeds human performance. The human visual system, however, operates robustly under a far broader range of conditions.

7.1.4 Effect of Surface Geometry

The geometry of a surface affects the aforementioned image statistics, so ignoring surface scale and geometry degrades the performance of a reflectance classifier. We demonstrate a method to incorporate known geometry into reflectance estimation. This method exploits the observation that, given a distant viewer and distant illumination, the radiance of a surface patch of a particular reflectance depends only on its orientation. One can therefore map the observed image to the image of a hypothetical sphere of the same reflectance under the same illumination, and apply a reflectance classifier to that sphere. Reflectance classification based on this technique proves robust to errors in assumed geometry for convex surfaces, but much less so for non-convex surfaces.

7.2 Suggestions for Future Work

7.2.1 Refinement of Reflectance Estimation Algorithms

The reflectance classification algorithms presented in this thesis leave room for a number of potential algorithmic refinements and extensions. First of all, one could recast reflectance recognition as a parameter estimation or regression problem, as described in Section 3.3.1. Such a system might estimate Ward model parameters from an observed image, for example. The image statistics we use for reflectance classification are likely to prove useful for parameter estimation as well. One could train a parameter estimation system from sample images using appropriate machine learning techniques.

Although we have found an effective set of features for reflectance classification, we cannot claim that they are in any sense optimal. One could likely find additional and better features, particularly for classification of reflectances not captured by the Ward model. If one assumes a generative probabilistic model for illumination and a parametric model for reflectance, one could derive optimal estimators for reflectance parameters. An accurate generative model for natural images and illumination maps remains an open research problem in its own right, but even analysis using an existing image model (*e.g.*, [40, 58, 90]) may yield further insight into the reflectance estimation problem.

The best choice of features for reflectance classification depends on the specific classes to be distinguished. One could improve classifier performance by tailoring the feature set to the particular classification or parameter estimation problem at hand. For example, statistics based on wavelet coefficients at fine scales are important in distinguishing a mirrored surface from a rough metallic surface, but coefficients at coarser scales are more important in distinguishing rough metal from a matte surface. Selecting the optimal feature set for a particular classification task remains an open problem. Feature reduction via an unsupervised learning technique such as Principle Components Analysis (PCA) generally degrades classifier performance.

The confidence with which particular statistical features can be recovered depends on the geometry of the observed surface. A highly curved surface compresses the reflected image, for example, eliminating or aliasing local information about the finest-

scale features. A reflectance classifier will perform best on a wide range of surface geometries if it takes the uncertainty associated with various observed image statistics into account, basing its classification primarily on the higher-confidence statistics. The design of such a classifier represents an area for future research.

■ 7.2.2 An Alternative Approach to Compensating for the Effects of Surface Geometry

The approach to handling surface geometry described in Section 6.2.1 suffers from severe degradation in performance when the surface is non-convex and the assumed geometry is inaccurate. This degradation is due to inconsistencies between different image regions mapped onto the same part of the Gaussian sphere (Section 6.2.3). One could prevent these problems by avoiding explicit reconstruction of a Gaussian sphere.

Instead, one could compute bandpass filter outputs directly on neighborhoods of the observed image, compensating locally for the effects of surface geometry on the image. One can think of the neighborhood of any point in the observed image as a warped version of a neighborhood of a Gaussian sphere. The statistics of the Gaussian sphere itself are approximately stationary (Appendix C), but the statistics of the observed image are nonstationary because the mapping from the Gaussian sphere to the observed image depends on local surface geometry.

Appendix C shows that one can approximate the transformation locally as a linear warping that depends only on local surface orientation and curvature. If one knows the orientation and curvature associated with each point on the surface, one can apply the inverse warping transformation to a neighborhood of each point in the observed image before computing local wavelet filter outputs. Alternatively, one could warp the wavelet filters themselves before applying them to each neighborhood of the image. One could then compute statistics of marginal and joint wavelet coefficient distributions as if the coefficients had been evaluated directly on the Gaussian sphere.

Each wavelet coefficient computed using this scheme depends on only one neighborhood of the observed image. Differences in image intensity between distant surface points with the same normal will therefore not give rise to large spurious wavelet coefficients, as they did on the reconstructed Gaussian sphere. The local analysis approach is robust to variation in illumination across the surface, as long as the illumination in each region has characteristic real-world statistics and the changes in illumination are not abrupt. This approach to handling geometry provides a more reasonable model than explicit reconstruction of a Gaussian sphere for the mechanisms employed by the human visual system. For example, Ostrovsky *et al.* [76] have found that humans are insensitive to inconsistencies in illumination from one part of a scene to another. The local analysis approach also permits the use of oriented planar wavelet basis functions without the distortions of a global cartographic projection.

These advantages come at a computational price. The warped wavelet basis functions may differ at every point in the image. Instead of a traditional wavelet pyramid, one must therefore compute outputs of spatially varying filters. One may be able to

reduce the computational burden by approximating the desired filter outputs as linear combinations of outputs of a spatially invariant filter bank applied to the entire image, using for example the deformable kernel approach of Perona [78].

■ 7.2.3 Shadows, Unknown Geometry, and Spatially Varying Reflectance

For most practical applications involving recognition of materials and material properties from real-world photographs, one needs a vision system free of the assumptions of distant illumination, known geometry, and homogeneous surface reflectance properties. Relaxing the assumption that both direct and indirect sources of illumination are distant introduces complications due to cast shadows, which may lead to high-contrast edges even in images of diffuse surfaces. To deal with this issue in the present framework, we need to understand statistically how real-world illumination varies as one moves across a surface. Weiss [119] used one such statistical characterization to eliminate the effects of shadows, but he assumed the availability of multiple registered images of a scene under different illumination conditions.

In most vision applications, surface geometry is not known in advance.[2] We therefore wish either to recover surface geometry and reflectance properties simultaneously or to recognize reflectance in a manner that is invariant to surface geometry. One might approach these challenges by first addressing the easier problem of recovering surface geometry for a specular surface of known reflectance under unknown real-world illumination. If one thinks of real-world illumination patterns as instances of a random process with known statistical properties, this problem parallels that of shape-from-texture [17, 62]. Its solution may exploit the analysis of Appendix C.1, which relates the local statistics of a surface image under real-world illumination to its geometry as well as its reflectance.

One rarely begins image analysis with an image segmented into surfaces of different materials. Instead, one must distinguish reflectance boundaries between surfaces and between different materials within a surface. Strong indicators of reflectance properties, such as specular highlights, typically occupy only a small portion of the observed image of a surface. Image segmentation based on reflectance properties is therefore a difficult open problem. One may be able to make progress on this front by associating uncertainty measures with local reflectance estimates and then propagating reflectance estimates across the image using the framework proposed by Freeman *et al.* [36] and Tappen [107].

Surface reflectance also varies spatially across the surface of a textured material. In this case, the pattern of spatial variation itself represents an important visual characteristic of the material. We discuss textured surfaces in more detail in Section 7.2.5.

Removing the assumptions of distant illumination, known geometry, and spatial homogeneity makes the reflectance recognition problem more difficult. Fortunately, one

[2] Geometry may be known in advance in some applications. Sato *et al.* [92] suggest the use of a camera equipped with a laser range finder that captures an optical image and the corresponding range image simultaneously.

can also exploit additional cues that make it easier.

■ 7.2.4 Exploiting Additional Image Cues

This thesis explored the problem of reflectance estimation given only a single monochrome image of the surface of interest, without image context. We restricted ourselves to this minimal problem in order to determine what information about reflectance properties one can glean from basic image structure. A practical reflectance recognition system should combine that information with cues from other sources, including color, motion, and image context.

Surface color represents an important material characteristic in its own right. Deducing surface color from image information under unknown illumination represents the challenging problem of color constancy, which has itself been the focus of extensive research (see Section 2.3). Image color also facilitates the separation of specular and diffuse reflections for a dielectric surface [97, 111], because for this class of materials, the diffuse reflections are colored by the surface while specular reflections are not.

Image motion in a video sequence provides additional information about reflectance. As either the observer or the surface moves, the changes in the observed image depend on the reflectance properties of the surface under observation. In particular, specular and diffuse components of the reflected image will move differently. Szeliski *et al.* [83] used image motion to separate a sharp reflection in a glass plane from a photograph behind the pane, treating the observed image sequence as a sum of two transparent layers moving at different velocities. A popular model for surfaces in computer graphics assumes that the specular reflectance properties are homogeneous while the diffuse reflectance varies spatially [56]. The specular and diffuse reflections of such a surface will move as if they were separate transparent layers similar to those of Szeliski *et al.*, and one may be able to separate them using similar techniques.

In estimating the reflectance properties of a surface, one can make use of the entire image, not just the portion of the image corresponding to the surface of interest. We discussed the importance of context in estimating overall illumination strength in Section 3.4. Image context can also provide information about the spatial structure of illumination, because the illumination of nearby surfaces in an image tends to be similar, and because some sources of direct and indirect illumination may actually be visible in the image. For example, the reflections visible in a chrome surface typically resemble the surrounding image.

■ 7.2.5 Relationship Between Reflectance and Texture

To recognize materials and their properties, one cannot rely on reflectance alone. One must also exploit texture, the fine-scale variations in geometry and reflectance that characterize materials such as burlap, granite, plaster, and sand. Most researchers have treated texture as a two-dimensional image property, as if it were due to variation in the albedo of a Lambertian surface. Recent authors have noted, however, that reflectance and texture are intricately connected [21, 22]. Surface appearance depends

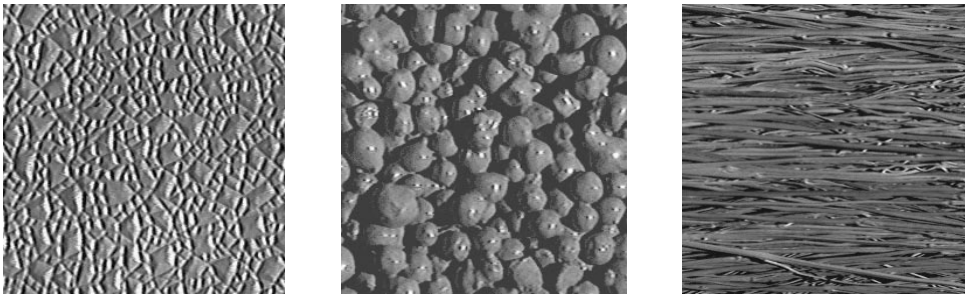


Figure 7.1. Images from the Columbia-Utrecht Reflectance and Texture database [22], which has formed the basis for recent work on discrimination of three-dimensional textures. The images in the database are all photographed under point source illumination, but with a variety of illumination directions and from a variety of view directions.

on illumination and on the position of the viewer both because of non-Lambertian reflectance properties and because of fine-scale variation in three-dimensional surface geometry. The distinction between reflectance and texture is largely a matter of scale. When one views grass at a distance, one cannot see the individual blades, but both their geometric structure and their reflectance properties will contribute to the BRDF of the observed surface. In fact, many physically-based reflectance models are formulated in terms of surface texture below the visible scale (*e.g.*, [42, 54, 69]). The remote sensing community also uses BRDFs to capture the average reflectance properties of structured surfaces [104].

Recent work by Leung and Malik [59], Cula and Dana [19], and Varma and Zisserman [116] has shown that one can classify three-dimensional textures from images using the statistical distribution of filter responses. Figure 7.1 shows examples of the images they considered. While these authors have focused on recognition of textured surfaces under point source illumination, this thesis has concentrated on recognition of untextured surfaces under complex illumination. A material recognition system should be able to combine image cues due to surface texture and those due to the texture-like properties of real-world illumination.

In this broader context of material recognition, a wider range of image statistics becomes relevant. Statistics that were not useful in distinguishing the reflectances of texture-free surfaces from one another may allow one to distinguish those texture-free surfaces from textured surfaces. In Section 5.1.2, for example, we noted that the kurtosis of wavelet coefficients in a particular band is of limited use in reflectance recognition for flat surfaces, because it does not change significantly with the parameters of a reflectance model. Textured surfaces, on the other hand, may have wavelet coefficients with less kurtotic wavelet distributions, so these kurtoses could be useful in recognizing that a reflective surface is free of texture in the first place.

High-frequency variations in an image due to reflected illumination typically exhibit edge-like structures (Section 5.1.3), while those due to geometric texture may not. Statistics based on correlations of wavelet coefficient magnitudes at adjacent scales,

orientations, and positions may therefore be useful in distinguishing smooth specular surfaces from textured surfaces. A potentially fruitful line of further research involves development of a material recogntion system that handles both textured and untextured surfaces under real-world illumination.

Appendix A

Expression of a Specular Image as a Convolution

For a distant viewer, the image produced by the specular component of the Ward reflectance model described in Section 2.1.2 can be approximated by a convolution over illumination. This convolution differs from that relating the illumination to the reflected light field, derived in Section 3.6. The expression of a specular image as a convolution depends on the fact that a mirrored surface patch with normal elevation angle γ reflects light from an elevation angle 2γ toward a vertically upward viewer (Figure A.1). Under Ward-like reflectance models, the specularity visible on the surface at a position specified by surface normal (γ, β) corresponds to to an integral over an illumination region centered around direction $(2\gamma, \beta)$. While we utilize the Ward model BRDF in the derivation below, a similar derivation applies to the specular components of the Phong and Blinn-Phong shading models.

Let $I(\gamma, \beta)$ represent the observed image of a convex surface under distant illumination, parameterized by surface normal (γ, β). This image corresponds to a cross-section

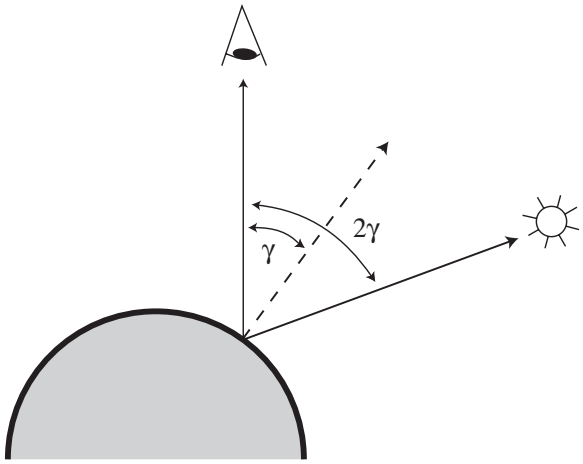


Figure A.1. A viewer at elevation angle 0 observing a surface patch whose normal has elevation angle γ will see the specular reflection of a light source at elevation angle 2γ.

of the reflected light field $B(\gamma, \beta; \theta'_r, \phi'_r)$. For a distant viewer in direction (θ_r, ϕ_r), it can be expressed as

$$I(\gamma, \beta) = B(\gamma, \beta; R^{-1}_{\gamma,\beta}(\theta_r, \phi_r)),$$

where $R^{-1}_{\gamma,\beta}(\theta_r, \phi_r)$ is a rotation operator mapping local coordinates to global coordinates, as defined by Equations 3.2 and 3.3 of Section 3.1.2.

Because we are concerned only with a single image, we can assume without loss of generality that the observation direction (θ_r, ϕ_r) is $(0, 0)$. We can then write the image as

$$\begin{aligned} I(\gamma, \beta) &= B\left(\gamma, \beta; R^{-1}_{\gamma,\beta}(0,0)\right) \\ &= \int_S L(\theta_i, \phi_i) \tilde{f}\left(\theta'_i, \phi'_i; R^{-1}_{\gamma,\beta}(0,0)\right) d\omega \\ &= \int_S L(\theta_i, \phi_i) \tilde{f}\left(\theta'_i, \phi'_i; R_y(-\gamma) R_z(-\beta)(0,0)\right) d\omega \\ &= \int_S L(\theta_i, \phi_i) \tilde{f}\left(\theta'_i, \phi'_i; \gamma, \pi\right) d\omega. \end{aligned} \tag{A.1}$$

We draw upon the properties of specular reflection to further rewrite this expression. The specular component of the Ward model has BRDF

$$f(\theta'_i, \phi'_i; \theta'_r, \phi'_r) = \rho_s \frac{1}{\sqrt{\cos\theta'_i \cos\theta'_r}} \frac{\exp(-\tan^2\delta/\alpha^2)}{4\pi\alpha^2}.$$

The corresponding modified reflectance function, defined as in Equation (3.5), is

$$\tilde{f}(\theta'_i, \phi'_i; \theta'_r, \phi'_r) = \begin{cases} \frac{\rho_s}{4\pi\alpha^2} \sqrt{\frac{\cos\theta'_i}{\cos\theta'_r}} \exp(-\tan^2\delta/\alpha^2) & \text{if } \theta'_i \le \frac{\pi}{2} \\ 0 & \text{if } \theta'_i > \frac{\pi}{2}. \end{cases}$$

In Section 2.1.2 and Figure 2.3, we defined δ as the angle between the surface normal and the vector bisecting the illumination direction (θ'_i, ϕ'_i) and view direction (θ'_r, ϕ'_r). As long as α is small, $\exp(-\tan^2\delta/\alpha^2)$ is significantly nonzero only when δ is small. This angle δ is small only when $\theta'_i \approx \theta'_r$, so we conclude that $\sqrt{\frac{\cos\theta'_i}{\cos\theta'_r}} \approx 1$ except where $\tilde{f}$ is close to 0 or θ'_r is close to $\pi/2$.

If $(\theta'_r, \phi'_r) = (\gamma, \pi)$, then δ is small only when (θ'_i, ϕ'_i) is close to $(\gamma, 0)$. In that case, the angle δ between $(0,0)$ and the bisector of (θ'_i, ϕ'_i) and (γ, π) is approximately half of the angle between (θ'_i, ϕ'_i) and $(\gamma, 0)$. Therefore δ is approximately half the angle between $R_y(-\gamma)(\theta'_i, \phi'_i)$ and $(0,0)$. Combining this observation with the above observation that $\sqrt{\frac{\cos\theta'_i}{\cos\theta'_r}} \approx 1$ when δ is small, we can conclude that

$$\tilde{f}\left(\theta'_i, \phi'_i; \gamma, \pi\right) \approx \tilde{f}\left(R_y(-\gamma)(\theta'_i, \phi'_i); 0, 0\right).$$

To rewrite Equation (A.1) as a convolution, note that

$$\begin{aligned}(\theta_i', \phi_i') &= R_{\gamma,\beta}^{-1}(\theta_i, \phi_i) \\ &= R_y(-\gamma)R_z(-\beta)(\theta_i, \phi_i)\end{aligned}$$

and therefore

$$\begin{aligned}R_y(-\gamma)(\theta_i', \phi_i') &= R_y(-\gamma)R_y(-\gamma)R_z(-\beta)(\theta_i, \phi_i) \\ &= R_y(-2\gamma)R_z(-\beta)(\theta_i, \phi_i) \\ &= R_{2\gamma,\beta}^{-1}(\theta_i, \phi_i).\end{aligned}$$

Equation (A.1) therefore becomes

$$I(\gamma, \beta) \approx \int_S L(\theta_i, \phi_i)\tilde{f}\left(R_{2\gamma,\beta}^{-1}(\theta_i, \phi_i); 0, 0\right) d\omega. \tag{A.2}$$

While defined on a sphere, the convolution of Equation (A.2) differs from that of Equation (3.6) in that it depends on the rotation operator $R_{2\gamma,\beta}^{-1}$ rather than $R_{\gamma,\beta}^{-1}$. One could rewrite it using a standard spherical convolution operator by defining a new illumination, $\tilde{L}(\theta_i, \phi_i) = L(2\theta_i, \phi_i)$. This modified illumination depends on the view direction.

The approximation of Equation (A.2) breaks down when the Ward model roughness parameter α becomes large or when the surface becomes nearly perpendicular to the viewer, with γ approaching $\frac{\pi}{2}$. However, these cases are of limited interest in practice. Ward intended his model for use with small α values; the model is properly normalized only when α is small. Portions of the surface that are nearly perpendicular to the viewer will be highly compressed in the image, and therefore are of limited value in reflectance estimation.

The complete Ward model includes a Lambertian component as well as a specular term. The image observed by a distant viewer, parameterized by the surface normal, can therefore be expressed as the sum of two spherical convolutions, one over the illumination $L(\theta_i, \phi_i)$ and the other over a modified illumination $\tilde{L}(\theta_i, \phi_i) = L(2\theta_i, \phi_i)$. This formulation allows analysis of the ambiguities involved in recovering Ward model reflectance parameters from a single image under unknown illumination. First, one cannot distinguish an overall amplitude scaling factor in the illumination from an overall amplitude scaling factor in the reflectance. Second, blurring the illumination has the same effect as blurring the specularity (increasing α); as discussed in Section 3.2.1, Ramamoorthi and Hanrahan [86] demonstrated that this ambiguity exists even when the entire light field is available. A third ambiguity stems from the fact that the specular reflection of low-frequency illumination may produce an image identical to that of the diffuse reflection under a different illumination. For example, one could match the image of a Lambertian surface under point source illumination by a specular surface of any desired roughness α, given the right low-frequency illumination pattern. This ambiguity generally disappears when images of the surface are available from all directions.

Appendix B

Rendering Under Photographically-Acquired Illumination

We used Ward's freely available[1] rendering package, *Radiance*, to render surfaces under photographically-acquired illumination maps. Unlike traditional ray tracing packages, *Radiance* takes into account illumination incident from all directions. Rendering a surface accurately under such *global illumination* is computationally expensive, because evaluation of each pixel in the rendered image requires a weighted integral over all possible illumination directions (Equation (3.1)). *Radiance* performs this integration efficiently using stochastic sampling and irradiance caching techniques, as described by Ward and Shakespeare [118]. These techniques introduce noise to the rendering process. For the benefit of a reader who intends to render surfaces under photographically-acquired illumination maps in *Radiance*, we describe the methods we used to minimize the effects of this noise. Debevec provides a more general discussion of rendering under image-based lighting [23].

When rendering a surface, we mapped the illumination to the inside of a spherical surface of infinite distance, a "source" in *Radiance*. We used the "glow" material type for this source such that *Radiance* would rely exclusively on indirect lighting calculations. When rendering in *Radiance*, we used surfaces with reflectance properties described by the Ward model (Section 2.1.2). *Radiance* employs separate algorithms to evaluate the contribution of the specular and diffuse reflectance components at each point on the surface. For a particular surface geometry and a particular illumination map, we first rendered a purely diffuse surface and purely specular surfaces with each desired value of the specular blur parameter α, and then combined the resulting images into renderings

[1] Source code and binaries for *Radiance* are available at *http://radsite.lbl.gov/radiance/*.

of the surface with each desired reflectance.[2]

As discussed in Chapter 3, the light field reflected by a surface can be expressed as a convolution of the illumination map and the BRDF. If the BRDF lacks power at high spatial frequencies, one can filter the illumination map to remove high frequency components without affecting the resulting light field. Filtering the illumination maps prior to rendering reduces noise introduced by stochastic sampling of the illumination map. To render the diffuse component, we prefiltered the illumination map by preserving only spherical harmonics through second order [84]. To render the specular component, we blurred the illumination map using a filter whose bandwidth depends on the specular blur parameter α. For a chrome surface with $\alpha = 0$, the illumination map was not blurred at all.

To further reduce image noise, we rendered images that were larger than the desired image size, and then downsampled these images. For specular components with nonzero values of α, we oversampled by a factor of eight in each direction.

Finally, the default compilation of *Radiance* evaluates the indirect specular reflection component using a sampling method based on a low-discrepancy (quasirandom) sequence of ray directions. We found that this technique introduced anisotropic, structured artifacts that remained evident even when we employed the previously described prefiltering and oversampling techniques. We recompiled *Radiance* with the -DMC flag, which forces it to use strict Monte Carlo sampling in evaluating the specular reflection.

[2] For convex surfaces, rendering the diffuse and specular components separately and then summing the resulting images is equivalent to rendering a single image with both specular and diffuse reflectance components. For non-convex surfaces, these two methods lead to slightly different images. If one sums separately rendered specular and diffuse components, the resulting image will fail to capture certain interreflections, such as a specular reflection of light reflected diffusely from another part of the surface. For the image sets presented in this thesis, we performed renderings using both methods to verify that the difference in the results was minor.

Appendix C

Effect of Geometry on Image Statistics

This appendix complements Sections 6.2.2 and 7.2.2 by extending the analysis of the relationship between surface geometry and observed image statistics. Section C.1 considers the transformation from a neighborhood of the Gaussian sphere to a neighborhood of the observed image in the three-dimensional world. Section C.2 discusses the assumption that the statistics of the Gaussian sphere itself are stationary.

■ C.1 Distortion Due to Geometry: Analysis in Three Dimensions

In Section 6.2.2, we considered a neighborhood of a one-dimensional image in a two-dimensional world as a warped neighborhood of a Gaussian sphere. We showed that one could approximate this warping locally as a compression or dilation that depends only on local surface orientation and curvature. In this section, we perform corresponding analysis for a two-dimensional image of a surface in a three-dimensional world. Savarese and Perona [93,94] have also recently analyzed the relationship between the geometry of a reflective surface and its image, although they assume a mirrored surface and known illumination.

We introduce a coordinate system (x, y, z), where the z-axis is in the direction of a distant viewer and the x and y axes are orthogonal to the view direction and to one another. One can parameterize the observed image by x and y. If the surface is differentiable at a point, one can model the surface in a neighborhood of that point as a quadratic function,

$$z = \frac{1}{2}ax^2 + bxy + \frac{1}{2}cy^2 + dx + ey + f, \tag{C.1}$$

where we assume the neighborhood of interest surrounds the point at $(x, y) = (0, 0)$. The first and second derivatives at $(x, y) = (0, 0)$ are

$$\begin{pmatrix} \frac{\partial z}{\partial x} \\ \\ \frac{\partial z}{\partial y} \end{pmatrix} = \begin{pmatrix} d \\ \\ e \end{pmatrix} \qquad \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \\ \frac{\partial^2 z}{\partial x \partial y} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} = \begin{pmatrix} a & b \\ \\ b & c \end{pmatrix}.$$

We define the vector-valued function $\mathbf{n}$ by

$$\mathbf{n}(x,y) \equiv \begin{pmatrix} -\frac{\partial z}{\partial x} \\ -\frac{\partial z}{\partial y} \\ 1 \end{pmatrix}. \tag{C.3}$$

Note that $\mathbf{n}(x,y)$ is normal to the surface at (x,y). At $(x,y)=(0,0)$, Equation (C.3) reduces to $\mathbf{n}(0,0) = \begin{pmatrix} -d \\ -e \\ 1 \end{pmatrix}$. Taking derivatives of $\mathbf{n}$ with respect to x and y gives

$$\frac{\partial \mathbf{n}}{\partial x} = \begin{pmatrix} -\frac{\partial^2 z}{\partial x^2} \\ -\frac{\partial^2 z}{\partial x \partial y} \\ 0 \end{pmatrix} \qquad \frac{\partial \mathbf{n}}{\partial y} = \begin{pmatrix} -\frac{\partial^2 z}{\partial x \partial y} \\ -\frac{\partial^2 z}{\partial y^2} \\ 0 \end{pmatrix}. \tag{C.4}$$

Evaluating Equation (C.4) at $(x,y)=(0,0)$ gives

$$\frac{\partial \mathbf{n}}{\partial x}\Big|_{(x,y)=(0,0)} = \begin{pmatrix} -a \\ -b \\ 0 \end{pmatrix} \qquad \frac{\partial \mathbf{n}}{\partial y}\Big|_{(x,y)=(0,0)} = \begin{pmatrix} -b \\ -c \\ 0 \end{pmatrix}. \tag{C.5}$$

The local tangent projection of the Gaussian sphere at any point can be parameterized by two variables. For convenience, we parameterize that projection locally using the following two unit vectors,[1] which are orthogonal to each other and to the local unit normal $\mathbf{n}_0 = \frac{1}{\sqrt{1+d^2+e^2}} \begin{pmatrix} -d \\ -e \\ 1 \end{pmatrix}$:

$$\begin{aligned} \mathbf{v}_1 &\equiv \frac{1}{\sqrt{(d^2+e^2)(1+d^2+e^2)}} \begin{pmatrix} d \\ e \\ d^2+e^2 \end{pmatrix} \\ \mathbf{v}_2 &\equiv \frac{1}{\sqrt{d^2+e^2}} \begin{pmatrix} -e \\ d \\ 0 \end{pmatrix}. \end{aligned}$$

We can expand $\frac{\partial \mathbf{n}}{\partial x}$ and $\frac{\partial \mathbf{n}}{\partial x}$ from Equation (C.5) in terms of $\mathbf{n}_0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ as

$$\begin{aligned} \frac{\partial \mathbf{n}}{\partial x} &= \frac{be+da}{\sqrt{e^2+d^2+1}}\mathbf{n}_0 + \frac{-be-da}{\sqrt{(e^2+d^2)(e^2+d^2+1)}}\mathbf{v}_1 + \frac{ea-db}{\sqrt{e^2+d^2}}\mathbf{v}_2 \\ \frac{\partial \mathbf{n}}{\partial y} &= \frac{db+ec}{\sqrt{e^2+d^2+1}}\mathbf{n}_0 + \frac{-db-ec}{\sqrt{(e^2+d^2)(e^2+d^2+1)}}\mathbf{v}_1 + \frac{be-dc}{\sqrt{e^2+d^2}}\mathbf{v}_2. \end{aligned} \tag{C.6}$$

[1] If d = e = 0, let $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Let $\tilde{\mathbf{n}}$ be the unit vector in the direction of $\mathbf{n}$, defined by $\tilde{\mathbf{n}}(x,y) \equiv \frac{\mathbf{n}(x,y)}{\|\mathbf{n}(x,y)\|}$. Then $\frac{d\tilde{\mathbf{n}}}{dx}$ is the projection of $\frac{d\mathbf{n}}{dx}$ onto the plane perpendicular to $\mathbf{n}$, normalized by $\|\mathbf{n}\|$. When $(x,y) = (0,0)$, Equation (C.6) implies that

$$\begin{aligned}
\frac{\partial\tilde{\mathbf{n}}}{\partial x} &= \frac{1}{\sqrt{(e^2+d^2)(e^2+d^2+1)}}\left(\frac{-be-da}{\sqrt{e^2+d^2+1}}\mathbf{v}_1 + (ea-db)\mathbf{v}_2\right) \\
\frac{\partial\tilde{\mathbf{n}}}{\partial y} &= \frac{1}{\sqrt{(e^2+d^2)(e^2+d^2+1)}}\left(\frac{-db-ec}{\sqrt{e^2+d^2+1}}\mathbf{v}_1 + (be-dc)\mathbf{v}_2\right). \qquad \text{(C.7)}
\end{aligned}$$

Additional algebra yields the more elegant form

$$\begin{pmatrix} \frac{\partial\tilde{\mathbf{n}}}{\partial x} & \frac{\partial\tilde{\mathbf{n}}}{\partial y} \end{pmatrix} = -\begin{pmatrix} \frac{\mathbf{v}_1}{e^2+d^2+1} & \frac{\mathbf{v}_2}{\sqrt{e^2+d^2+1}} \end{pmatrix} \begin{pmatrix} \frac{d}{\sqrt{d^2+e^2}} & \frac{e}{\sqrt{d^2+e^2}} \\ -\frac{e}{\sqrt{d^2+e^2}} & \frac{d}{\sqrt{d^2+e^2}} \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \qquad \text{(C.8)}$$

Equations (C.7) and (C.8) specify the change in surface normal direction as one moves horizontally or vertically across the image. In other words, they specify the local linear approximation of the warping from the neighborhood of $\mathbf{n}_0$ on the Gaussian sphere to the neighborhood of $(x,y) = (0,0)$ on the image.

The relationship between $\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_1 \end{pmatrix}$ and $\begin{pmatrix} \frac{\partial\tilde{\mathbf{n}}}{\partial x} & \frac{\partial\tilde{\mathbf{n}}}{\partial y} \end{pmatrix}$ specified by Equation (C.8) can be summarized by a single two-by-two matrix and therefore represents a four-parameter family of possible linear transformations.[2] The transformation depends only on the first and second derivatives of the surface at the point, or, equivalently, on the local surface orientation and curvature.

If a quadratic surface of the form in Equation (C.1) has $b = c = e = 0$, then it varies in the x-direction, but not in the y-direction. In this case, one would expect that the surface orientation will not change as one moves in the y-direction, and that the rate of change as one moves in the x-direction will reduce to the one-dimensional case. Indeed, when $b = c = e = 0$, Equation (C.7) implies that

$$\begin{aligned}
\left\|\frac{\partial\tilde{\mathbf{n}}}{\partial y}\right\| &= 0 \\
\left\|\frac{\partial\tilde{\mathbf{n}}}{\partial x}\right\| &= \frac{|a|}{1+d^2}
\end{aligned}$$

in agreement with Equation (6.11), which gives the rate of change of orientation with image position in the one-dimensional case.

Section 7.2.3 points out that because a real-world illumination map can be viewed as a type of "texture" with regular statistical properties, the problem of recovering shape

[2] If the statistics of the Gaussian sphere are isotropic, then one can ignore rotation in the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, so that one can effectively specify the warping transformation with three parameters. If illumination statistics depend on orientation, rotation in this plane may affect the image statistics.

from a reflective surface under unknown illumination parallels the traditional shape-from-texture problem. The analysis of the present section highlights one important difference. In traditional shape-from-texture, the statistics of the texture are assumed stationary on the surface. The local linear warping from the surface to the observed image depends only on the local orientation of the surface. For a reflective surface under regular illumination, on the other hand, the statistics of the Gaussian sphere are assumed stationary. The local linear warping from the Gaussian sphere to the observed image depends not only on local surface orientation, but also on local surface curvature.

■ C.1.1 Special Case of a Sphere

In order to better understand the distortion that results from orthogonal projection of the Gaussian sphere onto a plane, we evaluate Equation (C.8) for the special case where the surface under observation is a sphere of unit radius described by $x^2 + y^2 + z^2 = 1$. The derivatives at (x_0, y_0, z_0) are given by

$$\begin{pmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{pmatrix} = \begin{pmatrix} -\frac{x_0}{z_0} \\ -\frac{y_0}{z_0} \end{pmatrix} \qquad \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial x \partial y} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} = \begin{pmatrix} -\frac{x_0^2+z_0^2}{z_0^3} & -\frac{x_0 y_0}{z_0^3} \\ -\frac{x_0 y_0}{z_0^3} & -\frac{y_0^2+z_0^2}{z_0^3} \end{pmatrix}.$$

In other words, a local quadratic fit of the form of Equation (C.1) has

$$a = -\frac{x_0}{z_0}, \quad b = -\frac{y_0}{z_0}, \quad c = -\frac{x_0^2 + z_0^2}{z_0^3}, \quad d = -\frac{x_0 y_0}{z_0^3}, \quad e = -\frac{y_0^2 + z_0^2}{z_0^3}. \tag{C.10}$$

We reparameterize the observed image in terms of local coordinates p and q, where p is in the gradient (radial) direction and q is perpendicular to it. That is,

$$\begin{pmatrix} p \\ q \end{pmatrix} = \frac{1}{\sqrt{e^2 + d^2}} \begin{pmatrix} d & e \\ -e & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{e^2 + d^2}} \begin{pmatrix} d & -e \\ e & d \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}.$$

Therefore

$$\begin{pmatrix} \frac{\partial \tilde{\mathbf{n}}}{\partial p} & \frac{\partial \tilde{\mathbf{n}}}{\partial q} \end{pmatrix} = \begin{pmatrix} \frac{\partial \tilde{\mathbf{n}}}{\partial x} & \frac{\partial \tilde{\mathbf{n}}}{\partial y} \end{pmatrix} \begin{pmatrix} d & -e \\ e & d \end{pmatrix}. \tag{C.12}$$

Combining Equation (C.12) with Equation (C.8) gives

$$\begin{pmatrix} \frac{\partial \tilde{\mathbf{n}}}{\partial p} & \frac{\partial \tilde{\mathbf{n}}}{\partial q} \end{pmatrix} = -\begin{pmatrix} \frac{\mathbf{v}_1}{e^2+d^2+1} & \frac{\mathbf{v}_2}{\sqrt{e^2+d^2+1}} \end{pmatrix} \begin{pmatrix} \frac{d}{\sqrt{d^2+e^2}} & \frac{e}{\sqrt{d^2+e^2}} \\ -\frac{e}{\sqrt{d^2+e^2}} & \frac{d}{\sqrt{d^2+e^2}} \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} d & -e \\ e & d \end{pmatrix}.$$

Substituting in the values of a, b, c, d, and e from Equation (C.10) above and simplifying gives

$$\begin{pmatrix} \frac{\partial \tilde{\mathbf{n}}}{\partial p} & \frac{\partial \tilde{\mathbf{n}}}{\partial q} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} \frac{1}{z_0} & 0 \\ 0 & 1 \end{pmatrix},$$

or, equivalently,

$$\frac{\partial \tilde{\mathbf{n}}}{\partial p} = \frac{\mathbf{v}_1}{z_0}, \qquad \frac{\partial \tilde{\mathbf{n}}}{\partial q} = \mathbf{v}_2.$$

In other words, a neighborhood in the image represents a neighborhood of the Gaussian sphere that has been compressed only in the radial direction p. The compression increases as one moves from the center of the image, where $z = 1$, to the image boundary, where $z = 0$. The sharp increase in the compression factor $\frac{1}{z}$ near $z = 0$ leads us to omit a region of the original image near that boundary when unwrapping the annulus described in Section 6.1.2.

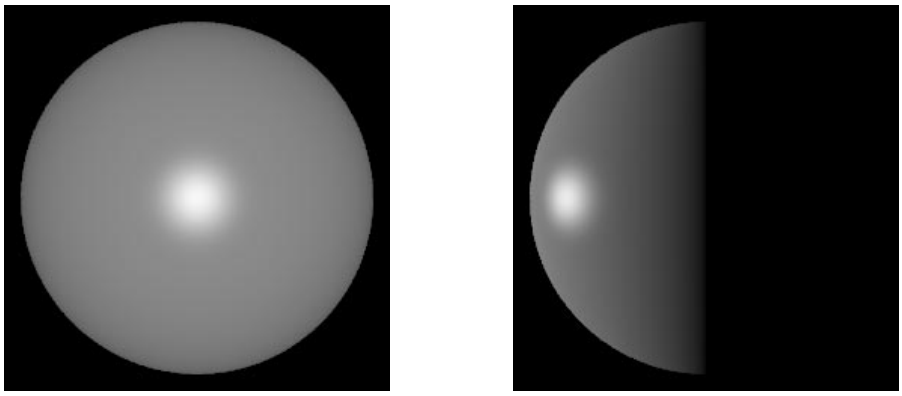
The warping associated with each neighborhood of the observed image is anisotropic, with compression only in the radial direction. This anisotropic compression affects the orientation of edge-like structures in the observed image, such that most edges tend to be oriented perpendicular to the radial direction, particularly near the boundary of the image where the compression factor is highest (for example, see the chrome spheres in Figure 6.5). The only edges whose orientations are not affected by this compression are those originally oriented in the radial direction and perpendicular to it. If we "unwrap" the image about its center as described in Section 6.1.2, these two directions map to the horizontal and vertical directions at every point in the image. This is a desirable property because we analyze the image using horizontally- and vertically-oriented filters. In Section 6.1.2, we discard the central region of the image, because it is distorted most severely in the unwrapping procedure.

■ C.2 Assumption of Stationarity on Gaussian Sphere

When discussing the effects of geometry on observed image statistics, we have assumed that the statistics of the Gaussian sphere are stationary. In particular, we have assumed that the expected value of a statistic computed on a neighborhood of the Gaussian sphere is invariant to the location of that neighborhood on the sphere. This assumption is approximately but not exactly correct. For completeness, we point out two potential sources of non-stationarity on the Gaussian sphere.

First, illumination itself is not stationary. As noted in Section 4.4.2, for example, the expected intensity of illumination depends on elevation angle. One could potentially take such nonstationarities into account if observing a surface with known orientation in global coordinates, but our reflectance recognition algorithms do not do so.

Second, the Gaussian sphere may have non-stationary statistics even when the illumination statistics are stationary. As an example, imagine a world in which illumination always consists of a point source of fixed intensity in a random direction, with all illumination directions equally likely. Consider a surface whose reflectance is described by the Ward model with non-zero specular and diffuse components. Figure C.1 shows two images of such a surface, with the light source in different locations. When the light source is directly behind the viewer (Figure C.1(a)), the diffuse and specular components peak in the same place. When the light source direction is different from the



(a) (b)

Figure C.1. (a) A sphere illuminated from directly behind the viewer. (b) The same sphere, illuminated from above. In each case, one hemisphere is illuminated. The intensity of the diffuse component peaks where the surface normal is in the same direction as the light source, while the specular component peaks where the surface normal bisects the light direction and the view direction. In (a), the specularity is in the center of the illuminated hemisphere, while in (b), it is not.

view direction (Figure C.1(b)), the diffuse and specular components peak in different places. The fact that the relative location of these two image components depends on light source position leads to nonstationarity in the Gaussian sphere. For example, the maximum intensity of a surface patch whose normal points toward the viewer is larger than the maximum intensity of patches with other surface orientations. More generally, the fact that the light field $B(\gamma, \beta; \theta'_r, \phi'_r)$ can be written as a convolution over illumination (Section 3.1.2) implies that a cross-section of B with constant values of θ'_r and ϕ'_r has stationary statistics. An observed image of a sphere, however, does not correspond to a planar cross-section of the light field B (see Section 3.1.3). The statistics of the Gaussian sphere are therefore not necessarily stationary. In practice, the effect of this nonstationarity on the statistics we use for reflectance classification is minor or negligible relative to the effect of the distortions discussed in Section C.1.

Bibliography

[1] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge, MA, 2000.

[2] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In M. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing*. MIT Press, Cambridge, MA, 1991.

[3] K. Amaratunga and J. E. Castrillon-Candas. Surface wavelets: a multiresolution signal processing tool for 3D computational modeling. *Int. J. Numer. Meth. Engng*, 52:239–271, 2001.

[4] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. In *Proceedings of the International Conference on Computer Vision*, 2001.

[5] J. Beck. *Surface color perception*. Cornell University Press, Ithaca, N.Y., 1972.

[6] J. Beck and S. Prazdny. Highlights and the perception of glossiness. *Perception & Psychophysics*, 30:407–410, 1981.

[7] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[8] J. F. Blinn. Models of light reflection for computer synthesized pictures. *Computer Graphics (SIGGRAPH)*, 11, 1977.

[9] S. Boivin and A. Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. *Computer Graphics (SIGGRAPH)*, 2001.

[10] J. R. Boyack and A. K. Juenger. Brightness adjustment of images using digital scene analysis. U.S. Patent US5724456, March 1998.

[11] D. H. Brainard, B. A. Wandell, and E.-J. Chichilnisky. Color constancy: from physics to appearance. *Cur. Dir. Psychol. Sci.*, 2:165–170, 1993.

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

[13] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8:1688–1701, 1999.

[14] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.

[15] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:955–974, 1998.

[16] Frank Canters and Hugo Decleir. *The World in Perspective: A Directory of World Map Projections*. John Wiley & Sons, New York, 1989.

[17] M. Clerc and S. Mallat. The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):536–549, 2002.

[18] R. Collobert and S. Bengio. Svmtorch: Support vector machines for large-scale regression problems. *Journal of Machine Learning Research*, 1:143–160, 2001.

[19] O. G. Cula and K. J. Dana. Recognition methods for 3D textured surfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[20] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18:1–34, 1999.

[21] Kristin J. Dana, Shree K. Nayar, Bram. van Ginneken, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 151–157, 1997.

[22] 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, January 1999.

[23] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *Computer Graphics (SIGGRAPH)*, 1998.

[24] P. E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. *Computer Graphics (SIGGRAPH)*, 2000.

[25] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Computer Graphics (SIGGRAPH)*, pages 369–78, 1997.

[26] J. Dorsey and P. Hanrahan. Digital materials and virtual weathering. *Scientific American*, 282(2):64–71, 2000.

[27] M. D'Zmura. Shading ambiguity: Reflectance and illumination. In M. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing*. MIT Press, Cambridge, MA, 1991.

[28] M. D'Zmura and G. Iverson. Color constancy. I. basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, 10:2148–2165, 1993.

[29] R. M. Evans. *The Perception of Color*. Wiley, New York, 1974.

[30] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1):1–50, 2000.

[31] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4:2379–94, 1987.

[32] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[33] R. W. Fleming, R. O. Dror, and E. H. Adelson. How do humans determine reflectance properties under unknown illumination? In *Proceedings of the CVPR Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation*, 2001.

[34] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real world illumination and the perception of surface reflectance properties. In preparation.

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

[36] W.T. Freeman, , E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[37] W.T. Freeman. Exploiting the generic viewpoint assumption. *International Journal of Computer Vision*, 20(3):243, 1996.

[38] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou. An anchoring theory of lightness perception. *Psychological Review*, 106:795–834, 1999.

[39] Andrew S. Glassner. *Principles of Digital Image Synthesis*, volume 2. Morgan Kaufmann, San Francisco, 1995.

[40] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:424–429, 2001.

[41] B. Hartung and D. Kersten. Distinguishing shiny from matte. In *Proceedings of Second Annual Meeting of the Vision Sciences Society*, page 191, Sarasota, Florida, 2002.

[42] X. D. He, K. E. Torrance, F. S. Sillion, and D. P. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics (SIGGRAPH)*, 25(4):175–86, 1991.

[43] D. J. Heeger and J. R. Bergen. Pyramid-based texure analysis/synthesis. *Computer Graphics (SIGGRAPH)*, 1995.

[44] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[45] B. K. P. Horn and M. J. Brooks. Introduction. In B. K. P. Horn and M. J. Brooks, editors, *Shape From Shading*, chapter 1, pages 1–28. MIT Press, Cambridge, MA, 1989.

[46] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[47] J. Huang and D. Mumford. Statistics of natural images and models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 541–47, 1999.

[48] T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: A sketching interface for 3D freeform design. *Computer Graphics (SIGGRAPH)*, 1999.

[49] T. Inui, Y. Tanabe, and Y. Onodera. *Group Theory and Its Applications in Physics*. Springer, Berlin, Heidelberg, second, corrected printing edition, 1996.

[50] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. *Computer Graphics (SIGGRAPH)*, 2001.

[51] J. Jensen, H. W. Buhler. A rapid hierarchical rendering technique for translucent materials. *Computer Graphics (SIGGRAPH)*, 2002.

[52] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf and C. J. C. Burges, editors, *Advances in Kernel Methods — Support Vector Learning*. MIT Press, 1999.

[53] J. D. Johnston. A filter family designed for use in quadrature mirror filter banks. In *Proc. Int. Conf Acoustics, Speech, and Signal Processing*, pages 291–294, 1980.

[54] Jan J. Koenderink, Andrea J. Van Doorn, Kristin J. Dana, and Shree Nayar. Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31:129–144, 1999.

[55] E. H. Land and J. J. McCann. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61:1–11, 1971.

[56] G. W. Larson and R. Shakespeare. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaugmann, San Francisco, 1998.

[57] S. B. Laughlin. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, 36c:910–912, 1981.

[58] A. Lee, J. Huang, and D. Mumford. Random collage model for natural images. *International Journal of Computer Vision*, submitted.

[59] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. In *Proceedings of the International Conference on Computer Vision*, pages 1010–1017, 1999.

[60] R. C. Love. *Surface Reflection Model Estimation from Naturally Illuminated Image Sequences*. PhD thesis, University of Leeds, 1997.

[61] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of 7th International Conference on Artificial Intelligence*, pages 121–130, 1981.

[62] J. Malik and R. Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *International Journal of Computer Vision*, 23:149–168, 1997.

[63] L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectances. *Journal of the Optical Society of America A*, 3:29–33, 1986.

[64] D. Marr. The computation of lightness by the primate retina. *Vision Research*, 14:1377–1388, 1974.

[65] S. R. Marschner. *Inverse Rendering for Computer Graphics*. PhD thesis, Cornell University, Ithaca, NY, 1998.

[66] S. R. Marschner, S. H. Westin, E. P. F. Lafortune, K. E. Torrance, and D. P. Greenberg. Image-based BRDF measurement including human skin. In *Proceedings of 10th Eurographics Workshop on Rendering*, pages 139–52, Granada, Spain, 1999.

[67] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Computer Graphics (SIGGRAPH)*, pages 39–46, 1995.

[68] G. Miller and Hoffman. R. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH Course Notes for Advanced Computer Graphics Animation*, 1984.

[69] S. K. Nayar and M. Oren. Visual appearance of matte surfaces. *Science*, 267:1153–6, 1995.

[70] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometrical considerations & nomenclature for reflectance. NBS Monograph 160, National Bureau of Standards, U.S. Department of Commerce, Washington, D. C., October 1977.

[71] S. Nishida and M. Shinya. Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A*, 15:2951–2965, 1998.

[72] K. Nishino. *Photometric Object Modeling: Rendering from a Dense/Sparse Set of Images*. PhD thesis, University of Tokyo, 2001.

[73] K. Nishino, Z. Zhang, and K. Ikeuchi. Determining reflectance parameters and illumination distributions from a sparse set of images for view-dependent image synthesis. In *Proceedings of the International Conference on Computer Vision*, pages 599–606, 2001.

[74] M. Oren and S.K. Nayar. Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, April 1995.

[75] Michael Oren and Shree K. Nayar. Generalization of Lambert's reflectance model. In Andrew Glassner, editor, *Proceedings of SIGGRAPH '94 (Orlando, Florida, July 24–29, 1994)*, Computer Graphics Proceedings, Annual Conference Series, pages 239–246. ACM SIGGRAPH, ACM Press, July 1994. ISBN 0-89791-667-0.

[76] Y. Ostrovsky, P. Cavanagh, and P. Sinha. Perceiving illumination inconsistencies in scenes. In preparation.

[77] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. *Computer Graphics (SIGGRAPH)*, 2000.

[78] P. Perona. Deformable kernels for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:488–499, 1995.

[79] Bui-Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[80] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf and C. J. C. Burges, editors, *Advances in Kernel Methods — Support Vector Learning*. MIT Press, 1999.

[81] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, 2000.

[82] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain. In *Proc. Int. Conf. Image Processing*, 2001.

[83] Szeliski R., Avidan S., and Anandan P. Layer extraction from multiple images containing reflections and transparency. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000.

[84] R. Ramamoorthi and P. Hanrahan. An efficient representation for environment irradiance maps. *Computer Graphics (SIGGRAPH)*, 2001.

[85] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A*, 10:2448–2459, 2001.

[86] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. *Computer Graphics (SIGGRAPH)*, 2001.

[87] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Science*, 98, 2001.

[88] R. M. Rifkin. *Everything Old Is New Again: A Fresh Look at Historical Approaches in Machine Learning*. PhD thesis, MIT, Cambridge, MA, August 2002.

[89] D. L. Ruderman. The statistics of natural images. *Network: Comput. Neural Syst.*, 5:517–548, 1994.

[90] D. L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37:3385–3395, 1997.

[91] S. Rusinkiewicz. A new change of variables for efficient BRDF representation. In *Rendering Techniques (Proceedings of the Eurographics Workshop on Rendering)*, pages 11–22, 1998.

[92] Y. Sato, M. D. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. In *Computer Graphics (SIGGRAPH)*, 1997.

[93] S. Savarese and P. Perona. Local analysis for 3D reconstruction of specular surfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[94] S. Savarese and P. Perona. Local analysis for 3D reconstruction of specular surfaces — part II. In *Proceedings of the European Conference on Computer Vision*, 2002.

[95] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. *Computer Graphics (SIGGRAPH)*, 1995.

[96] O. Schwartz and E. P. Simoncelli. Natural signal statistics and sensory gain control. *Nature: Neuroscience*, 4:819–825, 2001.

[97] S. Shafer. Using color to separate reflection components. *Color Research and Application*, 10:210–218, 1985.

[98] E. P. Simoncelli and E. H. Adelson. Subband transforms. In John W Woods, editor, *Subband Image Coding*, chapter 4, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.

[99] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Proc. Int. Conf. Image Processing*, 1996.

[100] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans. Information Theory*, 38:587–607, March 1992.

[101] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001.

[102] E.P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc SPIE, 44th Annual Meeting*, volume 3813, Denver, July 1999.

[103] P.-P. Sloan, W. Martin, A. Gooch, and B. Gooch. The Lit Sphere: A model for capturing NPR shading from art. In *Graphics Interface*, June 2001.

[104] W. C. Snyder. Definition and invariance properties of structured surface BRDF. *IEEE Trans. on Geocscience and Remote Sensing*, 40:1032–1037, 2002.

[105] S. K. Stein. *Calculus and Analytic Geometry.* McGraw-Hill, brief edition, 1987.

[106] M. F. Tappen. Unpublished, 2001.

[107] M. F. Tappen. Recovering shading and reflectance from a single image. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, 2002.

[108] S. Teller, M. Antone, M. Bosse, S. Coorg, M. Jethwa, and N. Master. Calibrated, registered images of an extended urban area. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[109] D. J. Tolhurst, Y. Tadmor, and T. Chao. Amplitude spectra of natural images. *Ophthalmology and Physiological Optics*, 12:229–232, 1992.

[110] S. Tominaga, E. Takahashi, and N. Tanaka. Parameter estimation of a reflection model from a multi-band image. In *Proceedings of the Workshop on Photometric Modeling for Computer Vision and Graphics*, pages 56–63, Fort Collins, Colorado, 1999.

[111] S. Tominaga and N. Tanaka. Estimating reflection parameters from a single color image. *IEEE Computer Graphics and Applications*, 20:58–66, September/October 2000.

[112] A. Torralba. Personal communication, August 2001.

[113] A. Torralba and P. Sinha. Contextual priming for object detection. MIT Artificial Intelligence Laboratory Memo 2001-020, 2001.

[114] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B*, 265:359–366, 1998.

[115] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1995.

[116] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the European Conference on Computer Vision*, 2002.

[117] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades on wavelet trees and their use in analyzing and modeling natural images. *Applied and Computational Harmonic Analysis*, 11:89–123, 2001.

[118] G. J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH)*, 26(2):265–272, 1992.

[119] Y. Weiss. Deriving intrinsic images from image sequences. In *Proc. Int. Conf. Computer Vision*, 2001.

[120] C.-J. Westelius. *Focus of Attention and Gaze Control for Robot Vision.* PhD thesis, Linköping University, Linköping, Sweden, 1995.

[121] C.-F. Westin, K. Nordberg, and H. Knutsson. On the equivalence of normalized convolution and normalized differential convolution. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume V, pages 457–460, 1994.

[122] L. Williams. Shading in two dimensions. In *Graphics Interface '91*, pages 143–151, 1991.

[123] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 215–224, Los Angeles, 1999.

[124] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 207–217, Orlando, Florida, 1998.