Radiometric Scene Decomposition:

Estimating Complex Reflectance and Natural Illumination from Images

A Thesis

Submitted to the Faculty

of

Drexel University

by

Stephen Anthony Lombardi

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

September 2015
Dedications

This thesis is dedicated to my parents.
Acknowledgments

I would first like to thank my advisor, Dr. Ko Nishino, for his guidance throughout this process. He has helped me in many aspects: from assisting me in finding an interesting research topic, lending technical and theoretical help, and giving critical writing and presentation feedback. This thesis would not be possible without his support and his expectation of the highest quality.

I would also like to thank my parents. Their encouragement and support in pursuing my education and throughout my life has enabled me to reach this point. I can’t thank them enough for all they have done for me.

I want to especially thank my wife, Michele Lombardi, for her unending support and devotion. When things are difficult, I can always count on her to grab a beer with me and help me relax. Finishing this degree would not have been possible without her love and encouragement.

I’d like to thank my labmates, past and present, who have often provided technical insight and humorous distractions. Most notably this includes Louis Kratz, Geoff Oxholm, Gabe Schwartz, and Leizer Teran. I’d also like to acknowledge my brother, Alex Lombardi, and my friends Chaz Ruggieri, Brandon McKoy, and Chris Laucius, who have helped me decompress through trips to all-you-can-eat sushi, bowling, and video games.

I must also thank the funding institutions that have allowed me to pursue these research topics, including the National Science Foundation, the Office of Naval Research, as well as the Drexel Department of Computer Science, the George Hill, Jr. Endowment, and the Harry Brown, Jr. Endowment.
Table of Contents

LIST OF TABLES ................................................................. viii
LIST OF FIGURES ............................................................. ix

ABSTRACT ................................................................. xv

1. INTRODUCTION ..................................................... 1
   1.1 Contributions .................................................. 3

2. BACKGROUND ...................................................... 9
   2.1 Reflectance ..................................................... 9
      2.1.1 BRDF Models ........................................... 11
      2.1.2 Measuring BRDFs ...................................... 13
   2.2 Illumination .................................................. 14

3. RELATED WORK ................................................... 16
   3.1 Early Radiometric Scene Decomposition ................. 16
   3.2 Intrinsic Images ............................................. 18
   3.3 State of the Art ............................................. 19

4. DIRECTIONAL STATISTICS BRDF ............................. 24
   4.1 Isotropic BRDF ............................................... 24
   4.2 Hemispherical Exponential Power Distribution .......... 26
   4.3 Modeling Color ............................................... 30
   4.4 Characterizing the Variation of Real-World BRDFs ....... 30
      4.4.1 A Prior on the Reflectance Bases .................. 34
      4.4.2 A Prior on Lobe Chromaticity ....................... 36
4.5 Conclusion .............................................................. 38

5. Reflectance Estimation under Point Light Illumination ............ 40
5.1 Inferring Reflectance .................................................. 40
5.2 Experimental Results .................................................. 42
5.3 Conclusion ............................................................... 44

6. Single Image Multimaterial Estimation ................................. 45
6.1 A Layered MRF Formulation ......................................... 47
6.2 Inferring Material Reflectance and Region .......................... 48
6.3 Experimental Results ................................................... 51
6.4 Conclusion ............................................................... 53

7. Reflectance and Natural Illumination from a Single Image ......... 54
7.1 Modeling Natural Illumination ......................................... 56
7.1.1 A Non-parametric Illumination Model ............................ 57
7.1.2 Illumination Priors .................................................. 57
7.1.3 The Entropy Increase by Reflection ............................... 59
7.1.4 Resolving the Color Ambiguity .................................... 61
7.2 Inferring Reflectance and Natural Illumination ..................... 63
7.3 Experimental Results ................................................... 64
7.3.1 Synthetic Results .................................................... 64
7.3.2 Real scenes .......................................................... 69
7.4 Conclusion ............................................................... 73

8. Spatially-varying Reflectance, Natural Illumination, and Refined
   Geometry from RGB-D .................................................. 74
8.1 Radiometric Scene Decomposition .................................... 77
8.2 Modeling Indirect Illumination ....................................... 78
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3 Spatial Segmentation</td>
<td>80</td>
</tr>
<tr>
<td>8.4 Spatially-varying Texture</td>
<td>83</td>
</tr>
<tr>
<td>8.5 Natural Illumination</td>
<td>85</td>
</tr>
<tr>
<td>8.6 Reflectance-aware Illumination Estimation</td>
<td>85</td>
</tr>
<tr>
<td>8.7 Optimizing Rough Geometry</td>
<td>86</td>
</tr>
<tr>
<td>8.8 Experimental Results</td>
<td>87</td>
</tr>
<tr>
<td>8.8.1 Importance of Interreflection</td>
<td>89</td>
</tr>
<tr>
<td>8.8.2 Evaluation of Geometry Refinement on Synthetic Blob Dataset</td>
<td>91</td>
</tr>
<tr>
<td>8.8.3 Decomposition on Synthetic Scenes</td>
<td>98</td>
</tr>
<tr>
<td>8.8.4 Decomposition on Real Scenes</td>
<td>98</td>
</tr>
<tr>
<td>8.9 Conclusion</td>
<td>101</td>
</tr>
<tr>
<td>9. Conclusion</td>
<td>102</td>
</tr>
<tr>
<td>9.1 Limitations</td>
<td>104</td>
</tr>
<tr>
<td>9.2 Future Work</td>
<td>105</td>
</tr>
<tr>
<td>Bibliography</td>
<td>107</td>
</tr>
<tr>
<td>Appendix A: Important Gradients for Optimization</td>
<td>112</td>
</tr>
<tr>
<td>A.1 Gradient of the Negative Log Likelihood</td>
<td>112</td>
</tr>
<tr>
<td>A.2 Gradient of the DSBRDF Model</td>
<td>117</td>
</tr>
<tr>
<td>vita</td>
<td>118</td>
</tr>
</tbody>
</table>
List of Tables
List of Figures

2.1 Diagram of the vectors used to define the BRDF. The left column shows an example outgoing direction $\omega_o$ and incident direction $\omega_i$ on the hemisphere. The right column shows a reparameterization of the BRDF using the half vector $h$ [46]. Here $n$ is the surface normal and $t$ is the surface tangent vector. Representing BRDFs using $(\theta_d, \phi_d)$ and $(\theta_h, \phi_h)$ is significantly easier than $(\theta_i, \phi_i)$ and $(\theta_o, \phi_o)$ for many real-world materials. .......................... 12

4.1 Hemispherical Exponential Power Distribution. The hemi-EPD model can represent a wide variety of hemispherical directional distributions. Here we show several example sets of parameters to visualize the expressive power of the hemi-EPD. .......................................................... 26

4.2 Synthetic spheres rendering using the DSBRDF model with parameters fit from three MERL BRDFs (nickel, specular-blue-phenolic, and orange-paint) [32]. The left column shows the spheres rendering using the measured BRDF data. The next five columns show the spheres rendered using the DSBRDF model with 1 to 5 lobes from left to right, respectively. .......................... 28

4.3 Synthetic spheres rendered with four MERL BRDFs [32], the 3-lobe DSBRDF model, and the first to third individual lobes of the DSBRDF model. The left column shows renderings of four MERL BRDFs, the second column shows the 3-lobe DSBRDF model fit to the MERL BRDF data, the third, fourth, and fifth column show the first, second, and third DSBRDF lobe, respectively. .......................... 29

4.4 Comparison of the DSBRDF model to the non-parametric bivariate model [44] and Cook-Torrance [9]. This figure shows the log-space RMSE of fitting Lambertian and 1 lobe of Cook-Torrance (10 free parameters), Lambertian and 3 lobes of Cook-Torrance (24 free parameters), the DSBRDF using only 7 learned bases (13 free parameters), the full DSBRDF model (42 free parameters), and the non-parametric bivariate BRDF (24,300 parameters). The figure demonstrates that the DSBRDF model accurately captures real-world reflectance functions with a low-order parameterization. .......................... 34

4.5 Reflectance prior overlaid onto a 2D slice of the DSBRDF space. The BRDFs of the MERL database are visualized by the first two eigenfunction coefficients in the DSBRDF model. The ellipses of the Gaussian mixture are shown. We can see that the mixture of Gaussians is a good model for this distribution. .......................... 36
4.6 Empirical support for the novel BRDF chromaticity prior. The left two rows show the lobe chromaticity values fit to four example BRDFs from the MERL database. Solid dots represent the chromaticity values. The dashed line is Eq. 4.16 as $\alpha$ is varied. The right plot shows the distribution of the angles between chromaticity hue vectors for each unique pair of lobes among all MERL BRDFs. We then fit a von Mises distribution to this (cyan curve). The strong correlation of lobe chromaticity hue vectors shows that the novel BRDF chromaticity prior effectively constrains the BRDF color to reflect that of real-world BRDFs.

5.1 Quantitative evaluation of single point light experiments and select results. Subfigure 5.1a shows the log-space RMSE (Eq. 4.13) of recovered BRDFs from 500 single point light experiments. Each BRDF from the MERL database was illuminated by a point light at 0, 30, 60, 90, and 120 degrees from the viewing direction. Subfigure 5.1b highlights several example results to illustrate how certain log-space RMSE values correspond to perceptual accuracy. The BRDFs on the right are visualized by rendering a sphere under several point light directions.

5.2 Quantitative evaluation of point light direction estimation. This figure shows the distribution of point light direction error in degrees. As shown, all estimates are within a single degree of the ground-truth lighting direction.

6.1 Multimaterial estimation results for three real scenes. Each row, from left to right, shows the input image, a synthesized image of the scene using estimated reflectance and light source, the material segmentation result, a relit image of the object, and a ground truth image of the relighting result for each scene. The results demonstrate that the method successfully recovers the reflectance for complex scenes, for instance the gold paint on the mask, except when interreflections and shadowing are prevalent in the scene.

6.2 The reflectance and spatial extents of the five materials in the mask scene. Each column represents a different material; the top row shows the geometry of the mask rendered with that material and the bottom row shows the material assignment $m_{k,p}$.

7.1 Gradient Magnitude Histogram of Natural Illumination. Subfigure 7.1a shows an example natural illumination map. Subfigure 7.1b shows the gradient magnitude histogram of this natural illumination map with a hyper-Laplacian distribution overlaid. We can see that the hyper-Laplacian distribution is a good model for the gradient magnitudes.

7.2 The BRDF always causes an increase in entropy of the reflected radiance. The most specular materials (e.g., nickel) cause the least increase while the most diffuse (e.g., green-latex) cause the greatest increase. Only a perfect mirror (e.g., chrome), will not increase entropy.

LIST OF FIGURES
7.3 Evaluation of synthetic experiments under natural illumination. The bottom plot shows the log-space RMSE for 100 MERL BRDFs under four natural illumination environments. Highlighted are eight results from that plot shown at the top. This plot demonstrates the ability of the model to successfully infer reflectance and illumination in a variety of illumination environments for many different materials.

7.4 Quantitative evaluation of natural illumination estimation versus BRDF specularity. This figure shows the RMS error of the natural illumination estimates in blue for 100 MERL BRDFs in four illumination environments. In red is the “specularity” of the BRDF, as measured by the $\gamma$ value of the most specular lobe. This figure illustrates the general trend that high specularity BRDFs enable more accurate illumination estimation.

7.5 Predicting the appearance of materials with recovered illumination. We use the recovered illumination map to predict the appearance of materials with lower frequency reflectance. The top row shows the ground truth and the bottom row shows the predicted appearances. The input image for each subfigure is the top-left image. These results demonstrate the ability to accurately predict object appearance with the recovered illumination map.

7.6 Predicting object appearance from different views. By using the recovered illumination map and reflectance estimate, we can show the object as it would appear from different viewing directions. The first row of each subfigure contains ground truth renderings for various BRDFs, using the measured BRDF values and true environment maps, and the second row contains the predicted appearance using the recovered reflectance and illumination from our method.

7.7 Relighting comparisons. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

7.8 Relighting comparisons continued. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

7.9 Relighting comparisons continued. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.
7.10 Relighting comparisons continued. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

7.11 Results on “lobby” and “spiralStairs” illumination environments. Each row shows the recovered reflectance and illumination of a different object with the ground-truth illumination for comparison. Note how features of the reflectance function are accurately captured: the “apple”, “bear”, “horse”, and “milk” objects are shiny and the recovered reflectance function has sharp specular highlights; the “tree” object has a softer glossy reflection that is captured in the reflectance estimate.

7.12 Results on “garden”, “main”, and “picnic” illumination environments. Each row shows the recovered illumination of a different object with the ground-truth illumination for comparison. This illustrates that our algorithm is able to capture important features of the illumination environment. For example, in the “picnic” scene, the method captures the sun behind a building from the “apple” object.

7.13 Comparison to Romeiro and Zickler [45]. We compare our results qualitatively on a data set of four real spheres created by [45]. The top row shows the input image, the second row shows the predicted appearance from a novel viewpoint using the BRDF recovered using [45], the third row shows the ground truth appearance of the object from the novel viewpoint, the fourth row shows the predicted image from the novel viewpoint using the recovered reflectance function using our method and the ground truth illumination map, and the fifth row shows our result after manual color correction. As shown, our method in general recovers a more accurate BRDF.

8.1 Example geometric bases. This figure visualizes the geometric bases $g$ created by computing the eigenvectors of a similarity matrix of the triangles in the scene. By taking a linear combination of these bases, we can represent a large variety of segmentations of the geometry surface.

8.2 Evaluation of the importance of considering global light transport. Left: An image of a simple scene consisting of a sphere and a quad. Middle: Our algorithm understands interreflection and therefore correctly estimates the reflectance of the quad. Right: When interreflection is not modeled, the algorithm cannot correctly infer reflectance in this scene.

8.3 Three example scenes of our synthetic dataset. The first four columns show the input images. The fifth column shows the illumination map used to render the images. The sixth column shows the ground truth geometry used to render the images. The seventh column shows the geometry after we artificially perturb it by adding random noise to the vertices and then smoothing. This process is repeated several times to make the geometric noise more low-frequency.
8.4 Two Example Geometry Refinement Results. In this figure we show two results of geometry refinement on a synthetic scene. The first column shows the input image, the second column shows the predicted appearance of the scene before geometry estimation, the third column shows the predicted appearance of the scene after geometry estimation, the fourth column shows a difference image between the first column and the second column, and the fifth column shows a difference image between the first column and the third column. As shown, the geometry refinement process improves the rough initial geometry.

8.5 Quantitative evaluation of novel synthetic dataset. We have constructed a set of synthetic scenes that consist of three different “blob” meshes rendered with three different reflectance functions in three different illumination environments rendered from four viewpoints. We run our radiometric scene decomposition algorithm on each scene. Subfigure 8.5a on the left shows a table of log-space root-mean-square error of the inferred reflectance for each scene and on the right shows the three example scenes rendered with the estimated reflectance, illumination, and geometry from a novel view compared to ground truth. Subfigure 8.5b on the left shows a table of the error of the refined mesh for each scene and on the right shows the difference image between the scene rendered with the estimated reflectance, illumination, and geometry and the ground truth before and after geometry estimation. We can see that the geometry estimation greatly reduces the errors on the border of the geometry.

8.6 Results on Blob Experiments Under Three Illumination Environments. For each reflectance function and mesh, we show a novel view of the scene and use our inferred reflectance, illumination, and geometry to predict the appearance of the novel view. We also quantitatively evaluate the reflectance estimates by computing the log-space root-mean-square error compared to the measured BRDF. We also quantitatively evaluate the geometry by computing the distance between each vertex in the estimated mesh to the closest point on the ground-truth geometry and average all distances. We also show the percentage the geometry error is reduced by our method in parentheses (a negative number indicates a reduction in error). The general low-frequency quality of the illumination environment eases the accurate inference of reflectance and illumination in these scenes.

8.7 Results on a synthetic scene. Subfigure 8.7a shows the input images. Subfigure 8.7b shows comparisons to ground truth. The first column shows the estimated appearance using the inferred reflectance, illumination, and segmentation of the scene, the inferred illumination, and the estimated segmentation map. The second column shows the input appearance and ground truth appearance from a novel view, ground truth illumination, and ground truth segmentation. Subfigure 8.7c shows the inferred reflectance functions visualized as a series of spheres rendered with varying point-light directions.

8.8 Results on a synthetic scene. Results are presented as in Figure 8.8.
8.9 Results on a real scene. Subfigure 8.9a shows the three input images. Subfigure 8.10b shows intermediate results of the algorithm: after reflectance and illumination estimation, after texture estimation, and after geometry estimation. It’s important to note that our method correctly infers the specular highlights rather than using the diffuse texture to model them. Subfigure 8.9c shows the predicted appearance from a novel view and geometry, illumination, and segmentation estimates. The first column shows the estimated appearance using the inferred reflectance, illumination, and segmentation of the scene, the refined geometry, the inferred illumination, and the inferred segmentation map. The second column shows the ground truth appearance from a novel view, the input geometry, the ground truth illumination. Subfigure 8.9d shows the inferred reflectance functions visualized as a series of spheres rendered with varying point-light directions. Our method correctly recognizes the bright specular highlight on the table and uses other highlights in the scene to recover a detailed illumination map.

8.10 Results on another real scene. Results are presented as in Figure 8.9.

8.11 Results on another real scene. Results are presented as in Figure 8.9.

8.12 Results on another real scene. Results are presented as in Figure 8.9.

A.1 Compute Predicted Pixel Irradiance Gradient with Respect to Reflectance. This algorithm is a modification of the path tracing algorithm to keep track of the gradient with respect to reflectance.
The phrase, “a picture is worth a thousand words,” is often used to emphasize the wealth of information encoded into an image. While much of this information (e.g., the identities of people in an image, the type and number of objects in an image, etc.) is readily inferred by humans, fully understanding an image is still extremely difficult for computers. One important set of information encoded into images are radiometric scene properties—the properties of a scene related to light. Each pixel in an image indicates the amount of light received by the camera after being reflected, transmitted, or emitted by objects in a scene. It follows that we can learn about the objects of the scene and the scene itself through the image by thinking about the interaction between light and geometry in a scene.

The appearance of objects in an image is primarily due to three factors: the geometry of the scene, the reflectance of the surfaces, and the incident illumination of the scene. Recovering these hidden properties of scenes can give us a deep understanding of a scene. For example, the reflectance of a surface can give a hint at the material properties of that surface. In this thesis, we address the question of how to recover complex, spatially-varying reflectance functions and natural illumination in real scenes from one or more images with known or approximately-known geometry.

Recovering latent radiometric properties from images is difficult because of the severe underdetermined nature of the problem (i.e., there are many potential combinations of reflectance, light, and geometry that would produce identical input images) combined with
the overwhelming dimensionality of the problem. In the real world, reflectance functions are complex, requiring many parameters to accurately model. An important aspect of solving this problem is to create a compact mathematical model to express a wide range of surface reflectance. We must also carefully model scene illumination, which typically exhibits complex behavior as well. Prior work has often simply assumed the light incident to a scene is made up of one or more infinitely-distant point lights. This assumption, however, rarely holds up in practice as not only are scenes illuminated by every possible direction, they are also illuminated by other objects interreflecting one another. To accurately infer reflectance and illumination of real-world scenes, we must account for the real-world behavior of reflectance and illumination.

In this work, we develop a mathematical framework for the inference of complex, spatially-varying reflectance and natural illumination in real-world scenes. We use a Bayesian approach, where the radiometric properties (i.e., reflectance and illumination) to be inferred are modeled as random variables. We can then apply statistical priors to model how reflectance and illumination often exist in the real world to help combat the ambiguities created through the image formation process. We use our framework to infer the reflectance and illumination in a variety of scenes, ultimately using it in unrestricted real-world scenes. We show that the framework is capable of recovering complex reflectance and natural illumination in the real world.
Chapter 1: Introduction

Images captured by a camera are the result of physical interactions between surfaces and light sources. Light begins by emission from a source and is reflected, absorbed, or transmitted between surfaces in a scene before potentially entering the aperture of a camera and striking the film or charge-coupled device (CCD). The CCD then tells us how much light, for three different color channels, entered the camera. The amount of light and the path the light takes from emission from a source to absorption in the camera depends on the geometry of the scene as well as the radiometric properties of each surface in the scene.

When a ray of light is reflected or transmitted, it encodes information about the surface and the illumination environment into the resulting light ray. The reflected light ray will be attenuated in different spectra and sent along a unique distribution of directions depending on the reflectance properties of the surface. This process imprints information about the surface onto the outgoing light ray, making it possible to infer information about the surface (both the reflectance of and the shape of the surface) and about the light source by carefully examining the pixel values of the image.

Decoding the radiometric information of the scene imprinted onto the image is an important problem to solve. The reflectance of an object can give us important clues about what the object is made of. For example, an object made of gold has a distinctive appearance: it is a golden-yellow color, and it reflects light in an almost mirror-like way. If we could identify these qualities, we could narrow down the potential set of materials of that object. The illumination of a scene can also give us important information. If we knew that a scene was brightly lit from a single direction with a particular spectral signature, we
might conclude that the sun is illuminating the scene and that the scene, therefore, takes place outdoors. We could even take this a step further and draw conclusions about where in the world the picture was taken based on the angle of the sun and the current time. This rich information that we can extract from the radiometric properties of an image indicate that this is a important problem to solve.

The problem of recovering radiometric and geometric information from images has received considerable attention. One of the earliest works in the area, called shape-from-shading [21], is a method for learning the geometry of an object from an image under strict assumptions. This work has pioneered the way for more complex methods for recovering radiometric and geometric properties from a set of images. Subsequent work in this area has attempted to relax the conditions of the scene or to infer different subsets of radiometric and geometric information. Our goal is to tackle this problem in the most unrestrained setting yet. In this thesis, we propose a method to infer complex, spatially-varying reflectance, infinitely-distant natural illumination, and refine the geometry given a small set of high-dynamic range (HDR) photographs and geometry acquired from an RGB-D sensor.

A common shortfall of past methods has been the use of simplistic reflectance and illumination models. Barron and Malik have developed a method for inferring spatially-varying reflectance, spatially-varying illumination, and geometry from RGB-D [2]. The proposed method is limited, however, in that the reflectance of all surfaces is assumed to be purely diffuse. Zickler et al. develop a method to measure spatially-varying reflectance from a small set of images [56], but only consider illumination by an infinitely-distant point light. Limiting the reflectance and illumination models necessarily limits the amount of information we can recover from the scene.

In the real world, reflectance behavior and incident lighting is complex. For example,
real-world scenes have complex reflectance functions, often featuring off-specular peaks, retroreflection, and subsurface scattering effects, that can vary along a surface. Furthermore, scenes are illuminated from arbitrary directions in the scene—not only by single point light sources but by distant surfaces in the scene (e.g., the sky, the walls, or the ground). There are also significant radiometric effects that occur between objects like shadowing and interreflection that can thwart inference algorithms that don’t account for these effects.

To extract the maximum possible information from images, we must fully model the real-world behavior of surfaces and light. In other words, we must model complex real-world reflectance functions and natural illumination. We cannot rely on simplifying assumption like assuming surfaces have only Lambertian reflectance. This would prevent us from learning about the complex specular behavior of a surface which could potentially reveal that a surface is made of metal rather than paint or fabric. Without modeling complex reflectance and illumination behavior, we cannot fully understand scenes on a radiometric level.

In this thesis, we build a method for the inference of complex, real-world reflectance and natural illumination and the refining of geometry from a set of one or more images and rough or exact initial geometry. We develop a reflectance model and priors, illumination priors, spatial segmentation models and priors, and geometry refinement techniques to effectively solve the problem. We demonstrate our inference framework on a set of synthetic and real scenes and give qualitative and quantitative evaluation of our method.

1.1 Contributions

Our first task is to develop a sophisticated reflectance model that is capable of modeling real-world reflectance. This is crucial for recovering the maximum amount of information from images. We begin with the Directional Statistics Bidirectional Reflectance Distribution function (DSBRDF) and further enhance it to improve compactness of the representation.
For tractable reflectance inference, we must compactly encode the great variability of reflectance behavior of real-world materials. To do this, we construct a novel data-driven reflectance prior that accurately captures the variation of real-world BRDFs. We then incorporate the model and prior in a Bayesian formulation to perform inference.

We demonstrate single reflectance function inference under point-light illumination with a large synthetic dataset. We use the MERL BRDF dataset [32] to synthesize spheres with each of the one hundred BRDFs under 5 different point light directions. We give quantitative and qualitative results for both recovered reflectance accuracy and illumination direction accuracy.

We build on this work by introducing a spatial segmentation model that handles multiple different reflectance functions in a single scene. This works by assuming the scene is made of a small number of reflectance functions and modeling which of the reflectance functions are present at each pixel location. Rather than simply using a Markov Random Field (MRF) to model this spatial segmentation, we develop a new model that allows for greater spatial contiguity for material regions. We again use a Bayesian framework to infer the reflectance, illumination, and segmentation map.

We demonstrate our multi-material reflectance and point-light illumination inference framework on a novel dataset of real-world images in a dark room setting. We give qualitative comparisons to a novel view of the object to show that we are correctly inferring reflectance and spatial segmentation.

So far, we have assumed that the scenes are illuminated with infinitely-distant point lights. As we have discussed, this rarely holds true in practice. Especially in indoor scenes, objects are potentially illuminated by windows, lights, a television, or even indirectly from the sun shining brightly on a nearby wall. Furthermore, these illuminants can be any size or

**Chapter 1: Introduction**
color and they can be arbitrarily directed. Modeling this variability is key to decomposing scenes radiometrically in the real world.

To address the problem of natural illumination, we model the incident illumination as a panoramic image of the scene. This gives us the flexibility necessary to model a variety of indoor and outdoor scenes. With this flexibility, however, comes additional ambiguity. We therefore propose a set of priors on the natural illumination image to constrain the variability of this representation. We note that, a panoramic image of the scene is first an image and we may therefore place natural image statistic priors on the illumination image. We also note that the reflectance function increases the entropy of the reflected radiance. To compensate for this, we must constrain the illumination image to have minimum entropy. We use a Bayesian framework to tie the pieces together.

We demonstrate reflectance and natural illumination inference from a single image on a large synthetic dataset and a novel real-world dataset. Our synthetic dataset is a set of spheres rendered with each of the one hundred MERL BRDFs in four different natural illumination environments from Paul Debevec’s light probe gallery [12]. We give quantitative evaluation of the reflectance estimates and the illumination estimates. Our novel real-world dataset consists of 6 objects in 5 novel natural illumination environments with ground truth illumination. We show qualitative comparisons of the objects rendered under different illumination environments.

The next hurdle we must overcome is the inference of multiple complex real-world reflectance functions in scenes with complex geometry (i.e., multiple objects that can interreflect, concave objects, etc.) under natural illumination. Extending the problem in this way presents a significant challenge in that the interreflection and radiometric interaction between objects begins to significantly affect the image. Modeling this interreflection be-
havior is crucial to learning about specular surfaces because the primary indicator that a surface is specular is that it reflects other objects in the scene.

Our first task to solve this problem is to extend our image-based spatial segmentation model to the three-dimensional geometry of the scene to enable the use of multiple views. We do this by placing the segmentation over the triangles of the geometry rather than the pixels of the image. A consequence of this is that the image likelihood is no longer differentiable with respect to the segmentation parameters, which is necessary for gradient-based optimizers. Therefore, we must reduce the dimensionality of the segmentation space so that we may use a gradient-free optimizer to compute the maximum a posteriori estimate for the segmentation map.

Although most real-world scenes have a small number of fundamental reflectance functions, the diffuse texture of an object can vary greatly along the surface. For example, imagine a wooden desk—the whole desk may have a light sheen from a coating of lacquer that is constant along the surface, but the wood grain will impart a change in the absorption of light of certain areas of the surface. To extract the most radiometric information possible from a scene, we must consider this diffuse texture variation. We model this by allowing the diffuse lobe color to finely vary within material regions using a high-resolution texture map. This allows us to capture important texture detail in the scene.

In the final chapter, we assume that the geometry is input from an RGB-D sensor. RGB-D sensors typically have several types of noise that interfere with reflectance and illumination inference. For example, the limited resolution of commodity depth sensors means that fine geometric detail is typically lost. Furthermore, large-scale stitching errors can be introduced when multiple depth images are fused to form a single geometric model. We address large-scale stitching errors by developing a geometric refinement model that
stretches or shrinks large regions of geometry to best match the input images.

Finally, we address the problem of the indirect illumination (e.g., interreflection between objects, shadowing, etc.). Ignoring these effects is a considerable problem when inferring complex reflectance because an object reflecting other objects in the scene can be the biggest indicator of its reflectance behavior. It is important to model shadows because they can provide important clues about the illumination of the scene. When natural illumination is considered, however, it can be difficult to determine exactly where shadows should be unless the indirect illumination is fully simulated. We address these problems by fully simulating the indirect illumination by using an unbiased rendering algorithm and by developing an algorithm for computing error function gradients by modifying a path tracing algorithm.

We demonstrate our method on both synthetic and real scenes. We develop a novel synthetic dataset consisting of several scenes with various objects and rendered with PBRT [40] for physically-accurate rendering. We also capture a novel dataset of real-world scenes with ground truth illumination using an HDR camera and a Kinect sensor. The scenes span a range of different objects with varying geometry and reflectance functions that would typically occur in real-world scenes.

In this thesis, we develop a method for inferring multiple complex reflectance functions, natural illumination, and refine geometry from a small set of images. This is achieved through the following contributions:

- A reflectance prior that enables tractable reflectance inference.
- A spatial segmentation model in the image domain to model objects with multiple
reflectance functions from a single image in real-world scenes.

- A spatial segmentation model in the 3D geometric domain to model scenes with multiple reflectance functions.

- A texture estimation method for recovering fine diffuse texture in a scene.

- A natural illumination model and priors to enable illumination inference

- A method for handling indirect illumination during reflectance and illumination inference.

- A Bayesian inference framework that allows us to incorporate our priors and solve the problem by computing an estimate of the maximum of the posterior distribution (i.e., the distribution of the reflectance, texture, spatial segmentation, illumination, and geometry given the input images).

We show that these contributions allow us to decompose real scenes into their radiometric components.
Chapter 2: Background

Our goal is to make inferences about physical quantities of the world based on radiance values measured by one or more cameras. In this chapter we will explain the notions of radiance, reflectance, and illumination, and discuss previous work to infer radiometric quantities from images.

2.1 Reflectance

For the purposes of this work, we assume that geometric optics are sufficient for representing light transport (i.e., light can be thought of as a ray through space with an associated spectrum distribution). This will let us ignore quantum effects such as diffraction.

A fundamental part of what determines the appearance of a scene is the reflectance function of each surface. In general, there are several things that can happen when light strikes a surface. One possibility is that the light is absorbed. Another possibility is that the light is reflected away from the surface or transmitted through the surface. If light is transmitted into the surface, it can pass straight through (as it would in a piece of glass) or be scattered within the object, possibly emerging from a different point on the surface.

We can quantify this behavior by defining a function,

\[ f_r(x_i, x_o, \omega_i, \omega_o) = \frac{L_o(x_o, \omega_o)}{(N_{x_i} \cdot \omega_i) L_i(x_i, \omega_i)}, \]

that gives the fraction of light leaving the surface \( L_o(x_o, \omega_o) \) at surface point \( x_o \) in direction \( \omega_o \) when that surface receives \( L_i(x_i, \omega_i) \) light at surface point \( x_i \) in direction \( \omega_i \), where \( N_{x_i} \) is the surface normal at surface point \( x_i \). This function is called the Bidirectional Scattering
Distribution Function (BSDF) [3]. If we shined a light on a surface at a particular point in a particular direction and recorded the amount of light that is reflected at another point in another direction, for all possible pairs of points and directions, we would gain a complete description of how that surface reflects light.

In practice, this is an extremely difficult task because of the high dimensionality of the function. There are two two-dimensional position parameters and two two-dimensional angular parameters for a total of eight dimensions. It is easier in practice to assume that a light ray is either immediately absorbed or reflected. This will eliminate one of the positional parameters of the model.

Assuming that a light ray is immediately absorbed or reflected gives us a simpler model called the Bidirectional Reflectance Distribution Function (BRDF). The BRDF is written more simply,

\[
f_r(x, \omega_i, \omega_o) = \frac{L_o(x, \omega_o)}{(N_x \cdot \omega_i)L_i(x, \omega_i)},
\]

and only considers light that is reflected from the surface (i.e., it ignores transmission through the surface). Despite this simplification, the BRDF is still a powerful way to represent reflectance.

Although the BRDF is a simpler model, it is still a four-dimensional function, making it hard to represent in practice. It turns out that many of the additional degrees of freedom can be redundant for many types of materials. An isotropic BRDF is one that is invariant when the incident direction \(\omega_i\) and the exitant direction \(\omega_o\) are rotated an equal amount by the surface normal. This restriction reduces the BRDF to only three dimensions, making it significantly easier to model.
2.1.1 BRDF Models

There are a number of BRDF models invented for computer graphics and computer vision. Arguably the simplest BRDF is a Lambertian surface [27], which appears the same intensity regardless of the viewing direction, therefore making it invariant to $\omega_o$:

$$f_r(\omega_i, \omega_o) = \frac{\rho}{\pi}, \tag{2.3}$$

where $\rho$ is the albedo of the surface that determines its color. This simple BRDF can only model diffuse surfaces.

Separate models have been developed to represent surfaces with specular behavior. Cook-Torrance [9] is a microfacet BRDF model, meaning that it assumes the surface is composed of microscopic facets whose orientation follows a particular distribution. The analytical formula for the model is derived by aggregating the behavior over the distribution of facets and accounting for geometric effects between facets. Similar models have been derived using this microfacet approach, such as the Oren-Nayar model [38].

It’s also possible to combine two or more different BRDF models by adding them together. In this scenario, each term is referred to as a BRDF lobe. Lobes can be two different BRDF models entirely (e.g., one lobe for diffuse reflectance and one lobe for specular reflectance) or each lobe can use the same model but with different parameters. Using the multiple lobes of a single reflectance model is an easy way to create a more expressive model.

Many other BRDF models have been developed without an underlying physical theory. Instead, they simply provide a visually plausible reflectance model. One of the first examples

Chapter 2: Background
Figure 2.1: Diagram of the vectors used to define the BRDF. The left column shows an example outgoing direction $\omega_o$ and incident direction $\omega_i$ on the hemisphere. The right column shows a reparameterization of the BRDF using the half vector $h$ [46]. Here $n$ is the surface normal and $t$ is the surface tangent vector. Representing BRDFs using $(\theta_d, \phi_d)$ and $(\theta_h, \phi_h)$ is significantly easier than $(\theta_i, \phi_i)$ and $(\theta_o, \phi_o)$ for many real-world materials.

is the Blinn-Phong specular model [4]. This model writes the reflected radiance,

$$f_r(\omega_i, \omega_o) = (H \cdot N)^\alpha,$$

(2.4)

where $H = \frac{\omega_i + \omega_o}{\|\omega_i + \omega_o\|}$ is called the halfway vector, $N$ is the surface normal, and $\alpha$ controls the degree of specularity.

The halfway vector is a convenient way to formulate BRDFs. A popular reparameterization lets us write the BRDF as a set of angles relative to the halfway vector rather than relative to the incident and exitant directions [46]. We can use this reparameterization to write the BRDF,

$$f_r(\theta_h, \phi_h, \theta_d, \phi_d),$$

(2.5)

where $\theta_h$ is the angle between the halfway vector and the surface normal, $\phi_h$ is the azimuth angle of the halfway vector, $\theta_d$ is the angle between the halfway vector and the incident direction $\omega_i$, and $\phi_d$ is the azimuth angle of the incident direction with respect to the halfway...
vector. Figure 2.1 visualizes this reparameterization.

This parameterization has a number of benefits. First, isotropic BRDFs can be represented by simply dropping the dependency on $\phi_h$. Next, a second type of isotropy that is common in real materials can be represented by dropping the dependency on $\phi_d$. This is the same as stating that the BRDF has the same value when the incident direction and exitant direction are rotated about the halfway vector. These modifications allow us to write the BRDF as a two-dimensional function (dependent only on $\theta_h$ and $\theta_d$) that is dramatically simpler to model.

2.1.2 Measuring BRDFs

The reflectance function of real-world surfaces is more complex than can typically be modeled by simple parametric BRDF models like Lambert’s Law and Cook-Torrance. Advanced models of reflectance have been developed that can capture real-world reflectance behavior more accurately.

Properly measuring a real-world reflectance function requires taking many different angular samples of both the incident and exitant directions of light. Numerous methods have been developed to do this. Murray-Coleman and Smith [34] develop a gonioreflectometer to easily sample many incident and exitant light directions to acquire BRDF data. Ward [51] develop a curved mirror apparatus to easily capture multiple incident light directions at once. This apparatus is used to measure real-world reflectance functions and fit them to a parametric anisotropic BRDF model. Later, Dana [10] introduce an improved method that uses a concave parabolic mirror and enables the automatic changing of illumination direction. Marschner et al. [31] move the camera around a curved object illuminated by a point light to measure the reflectance function. This method is primarily designed for BRDF measurement in laboratory settings. Han and Perlin [18] develop a tapered kaleidoscope for
the capture of surfaces from multiple views with a single image.

Dana et al. [11] extend the concept of a BRDF by including the spatial domain, giving rise to the bidirectional texture function (BTF). They use a robotic arm to automatically rotate material samples to sufficiently cover the angular domain of the reflectance function. They use this apparatus to recover measured reflectance data from over 60 different materials with over 200 different combinations of viewing and illumination directions.

Matusik et al. [32, 33] introduce an image-based BRDF capture apparatus that relies on curved surfaces to provide many surface normals so that only the light source must be moved. They use their BRDF capture apparatus to recover densely sampled BRDFs of one hundred real-world materials. This has been an extremely important contribution to the field as it allows us to analyze the accuracy of many BRDF models and even develop new models from the data. These developments have made it possible to capture a large number of real-world reflectance functions from an image easily.

2.2 Illumination

The illumination incident to a surface $L_i(x, \omega_i)$ can take on many forms. In real scenes, surfaces are illuminated from every direction on the hemisphere, from either surfaces that emit light or from other surfaces reflecting light. This complexity is typically reduced in most computer vision applications in a number of ways. Very often, it is assumed that all surfaces are directly illumination by a small emitting source, such as a point light. Less often, methods account for shadowing. Even more rare are methods that account for full illumination complexity (i.e., modeling unlimited indirect light bounces).

In this work, we model the light source of a scene to be an infinitely-distant sphere surrounding the scene emitting light from every direction. These assumptions generally hold on natural scenes in practice when light sources are sufficiently far from imaged part.
of the scene.
Chapter 3: Related Work

Decomposing a scene into its radiometric ingredients has been an important problem through the history of computer vision. In this chapter, we discuss many of the method for inferring one or more radiometric property of a scene (i.e., reflectance, illumination, and geometry) from one or more images.

3.1 Early Radiometric Scene Decomposition

The problem of inferring reflectance and other scene properties from an image has received considerable attention in computer vision. Photometric stereo [53] and shape-from-shading [21] are two early algorithms that attempt to solve an instance of this problem. These two problems focused primarily on the recover of geometry but they can be seen as radiometric scene decomposition where some quantities (e.g., reflectance or illumination) are assumed to be known.

Photometric stereo is the problem of estimating the diffuse albedo and surface normals of a scene from a set of images from a single view under changing point-light illumination. When the illumination directions are known, recovering the albedo and surface normals can be accomplished by a simple matrix inversion. The problem is only convex, however, when the reflectance is assumed to be Lambertian and the illumination is assumed to be infinitely-distant point lights.

Shape-from-shading methods attempt to recover the geometry of an object based on variation in shading in the input image. This is typically done by finding the surface normals that produce an image that closely matches the input. Unfortunately, for simple
surfaces under point-light illumination, there are many surface normals that produce one particular image pixel value. Horn [21] attempts to constrain this ambiguity by using a spatial smoothness prior on the surface normals and an occluding boundary prior on the edge of the object. Many steady improvements have been made to this basic algorithm, including the enforcing of integrable surface estimates by Frankot and Chellappa [16], estimating the lighting direction and albedo [55], and extending it to surfaces that are piece-wise smooth [29].

These early methods, however, impose simplistic assumptions about objects (e.g., that they exhibit ideal Lambertian reflectance) that inherently limit their applicability. Although many of these restrictions have been relaxed as the body of work has grown, there is still a need for methods that function effectively in the wild.

Other methods are primarily interested in recovering reflectance and illumination and assume that the geometry is known. Ikeuchi and Sato develop a method for recovering reflectance properties and illumination direction from a single image of a uniform object with known geometry [23]. They model reflectance with a combination of Lambertian plus Torrance-Sparrow that allows them to represent and infer a more rich set of object reflectance than the original photometric stereo and shape-from-shading methods. This pioneering work opened the door for many more methods that focus on the recovery reflectance and illumination information from a small number of images.

Later, Sato et al. [47] developed a method to recover spatially-varying reflectance properties from the full geometry and image intensity from multiple views. They first acquire a series of range images of the object and align and merge them to form a full geometric model of the object. They then use the geometry, known light source direction, and image observations from multiple views to densely recover the diffuse albedo, specular component,
and refine the surface normals. This work is somewhat limited, however, in that it relies on having at least three observations of each triangle in the mesh and its assumption of known point-light illumination. In real scenes, objects may occlude one another, limiting the number of observations across the surface. In addition, the natural illumination of real scenes makes this type of approach infeasible in general.

3.2 Intrinsic Images

Intrinsic image decomposition in general are methods to extract per-pixel latent information from images. In the context of reflectance and illumination recovery, intrinsic images typically refer to a reflectance and a shading image that, when appropriately combined, will form the original input image. The reflectance image typically contains the diffuse albedo for each pixel in the scene, while the shading image represents the effects of both the illumination and geometry. If properly estimated, the intrinsic images can be used to synthetically insert objects into a scene or accurately segment an image based on reflectance. Because the goal is to retrieve two images from a single input image, additional constraints must be applied to solve this underdetermined problem.

One of the first works that attempted to recover intrinsic reflectance and shading images from an image is called Retinex from Land and McCann [28]. The method works by assuming that the reflectance image will be characterized by sharp jumps whereas the shading image will have smooth variations from curved surfaces. They use this assumption to filter the input image into a reflectance and shading image.

Tappen et al. classify each image gradient as being caused by either the reflectance image or shading image [49]. They use color and gray-scale image features as input to a classifier to determine whether an image gradient belongs to the reflectance or shading image. They then propagate this information to areas where the image features are ambiguous.
Chen and Koltun infer intrinsic images by breaking the shading image into a direct irradiance and indirect irradiance image so that they may use stronger priors on each [8]. They also leverage depth information to place a prior on the direct irradiance image. This method is interesting because it attempts to estimate the indirect illumination contribution. As we show in this thesis, when we estimate reflectance and illumination with known geometry of a scene, we can simulate the indirect illumination.

By combining the effects of illumination and geometry into a single image, intrinsic image methods limit the amount of information they can recover from a scene. They also typically only assume that surfaces exhibit Lambertian reflectance. Rather, we are interested in separately recovering each variable (i.e., reflectance, illumination, and geometry) using expressive models that don’t restrict our ability to model real-world scenes.

3.3 State of the Art

Many authors have adopted sophisticated reflectance models to recover object reflectance outside the laboratory. Zickler et al. develop a method to measure spatially-varying reflectance from a small set of images [56]. This work uses a non-parametric BRDF model but overcomes the need for many reflectance samples by “sharing” observations between the spatial and angular domain. Although these methods use more complex models for reflectance, expressive illumination models are also necessary for the real world.

Some past work has explored sophisticated illumination models for real-world radiometric scene decomposition. Marschner and Greenberg estimate a lighting distribution using a sum of basis functions from a single image, but assume Lambertian reflectance [30]. Nishino et al. recover a Lambertian plus Torrance-Sparrow reflectance model under an unknown lighting distribution from a small set of images [37]. Hara et al. derive a spherical Torrance-Sparrow reflection model to jointly estimate multiple point sources and the re-
flectance parameters through mixture modeling on a unit sphere [20]. The main limitation of this work is the use of polarization filters to manually separate specular highlights. Hara and Nishino model the spatial variation of just the specular reflection by superimposing a radial basis function network and by estimating the parameters of a spherical Torrance-Sparrow model at each node using variational inference [19]. Chandraker and Ramamoorthi study the conditions under which a 1D slice of a BRDF is uniquely estimable [7]. By contrast, we estimate full BRDFs by utilizing strong priors to constrain the solution space. For recovering reflectance and illumination in the wild, we need to represent both reflectance and illumination in a general way.

Ramamoorthi and Hanrahan introduced a signal-processing framework for unconstrained reflectance and illumination environments [42]. This work analyzed the theoretical ambiguities that exist between the reflectance and illumination by representing both with spherical harmonics and enumerated the situations under which they cannot be separated. A major practical concern of this work, however, is the large number of input images required and the limited expressiveness of generic bases like spherical harmonics that require excessively many coefficients to express high-frequency illumination and reflectance (e.g., strong directional light in a scene and specularities, respectively).

Barron and Malik construct a complete framework for joint spatially-varying reflectance, spatially-varying illumination, and geometry estimation [2]. They use a large-scale energy minimization of reflectance, illumination, and geometry with several unique priors to solve the problem in the real world. Although the method is able to estimate reflectance, illumination, and refine geometry, it has one important limitation: the assumption of pure Lambertian reflectance. While many scene may have a number of diffuse objects, the diffuse behavior of these objects themselves can be very complex (e.g., exhibiting retroreflection...
and even grazing angle specular behavior). Solving this problem in the real world requires dealing with all kinds of real objects with real, complex reflectance behavior.

An interesting and unique aspect of the work by Barron and Malik is the use of a spatially-varying illumination model. This makes the method capable of modeling nearby light sources. Additionally, it enables the method to gracefully handle strong indirect illumination. This is, however, at the cost of a large amount of additional variables to infer. On the other hand, our method simulates the indirect lighting bounces, obviating the need for such an expressive illumination model.

Yu et al. refine shape from RGB-D images by simultaneously estimating albedo and illumination [54]. They explicitly seek to refine the surface normals from noisy RGB-D sensors. This method assume that the objects exhibit Lambertian reflectance. While this is a convenient assumption, it eliminates much of the rich radiometric information we can infer from RGB-D images—for example, the reflectance behavior of an object can give us a clue about the material of that object, and the appearance of the specular highlights can tell us about the illumination environment.

Oxholm and Nishino recover geometry and reflectance from a small set of views in the real world with known illumination [39]. They combine surface normal likelihood distributions for each view with several shape priors to successfully refine geometry from an initial convex hull. Because our initial geometry comes from an RGB-D sensor rather than a convex hull, we must tailor our geometry refinement process appropriately. Furthermore, the method only handles objects with a single reflectance function.

Goldman et al. estimate geometry and spatially-varying BRDFs from a set of images under known point-light illumination [17]. Similar to our approach, the authors note that objects are typically composed of a small number of fundamental materials. They estimate
each fundamental material plus a material weight map that controls the mixture of the fundamental materials at each pixel. We take a similar approach, but with some important differences: the segmentation of the scene is imposed over the 3D geometry rather than in the image space to allow for multiple views of a scene; the segmentation is specified by a set of geometric bases which encourages contiguity among material segments. Furthermore, we don’t allow a mixture of fundamental materials (i.e., a surface point only exhibits one of \( K \) base materials). Rather, we argue that a linear combination of fundamental materials is rare but variation in the diffuse texture is common.

Romeiro and Zickler introduced a practical method to estimate real-world reflectance functions under known natural illumination from a single image [44]. This method is able to effectively tackle real-world scenes by using a non-parametric data-driven BRDF model. Later, Romeiro and Zickler extended this work to unknown natural illumination [45]. The main drawbacks of their method are that the illumination environment is not inferred but instead marginalized out and that the reflectance and illumination are actually monochrome (color is added later as a post-processing step). The illumination environment itself conveys a wealth of information as it is the scene itself surrounding the object; inferring it jointly with reflectance is therefore useful.

Dong et al. estimate spatially-varying reflectance and natural illumination from a video of a single object with known geometry [13]. Their method works by examining the change in appearance of an object as it rotates. Unfortunately, the method requires many input images and is not robust to noisy geometry from RGB-D sensors. Furthermore, the method only considers single objects rather than scenes where objects may interreflect and shadow one another. Rather, our method uses only approximately-known geometry from an RGB-D sensor.
Karsch et al. estimate reflectance, illumination, and geometry from a single image [26]. Their method uses a pipeline of methods to estimate each variable one at a time: first geometry, then reflectance, then illumination. This is limiting because each variable is estimated in isolation (i.e., there is no feedback between variables). Our method models the crosstalk between each of these variables. This is critical because the appearance of a scene is so interdependent on each of the variables. For example, the change in appearance along the surface of an object can be due to a change in surface geometry, an object reflecting a natural illumination environment, or a change in the reflectance function along the surface (i.e., texture). Isolating each variable by itself will render an algorithm unable to decide which aforementioned situation is really occurring.

Unlike previous work, we estimate complex spatially-varying reflectance and natural illumination from geometrically complex scenes (i.e., scenes prone to interreflection and other indirect illumination effects) from known or approximately-known geometry. We use expressive but tractable reflectance and illumination models with data-driven priors to reliably estimate radiometric scene properties. We show that our framework can tackle this problem in unrestricted real-world scenes.
Chapter 4: Directional Statistics BRDF

The goal of this chapter is to derive a low-dimensional parametric BRDF model that can achieve accuracy comparable to non-parametric representations. That is, we aim to accurately express the wide variety of real-world BRDFs with an analytical model consisting of a small number of parameters. In order to make the problem tractable, we focus on isotropic BRDFs. Our approach to this challenging, long-standing problem is based on a novel perspective of a BRDF. We view the BRDF as a directional statistics distribution, a probability density function that takes in an incident light ray direction and returns a distribution of reflected light ray directions.

We believe the novel DSBRDF model has direct implications in a broad range of applications. Most important, it enables the encoding of a wide variety of real-world isotropic BRDFs with a common low-dimensional analytical form. This is in sharp contrast to previous work in which the appropriate model had to be chosen and combined among different BRDF models or the representations were left high dimensional, incurring burden on further analysis.

4.1 Isotropic BRDF

As discussed in Chapter 2, the BRDF is a four-dimensional function defined as the ratio of the reflected radiance $L_o$ in a given exitant (view) direction $\omega_o$ to the incident irradiance $(N \cdot \omega_i) L_i$ due to light from direction $\omega_i$,

$$f_r(\omega_i, \omega_o) = \frac{L_o(\omega_o)}{(N \cdot \omega_i) L_i(\omega_i)},$$  \hspace{1cm} (4.1)
where \( \mathbf{N} \) is the surface normal at the surface point of interest. The BRDF is thus a 4D real-valued function \( f_r : \Omega \times \Omega \rightarrow \mathbb{R} \), where \( \Omega \) is the upper hemisphere with its origin at the surface point and its north pole (Z axis) aligned with the surface normal. The two directions \( \omega_i \) and \( \omega_o \) can be described in spherical coordinates \( (\theta_i, \phi_i) \) and \( (\theta_o, \phi_o) \) for an explicit 4D notation of the BRDF.

Although in general the BRDF has four dimensions, the intrinsic dimensionality of many real-world materials is less than four. One class of BRDFs with a lower intrinsic dimensionality are isotropic BRDFs. These BRDFs are invariant to azimuthal rotations of the incident and exitant direction about the surface normal. This property can be written,

\[
f_r(\theta_i, \phi_i, \theta_o, \phi_o) = f_r(\theta_i, \phi_i + c, \theta_o, \phi_o + c),
\]

for any real number \( c \). Using the reparameterization of the BRDF into half and difference vector, this can be represented more simply by dropping the dependency on \( \phi_h \),

\[
f_r(\theta_h, \phi_h, \theta_d, \phi_d) = f_r(\theta_h, c, \theta_d, \phi_d),
\]

for any real number \( c \).

Romeiro et al. have experimentally shown that many isotropic BRDFs can be accurately represented in only two dimensions [44]. They found that measured BRDFs tend not to vary greatly with \( \phi_d \) which leads to a writing of the BRDF as a function of only \( \theta_h \) and \( \theta_d \): \( f_r(\theta_h, \theta_d) \). We exploit this observation by developing an analytical BRDF model as a function of only \( \theta_h \) and \( \theta_d \).
4.2 Hemispherical Exponential Power Distribution

We now derive an analytical reflection model for representing real-world isotropic BRDFs. The key observation we make is that a 2D slice of a BRDF can be viewed as a statistical distribution of reflected light rays given an incident light ray ($\theta_d$), where the reflected radiance values $f_r(\theta_h, \theta_d)$ represent the probability of incident light being scattered into that specific direction ($\theta_h, \theta_d$).

With this insight, we must choose a directional statistics distribution with a small number of parameters that is also able to model real-world reflectance functions. We can leverage the fact that BRDFs typically peak when $\theta_h = 0$ because this will cause the incident and exitant directions to become the mirror direction of each other. This suggests that we should use a center-peaked isotropic probability density function of $\theta_h$ to model the BRDF. The domain of many conventional direction statistics distributions are unsuitable.
for modeling this variation; we must explicitly model it as a probability distribution over the hemisphere.

We use a hemispherical distribution analogous to the exponential power distribution for this task [36]. This directional distribution is referred to as the hemispherical exponential power distribution (hemi-EPD):

\[ p(\theta_h) = C(\kappa, \gamma) \left( \exp \left[ \kappa \cos \gamma \theta_h \right] - 1 \right), \quad (4.4) \]

where \( \kappa \) and \( \gamma \) are the parameters of the distribution and \( C \) is the normalizing factor.

The hemi-EPD has several advantages for modeling real-world BRDFs. As shown in Figure 4.1, it can represent a wide variety of hemispherical directional distributions. We refer to \( \kappa \) as the scale parameter and \( \gamma \) as the shape parameter. The \( \kappa \) parameter controls the height of the distribution and therefore has a similar effect as an albedo parameter. The \( \gamma \) parameter controls the concentration of the reflected light direction and therefore controls the specularity of the reflectance. The hemi-EPD can model perfect Lambertian surfaces when \( \gamma = 0 \) or a perfect mirror as \( \gamma \to \infty \).

Real materials often exhibit a mixture of different reflectance behavior. For this reason, a linear combination of BRDF models are typically used in practice to represent real-world surfaces (e.g., Lambertian and Torrance-Sparrow). We can use this same concept to take a linear combination of hemi-EPDs to represent the BRDF,

\[ \varrho(\theta_d, \theta_h) = \sum_r \exp \left[ \kappa_r(\theta_d) \cos^{\gamma_r}(\theta_d) \theta_h \right] - 1, \quad (4.5) \]

where \( \kappa_r \) is the shape parameter of the \( r \)-th lobe and similarly for \( \gamma \). Here we drop the normalization constant \( C \) because real-world BRDFs do not typically integrate to one due
Figure 4.2: Synthetic spheres rendering using the DSBRDF model with parameters fit from three MERL BRDFs (nickel, specular-blue-phenolic, and orange-paint) [32]. The left column shows the spheres rendering using the measured BRDF data. The next five columns show the spheres rendered using the DSBRDF model with 1 to 5 lobes from left to right, respectively.

to absorption of the incident light. Note that we have now made the parameters $\kappa$ and $\gamma$ a function of $\theta_d$, the angle between the incident direction and the half vector. This substantially increases the expressibility of the model.

A BRDF represented with the DSBRDF model can now be thought of as a set of $(\kappa, \gamma)$-curves, one for each lobe. The $(\kappa, \gamma)$-curves control the behavior of the reflectance function across different $\theta_d$ slices of the BRDF. We choose to model these $(\kappa, \gamma)$-curves with quadratic B-splines with nine knots based on an exploratory analysis of various functional bases with leave-one-out cross validation,

$$
\kappa_r(\theta_d) = \exp \left[ \sum_i \alpha_{\kappa,r,i} B_{i,2}(\theta_d) \right],
$$

$$
\gamma_r(\theta_d) = \exp \left[ \sum_i \alpha_{\gamma,r,i} B_{i,2}(\theta_d) \right],
$$
Figure 4.3: Synthetic spheres rendered with four MERL BRDFs [32], the 3-lobe DSBRDF model, and the first to third individual lobes of the DSBRDF model. The left column shows renderings of four MERL BRDFs, the second column shows the 3-lobe DSBRDF model fit to the MERL BRDF data, the third, fourth, and fifth column show the first, second, and third DSBRDF lobe, respectively.
for the knot vector $\left(0, 0, 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{\pi}{2}\right)$, where $B_{i,2}$ is a quadratic B-spline basis function. Importantly, we model the $(\kappa, \gamma)$-curves in log-space to ensure that $\kappa$ and $\gamma$ can never be negative.

### 4.3 Modeling Color

We represent the BRDF color by introducing a chromaticity vector for each lobe. The chromaticity, $c$, modulates the color of each lobe,

$$\varrho_{\lambda}(\theta_d, \theta_h) = \sum_{r} c_{r,\lambda} \left( \exp \left[ \kappa_{r}(\theta_d) \cos^{\gamma_{r}}(\theta_d) \theta_h \right] - 1 \right),$$

where $\lambda$ is the color channel, and with the constraint that $\sum_{\lambda} c_{r,\lambda} = 1$, which ensures that the chromaticity vector does not influence the overall luminance of the lobe. This explicit modeling of color is key to more tightly modeling real-world reflectance.

### 4.4 Characterizing the Variation of Real-World BRDFs

Because $(\kappa, \gamma)$-curves give a complete description of the BRDF, we can fit them to a set of measured reflectance data (such as the MERL BRDF database [32]). Capturing the variation of these curves is the key to achieving a compact but expressive model.

We fit the DSBRDF model by numerically optimizing the B-spline parameters to the $(\kappa, \gamma)$ curves. To do this, we optimize the following objective function:

$$\sum_{\theta_h, \theta_d, \phi_d} \left( \log f_r(\theta_h, \theta_d, \phi_d) - \log \varrho(\theta_h, \theta_d) \right)^2,$$

where $f_r(\theta_h, \theta_d, \phi_d)$ is the measured BRDF data and $\varrho(\theta_h, \theta_d)$ is the DSBRDF model. We found it necessary to optimize the logarithm of the BRDF values for several reasons. For one, it helps account for the large disparity in scale between specular peaks and diffuse areas.
of the BRDF. Second, it makes the objective invariant to the overall scale of the BRDF (i.e., dark BRDFs will not be penalized in the objective function differently than bright BRDFs).

Once we fit the DSBRDF model to measured data, we can establish correspondences among the $K$ individual lobes across different $\theta_d$ by simply sorting the lobes based on the shape parameter $\gamma$. This will separate the lobes into different classes of lobes, from most specular to least specular. Later, we will analyze how each reflectance lobe class varies in the real world.

Figure 4.2 shows the DSBRDF model fit to three MERL BRDFs when the number of lobes used varies from one to five. As shown, most of the main characteristics are captured in the first lobe of the BRDF. However, capturing more subtle detail of the BRDF requires sometimes up to three or more lobes. This is most noticeable in the middle row of the figure (specular-blue-phenolic) where even two lobes is not sufficient for capturing the subtle behavior of the reflectance function especially at grazing angles (notice how the two-lobe approximation is too dark along the edges). We empirically found that using three lobes gave a good compromise between accuracy of fit and the size of the parameterization. Compared to MERL BRDFs, which are modeled with $3 \times 90 \times 90 \times 180$ samples, modeling the DSBRDF with $(\kappa, \gamma)$-curves requires only $12K$ parameters. In the rest of the work, we will fix the model to have three lobes.

Figure 4.3 shows the individual lobe behavior for the DSBRDF model. From the figure we can see that each lobe captures a different aspect of the reflectance function: the last lobe typically captures diffuse behavior, whereas the first and second lobes capture very specular and glossy characteristics. For very glossy BRDFs, however, even the last lobe is dedicated to modeling specular reflection.
The DSBRDF model can represent a continuum of real-world isotropic BRDFs with the same low-dimensional parametric form. This enables an analysis of the space of isotropic BRDFs, represented with measured data of 100 real-world BRDFs. This ability is vital to study the space and characterize it for further use in various inverse or forward problems—for instance, to extract statistical priors to constrain ill-posed inverse problems.

We analyze the variability of the \((\kappa, \gamma)\) curves by performing multivariate functional principal component analysis (FPCA) [43] on all lobes of the 100 real-world BRDFs. To properly account for the scale disparity between \(\kappa\) and \(\gamma\), we modify the inner product used to perform FPCA between \((\kappa, \gamma)\) curves by weighting the \(\kappa\) and \(\gamma\) portions of the product by the reciprocal of their mean curve squared,

\[
\langle \xi_1, \xi_2 \rangle = \sum_r \frac{1}{||\mu^{\kappa_r}||^2} \int \xi_1^{\kappa_r}(\theta_d)\xi_2^{\kappa_r}(\theta_d)d\theta_d + \frac{1}{||\mu^{\gamma_r}||^2} \int \xi_1^{\gamma_r}(\theta_d)\xi_2^{\gamma_r}(\theta_d)d\theta_d, \tag{4.10}
\]

where \(\xi = [\xi^{\kappa_r}, \xi^{\gamma_r}] \in \{1 \ldots K\}\) are the \((\kappa, \gamma)\)-curves. The principal functions (eigenfunctions) not only capture the major variations of \(\kappa\) and \(\gamma\) separately across \(\theta_d\), but also capture how \(\kappa\) and \(\gamma\) vary together.

Note that the mean function and each basis function are 36-dimensional vectors that contain the B-spline values for all reflectance lobes and \(\kappa\) and \(\gamma\). If we denote the operation of extracting the B-spline coefficients for reflectance lobe \(r\), and function \(\kappa\) with \(b(\theta_d; \kappa, r)\) (similarly for \(\gamma\)), we can express the DSBRDF parameters \(\kappa\) and \(\gamma\) as a log-linear combination of these functions:

\[
\kappa_r(\theta_d) = \exp \left[ b_\mu(\theta_d; \kappa, r) + \sum_i \psi_i b_i(\theta_d; \kappa, r) \right], \tag{4.11}
\]

\[
\gamma_r(\theta_d) = \exp \left[ b_\mu(\theta_d; \gamma, r) + \sum_i \psi_i b_i(\theta_d; \gamma, r) \right]. \tag{4.12}
\]
where $b_\mu$ is the mean function, $b_i$ are the learned basis functions, and $\psi_i$ are the DSBRDF coefficients. Each coefficient $\psi_i$ for the $i$-th basis function is a scalar that controls the influence of the $m$-th eigenfunction to the DSBRDF parameters across all reflectance lobes.

Because of the nature of PCA, these principal basis functions are ordered by importance (the functional variance of the data) and we may use far fewer basis function than the full set and still retain accuracy. By using a subset of the basis functions, we can trade off accuracy and compactness. We found that as few as 7 bases are sufficient for accurately representing real-world BRDFs.

Figure 4.4 shows a comparison of the modeling error of BRDFs fit to measured data from the MERL database [32]. We compare the DSBRDF model to the bivariate model proposed by Romeiro and Zickler [44] and a simple Lambertian plus Cook-Torrance BRDF [9]. In this example, we use 7 reflectance bases for the DSBRDF and 2 parameters per lobe for chromaticity for a total of 13 free parameters. The bivariate BRDF model is a non-parametric representation that depends only on $\theta_d$ and $\theta_h$ and is computed by averaging over $\phi_d$. This bivariate model represents the best error we can possibly achieve when the BRDF is only modeled as a function of only $\theta_h$ and $\theta_d$. We measure accuracy using root-mean-square error in log-space,

$$E_{\text{RMS}} = \sqrt{\frac{\sum_{\theta_h, \theta_d, \phi_d} \left( \log f(\theta_h, \theta_d, \phi_d) - \log \varrho(\theta_h, \theta_d) \right)^2}{N}},$$

between the measured BRDF $f(\theta_h, \theta_d, \phi_d)$ and the fit BRDF $\varrho(\theta_h, \theta_d)$. As noted by [32], when comparing two BRDFs it is critical to take the log because the specular peaks of the BRDFs typically have values orders of magnitude greater than the diffuse portions.

The figure demonstrates the modeling power of the DSBRDF model compared to a sim-
Figure 4.4: Comparison of the DSBRDF model to the non-parametric bivariate model [44] and Cook-Torrance [9]. This figure shows the log-space RMSE of fitting Lambertian and 1 lobe of Cook-Torrance (10 free parameters), Lambertian and 3 lobes of Cook-Torrance (24 free parameters), the DSBRDF using only 7 learned bases (13 free parameters), the full DSBRDF model (42 free parameters), and the non-parametric bivariate BRDF (24,300 parameters). The figure demonstrates that the DSBRDF model accurately captures real-world reflectance functions with a low-order parameterization.

The DSBRDF model can model real-world BRDFs more accurately than Lambertian and several lobes of Torrance-Sparrow with a slightly larger parameterization (42 free parameters for the DSBRDF compared to 24 free parameters for Lambertian plus three lobes of Cook-Torrance). However, when we truncate the set of reflectance bases and only use the 7 most important bases, we get a model that is still more accurate than Lambertian plus three lobes of Cook-Torrance but with a more compact parameterization (only 13 free parameters). This shows that we have achieved a very expressive but also compact model, which will be important when we want to perform inference.

4.4.1 A Prior on the Reflectance Bases

With the entire model specified, we can develop specialized priors that will enable tractable inference. We first place a prior on the variation of the $(\kappa, \gamma)$-curves.
A straightforward way to constrain the variation is to exploit the fact that principal component analysis naturally assumes the projection of the data onto each principal component is characterized by a Gaussian distribution,

\[ p(\Psi) \sim \mathcal{N}(0, \Sigma), \quad (4.14) \]

where \( \Sigma \) is the covariance matrix derived from the PCA. This provides a fairly good fit to the data.

We can exploit our model better, however, by examining how the reflectance bases characterize the space of BRDFs in the MERL database. Figure 4.5 shows a visualization of the prior overlaid onto the DSBRDF space. The DSBRDF space is visualized by plotting the projections of each BRDF onto the first two basis functions \( b_0(\theta_d), b_1(\theta_d) \). We note that the projections give a natural clustering of the BRDFs, with pure diffuse lobes spread over the right side of the graph, metallic-looking materials tightly clustered towards the center, and plastic-looking materials toward the left. This tells us that the DSBRDF model is naturally separating different types of materials through the reflectance bases. It then makes sense to model this using a Gaussian mixture model,

\[ p(\Psi) = \sum_i \pi_i \mathcal{N}(\Psi | \mu_i, \Sigma_i), \quad (4.15) \]

to provide a stronger prior and to capture this variation of different types of materials. We overlay the ellipses of the Gaussian mixture in Figure 4.5 to demonstrate how it models the space of BRDFs.
4.4.2 A Prior on Lobe Chromaticity

We can also place a novel prior on the chromaticity vectors. The most important variation we should capture is the variation in color among the lobes—that is, the joint variation of lobe chromaticity. We observe that the chromaticity vectors in a BRDF will often have similar hue but will not necessarily have similar saturation. We use this observation to construct a simple prior distribution that quantifies the relationship of lobe chromaticity. This is difficult to do, however, as there is no obvious reparameterization of chromaticity into “hue” and “saturation”. We must therefore derive a hue representation.

Hue is typically encoded as a rotation around the center of a color triangle. Chromaticity is represented with Barycentric coordinates so it would be unnatural to try to compute this rotation as if it were a Euclidean space. Instead, we will use a function that gives a vector that is invariant to saturation. First, consider the function,

$$g_l(\mathbf{x}, \alpha) = \frac{x_i^\alpha}{\sum_{i'} x_{i'}^\alpha},$$

(4.16)
where \( x \) is a Barycentric coordinate. As \( \alpha \) grows, the point \( x \) will move from the center of the color triangle toward the nearest corner. The derivative of \( g \) with respect to \( \alpha \) at \( \alpha = 0 \),

\[
g'(x, 0) \propto \mathcal{L} \log x,
\]

(4.17)

where

\[
\mathcal{L} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix},
\]

(4.18)

will give a vector whose direction can be interpreted as “hue” and whose magnitude can be interpreted as “saturation”. Normalizing it will then give us “hue”:

\[
h(x) = \frac{\mathcal{L} \log x}{\| \mathcal{L} \log x \|}.
\]

(4.19)

Given that the hue vector lies on the unit circle, it is a natural choice to place a von Mises distribution [15] on it. In order to address the relationship of lobe chromaticity vectors in a BRDF, we will construct a prior distribution for pairs of lobe chromaticity vectors. The prior distribution for the chromaticity of lobe \( r \) given the chromaticity of lobe \( r' \) is,

\[
p(h(c_r) \mid h(c_{r'})) \propto \exp \left[ \kappa_h h^T(c_{r'})h(c_r) \right],
\]

(4.20)

where \( \kappa_h \) is the concentration parameter. Assuming there are three conditionally independent lobes and assuming a uniform marginal distribution, we can write the joint distribution for \( c \),

\[
p(c) = p(h(c_2) \mid h(c_1))p(h(c_3) \mid h(c_1)).
\]

(4.21)
Figure 4.6: Empirical support for the novel BRDF chromaticity prior. The left two rows show the lobe chromaticity values fit to four example BRDFs from the MERL database. Solid dots represent the chromaticity values. The dashed line is Eq. 4.16 as $\alpha$ is varied. The right plot shows the distribution of the angles between chromaticity hue vectors for each unique pair of lobes among all MERL BRDFs. We then fit a von Mises distribution to this (cyan curve). The strong correlation of lobe chromaticity hue vectors shows that the novel BRDF chromaticity prior effectively constrains the BRDF color to reflect that of real-world BRDFs.

Intuitively, this prior encourages lobes to have similar hues regardless of saturation.

Figure 4.6 visualizes the distribution of chromaticities among the lobes in several BRDFs and gives empirical evidence for our prior. For these example BRDFs, we also illustrate the path of the Eq. 4.16 as $\alpha$ is varied from zero to infinity. Any two lobes having the same dashed path will have the same chromaticity hue vector. Also shown is the distribution of the angles between chromaticity hue vectors for each unique pair of lobes for all MERL BRDFs. This data is used to fit the parameter $\kappa_h$ for the von Mises prior distribution. We can see that the distribution is peaked where the hue vectors are pointing in the same direction, indicating a strong correlation of hue among lobes.

4.5 Conclusion

In this chapter, we introduced the Directional Statistics Bidirectional Reflectance Distribution Function (DSBRDF). We gave a mathematical description of the model and showed
that it is capable of accurately representing a diverse set of real-world BRDFs. We also introduced novel reflectance priors that will help us to infer the reflectance parameters from images. In the following chapters, we will leverage the DSBRDF to infer complex reflectance functions from synthetic and real-world images.
Chapter 5: Reflectance Estimation under Point Light Illumination

In this chapter, we investigate inference under a simple point light illumination model. Although inference under a single point light is an unrealistic assumption, it will allow us to understand how well the model is able to extrapolate unseen slices of the reflectance function. This is because an object imaged under a single point light will only express a small portion of its reflectance function (i.e., $\theta_d$ is fixed). Any inference procedure that attempts to recover a full BRDF from an object illuminated by a single point light must therefore make assumptions about the other parts of the reflectance function that are not observed. In this section, we show that our model is making intelligent assumptions about the behavior of unseen BRDF slices.

5.1 Inferring Reflectance

We may now formulate single material estimation under a single point light as estimating the DSBRDF parameters $R = \{\Psi, c\}$ from the image. Note that, for single material estimation, the entire object surface shares a single set of parameters $R$. We achieve this with a Bayesian formulation that lets us fully leverage the distribution of real-world materials derived in Chapter 4. Given an image $I = \{I_p = (I_{R_p}, I_{G_p}, I_{B_p}) | p = (x, y)\}$, our goal is to find the maximum a posteriori estimate of the DSBRDF parameters $R$ and the point source $L = (L_x, L_y, L_z)$ including its direction and intensity

$$p(R, L|I) \propto \prod_p p(I_p|R, L)p(R)p(L),$$
where \( p(\mathbf{R}) = p(\Psi)p(\mathbf{c}) \). We assume a uniform distribution for the normalization \( p(I) \) and also for the prior probability of the light source direction \( p(L) \).

We write the likelihood distribution using a Laplacian distribution on the logarithm of the irradiance values,

\[
p(I|R, L) \propto \prod_{p, \lambda} \exp \left[ -\beta_I \left| \log I_{p, \lambda} - \log \hat{I}_\lambda(p; R, L) \right| \right], \tag{5.1}
\]

where \( \hat{I}_\lambda(p; R, L) \) is the mean of the distribution and represents a function that computes the irradiance at pixel \( p \) and color channel \( \lambda \) given parameters for reflectance \( R = \{\Psi, c\} \) and illumination \( L \) and \( \beta_I \) is a parameter controlling the scale of the distribution. For point light illumination, we can write the predicted irradiance,

\[
\hat{I}_\lambda(p; R, L) = \varrho(\theta_d(N_p, L), \theta_h(N_p, L); R) |L| \cos \theta_i(N_p, L), \tag{5.2}
\]

where the polar angles \( \theta_h \) and \( \theta_d \) and the incident angle \( \theta_i \) are computed using the known surface normal at pixel \( N_p \) and the light source direction \( \frac{L}{||L||} \). Here the illumination is assumed to be white.

We use a Laplacian distribution to model the image formation process because it allows for robust estimation [52]. It’s important to place the distribution on the logarithm of the radiance so that the likelihood is invariant to the scale of the input image pixel intensities which can be arbitrary due to the camera response function.

With the complete posterior distribution specified, we can now perform inference by computing a maximum a posteriori estimate. In practice, we will minimize the negative log of the posterior with an alternating minimization approach. After initialization, we fix the reflectance and chromaticity estimates and estimate the illumination. Then, we
fix the illumination estimate and estimate the reflectance and chromaticity. We alternate estimating reflectance and illumination in this fashion until convergence (determined by the difference in the negative log posterior between successive iterations), which typically takes 5 to 10 alternations. Although our priors help to keep the optimization on the right track, there are cases where the optimization gets stuck in a local minimum.

We use a simple procedure for estimating the point light parameters. We model the light as a point on a sphere with a non-negative intensity value. During the inference algorithm, when the reflectance estimate is fixed, we simply use Powell’s method [41] to update the lighting parameters.

We initialize our algorithm with a very straightforward approach. The reflectance coefficients $\Psi$ are initialized to all zeros. This can be interpreted as the “mean BRDF” because of the data-driven bases derived from the MERL database. The chromaticity $c$ is initialized to $1/3$ for each color. The light direction is initialized to be the same direction as the viewer.

5.2 Experimental Results

To perform this evaluation, we set up 500 single point light experiments. For each of the 100 BRDFs in the MERL database [32], we render a sphere with that BRDF under five point light directions: 0, 30, 60, 90, and 120 degrees from the viewer. We then run our inference algorithm for each rendered sphere and quantitatively evaluate the estimated BRDF. Note that we exclude the BRDF being estimated in the data-driven basis and prior computations.

To evaluate BRDF estimates, we again use the log-space RMSE (Eq. 4.13). As previously discussed, there is an ambiguity in scale between the light and reflectance function that cannot be known without additional information. Therefore, there is an unknown scale factor between the recovered BRDF and the ground-truth BRDF. Because we know the ground-truth lighting intensity, we can simply invert the estimated lighting intensity to
Figure 5.1: Quantitative evaluation of single point light experiments and select results. Subfigure 5.1a shows the log-space RMSE (Eq. 4.13) of recovered BRDFs from 500 single point light experiments. Each BRDF from the MERL database was illuminated by a point light at 0, 30, 60, 90, and 120 degrees from the viewing direction. Subfigure 5.1b highlights several example results to illustrate how certain log-space RMSE values correspond to perceptual accuracy. The BRDFs on the right are visualized by rendering a sphere under several point light directions.

arrive at the correct scale factor.

Figure 5.1 shows quantitative results for all 500 experiments. To give these quantitative results meaning, we compare several BRDFs to ground truth as a cascaded rendering of spheres from different point light directions. This helps visualize how log-space RMSE
values correspond to perceptual BRDF accuracy. From the plot we can see that specular BRDFs are more difficult to estimate. This is likely because the specular BRDFs have a greater degree of variation, especially as $\theta_d$ varies.

Figure 5.2 shows a quantitative evaluation of the point light estimates. As the figure shows, all estimates of the illumination direction are within one degree of the ground truth lighting.

5.3 Conclusion

The extensive synthetic results give us convincing evidence that our reflectance inference framework is performing well in this restricted case. The Bayesian framework is capable of recovering a large variety of reflectance functions and light source positions from a single image. This ability is necessary for recovering reflectance in the real world. Next, we broaden the applicability of this framework by considering scenes with multiple unique reflectance functions.
Chapter 6: Single Image Multimaterial Estimation

Identifying materials from their images is a long quested ability of computer vision systems to better interact with the real world. Estimating the radiometric properties of materials from their images (i.e., recovering the reflectance of object surfaces) is a fundamental step towards achieving this goal. If successful, the estimated reflectance can give robust cues to other descriptive attributes, such as its tactile and geometric properties (e.g., surface roughness and rigidity), and also greatly narrow down the search space for recognizing materials from their raw image cues.

In this chapter, we will focus on the challenging but practical task of estimating the reflectance of object surfaces made of multiple materials from as few as a single image. We assume that the reflectance of the surface can be approximated well with pointwise reflectance, namely the Bidirectional Reflectance Distribution Function (BRDF), and its geometry can be estimated beforehand, in our case with a laser-stripe range sensor. We also assume that the illumination consists of an unknown single directional light source.

Multimaterial estimation under these assumptions is still a very challenging task mainly for two reasons: material assignment and limited angular samples. Consider estimating the reflectance of an object surface that occupies $N$ pixels in its image. All $N$ combinations between the two extrema, the surface consisting of a single material and the image providing $N$ angular samples of its reflectance on one hand and the surface consisting of $N$ materials and the image providing on average only one angular sample of each material’s reflectance on another hand, are plausible solutions. The main challenge, however, lies in the assignment of the pixels to distinct materials as each $k$-th combination can have $k^N$ valid solutions each
of which corresponds to a different spatial segmentation of the object surface with up to $k$ materials.

Most important, as the number of materials increases, the angular sampling of each material’s reflectance decreases. The fact we have only a single point source compounds this problem as each angular sample would only capture a single combination of incident and exiting directions. As such, even if we could manually give a fairly reliable estimate of the number of materials (e.g., with the aid of color segmentation) segmenting the surface into individual materials while estimating the reflectance of each material poses significant challenges that require the interpretation of sparse and under-sampled angular observations of each reflectance. Surface normals of real object surfaces rarely cover the full frontal half of a sphere as often assumed in other works, which further exacerbates this problem.

We address these challenges by leveraging the DSBRDF and constructing a novel probabilistic formulation of multimaterial estimation that fully leverages those reflectance priors.

We derive a layered Markov random field formulation of multimaterial estimation, inspired by the work of Sun et al. [48], to fully leverage these reflectance priors. Each material is represented with an MRF in this formulation. The spatial extent of each material is modeled with a continuous latent layer that encodes soft assignments of pixels to that material. The reflectance of each material is then modeled using a set of DSBRDF parameter values. This formulation nicely captures the spatial segmentation of the multiple materials while allowing us to place constraints on the solution. We jointly estimate the material segmentation and each material reflectance, together with the strength and direction of a single point source.

We experimentally evaluate the effectiveness of the method on a number of synthetic and real object surfaces that consists of different numbers of materials. The results clearly
demonstrate that the method can achieve accurate segmentation and recovery of the reflectance. They also show that the method can handle real-world materials that exhibit reflectance that cannot be modeled with conventional parametric models. We believe the method yields an important step towards realizing radiometric material estimation in general images.

6.1 A Layered MRF Formulation

We solve single image multimaterial estimation by fully leveraging additional characteristics of real-world surfaces. We observe that adjacent pixels tend to belong to different materials when their intensities differ greatly. Although adjacent pixels in an image will hardly ever be the exact same color, they will usually differ smoothly due to lighting if they belong to the same material.

We derive a probabilistic formulation to exploit this observation together with the reflectance model and priors. The formulation, inspired by the work of Sun et al. [48], uses a layered graphical model where each \( k \)-th layer is responsible for both the spatial extent and surface geometry of the \( k \)-th material out of \( K \) distinct materials. Each layer models the spatial extent and surface geometry with a Markov random field (MRF), \( m_k \) respectively. The MRF \( m_k \) specifies the spatial region of the \( k \)-th material with continuous latent variables at each pixel \( p \). We denote the collection of this MRF for all \( K \) materials with \( m \).

We say that a pixel \( p \) belongs to the \( k \)-th material if and only if \( m_{k,p} (\in m_k) \) is the maximum value among all other layer values \( \{m_{k',p} | k = 1, \ldots, K\} \). For this, we define an indicator random field \( s \) that, for each pixel \( p \), encodes the material of the surface in a 1-of-\( K \) coding

\[
s_{k,p} = 1 \{ m_{k,p} = \max_k m_{k',p} \},
\]  

(6.1)
where $\mathbf{1}$ is an indicator function returning one when the expression inside the braces are true, zero otherwise.

The assumptions that the illumination will cause a material to vary smoothly across an image, whereas changes in material are more likely to happen between pixels with large intensity differences, can be modeled with a spatial prior on each material MRF

$$p(m) = \prod_k \prod_p \prod_{p' \in N(p)} \exp \left[ -w(p, p')(m_{k,p} - m_{k,p'})^2 \right], \quad (6.2)$$

where $N(p)$ denotes the 4-neighbors of pixel $p$ and the weights reflect the intensity differences

$$w(p, p') = \max \left\{ \exp \left[ -\frac{1}{2\sigma_c^2} \| I_p - I_{p'} \|^2 \right], \delta \right\}, \quad (6.3)$$

where the difference is the $l_2$ norm of RGB vectors and the parameters $\sigma_c$ and $\delta$ control the sharpness and penalty of material boundary edges.

We again write the likelihood distribution using a Laplacian distribution,

$$p(I|R, m, L, N) = \prod_k \prod_p \exp \left[ -s_{k,p} I_p - \hat{I}_{\lambda}(p; R_k, L, N) \right], \quad (6.4)$$

where $R = \{\Psi, c\}$ are the set of the DSBRDF parameters and $\hat{I}$ is a function that computes the irradiance at pixel $p$ (Eq. 5.1).

### 6.2 Inferring Material Reflectance and Region

We estimate the reflectance of each surface point from a single image by computing the maximum a posteriori estimates of the reflectance $R = \{R_k|k = 1 \ldots K\}$ and the spatial
extent \( m = \{ m_k | k = 1 \ldots K \} \) of each of the \( k \)-th material,

\[
p(R, m, L | I) \propto p(I | R, m, L, N) p(m) p(L) \prod_k p(R_k),
\]

where \( p(R_k) = p(\Psi_k) p(c_k) \). We assume a uniform distribution for the light source prior \( p(L) \). In practice, we compute the MAP estimate via energy minimization which corresponds to minimizing the negative log posterior while weighting the prior terms.

We alternate between estimating the light source \( L \), the reflectance parameters \( R \), and the material extents \( m \). We employ the same method described in Chapter 5 to estimate the light source. In the material estimation step, we compute the reflectance parameters for each material \( R_k \) given the current material segmentation \( m \) to compute the unary prior \( p(R_k) \). For each \( k \)-th material, this is equivalent to the single material estimation discussed in Chapter 5. The error term in an energy potential form corresponding to the unary prior \( p(R_k) \) is also weighted proportionally to the number of pixels assigned to material \( k \),

\[
E(R_k) = \sum_k \sum_p -s_{k,p} \log p(R_k)
\]

(6.6)

to avoid penalizing small material regions more heavily. Once we compute the reflectance for each material, given the material segmentation from the last iteration, we can estimate the spatial extent \( m_k \) of each material by minimizing the negative log posterior using the L-BFGS algorithm [6].

We compute an initial estimate of the light source direction by thresholding the image to roughly obtain the specular highlights and then by averaging all the reflected directions of the viewing direction mapped on the geodesic dome using the known surface normals of
these highlight pixels. Although this only gives a rough estimate as highlights can never be exactly localized with simple thresholding, we found it to be good enough to start the estimation.

We obtain an initial estimate of the material segmentation $m$ through $k$-means clustering on the chromaticity of the single input image. The clustering gives us a set of centroids $\{h_k \mid k = 1, \ldots, K\}$, where each centroid $h_k$ is a three-dimensional chromaticity vector $\bar{I} = \frac{1}{I_R + I_G + I_B} (I_R, I_G, I_B)$. We set the initial estimates $m^{(0)}$ to be inversely proportional to the distance of each pixel in the chromaticity image to the centroids

$$m^{(0)}_{k,p} = \exp \left[ - (\bar{I}_p - h_k)^2 \right]. \quad (6.7)$$

Finally, we approximate the max operator to compute the segmentation map $s$ from the material MRFs $m = \{m_k \mid k = 1, \ldots, K\}$ with a differentiable function so that the gradient can be computed when optimizing for $R$ and $m$.

$$s_{k,p} \approx \frac{\exp \gamma m_{k,p}}{\sum_{k'} \exp \gamma m_{k',p}}.$$ 

In this approximation, $\gamma$ controls its accuracy: the approximation becomes exact as $\gamma \to \infty$. We, however, keep $\gamma$ to a moderately large number to avoid arithmetic overflow. It is worth pointing out that $\sum_k s_{k,p} = 1$, indicating that our selection of $\gamma$ will never over- or under-penalize the pixel $p$. A consequence of this is that we may set $\gamma$ low to allow materials to blend together, which is useful for objects that do not have sharp material boundaries.
Figure 6.1: Multimaterial estimation results for three real scenes. Each row, from left to right, shows the input image, a synthesized image of the scene using estimated reflectance and light source, the material segmentation result, a relit image of the object, and a ground truth image of the relighting result for each scene. The results demonstrate that the method successfully recovers the reflectance for complex scenes, for instance the gold paint on the mask, except when interreflections and shadowing are prevalent in the scene.

6.3 Experimental Results

Figure 6.1 shows the results of multimaterial estimation for several real-world scenes. The results show that the method successfully segments the object surface into distinct materials and accurately estimates the reflectance at each surface point. The painted mask object has a very smoothly varying surface; because of this, normal refinement was not necessary. The cup scene, on the other hand, features particularly sharp specular highlights.

Some of these materials have reflectance that cannot be captured with conventional diffuse plus specular reflectance models. For instance, the gold paint on the forehead of...
the mask does not have any diffuse component and its color is actually in its specular reflection. The method successfully recovers the reflectance of such challenging real-world materials from the limited information that can be extracted from a single image, which truly demonstrates the power of the reflectance priors and the probabilistic formulation.

For the scenes in figure 6.1 the estimated light source direction was within 13.5 degrees of the ground truth direction. We believe these errors were mainly caused by the imperfect surface normal measurements.

Figure 6.2 gives some insight to how the method operates by showing the geometry of the mask rendered with each material, and the values of material assignment $m_{k,p}$. The method attempts to minimize the difference between adjacent values of $m_{k,p}$ while simultaneously minimizing the difference between the predicted and the observed data. It allows material regions to be discontinuous (jump in the values of $m_{k,p}$) when the difference between adjacent pixel intensities are high, and this effect is visible in the figure.

In this chapter, our method does not model global illumination effects like interreflection.
and cast shadows which can contribute greatly to the scene irradiance of more complex objects. This effect is noticeable in the cups scene: the center cup is reflected in the green cup and a heavy shadow cast on the red cup causes that area to be erroneously assigned to a dark material. In Chapter 8, we explore methods to predict these global effects by modeling the complete geometry of the scene rather than just the surface normals.

We note some limitations of our method as well. Our formulation assumes that a scene is made up of only a handful of materials. As a result, the method is not suited to scenes that feature smoothly varying BRDFs along the scene surface. As we mentioned, it is also unable to handle scenes with significant global illumination effects. Finally, a scene including materials with a limited number of angular samples (e.g., a scene composed of a set of planes) can cause the method to fail to accurately capture the BRDFs. Using multiple input images can alleviate this problem.

6.4 Conclusion

In this chapter we developed a spatial segmentation model to enable the inference of reflectance in scenes with multiple unique materials. We developed this model by creating a set of latent material layers that combine to form the indicator segmentation. This allowed us to easily place strong spatial priors on the latent layers that encourages large contiguous material regions. We demonstrated our method on a novel set of real scenes and showed that we can accurately decompose the reflectance functions of a scene.
Chapter 7: Reflectance and Natural Illumination from a Single Image

An image is a function of several scene components: object geometry, object reflectance, and scene illumination. So far, we have investigated the recovery of reflectance functions from scenes under simple, controlled illumination with known geometry. In general, however, scene illumination is more complex than a single infinitely-distant point light. In fact, real scenes are illuminated from every possible incident direction. Knowing this illumination environment could tell us about the scene itself—for example, it could be used to infer whether a scene is indoors or outdoors. Acquiring the reflectance of an object and the illumination environment would also enable the prediction of object appearance in a novel scene, helping support appearance-based object tracking and recognition. The most critical factor in solving this problem successfully is that we must design an approach with real-world complexity in mind.

In this chapter, we investigate the problem of estimating object reflectance and scene illumination from a single image given the geometry of that object. We would like to solve this problem for images taken in the wild—images of real-world objects taken under natural, complex illumination. To do this, we make as few limiting assumptions about reflectance and illumination as possible: we don’t assume that reflectance can be accurately represented by simple models like Lambert’s law [27] or Torrance-Sparrow [50] and we don’t assume that illumination can be modeled by a small set of point lights or a small set of generic linear bases. Rather, we will leverage highly expressive reflectance and illumination models that do not constrain us to simple synthetic scenes. This added flexibility significantly complicates the problem by introducing many additional variables. To address this difficulty, we must
analyze the natural variation of real-world reflectance and the effect of reflectance on the illumination environment.

Each reflectance and illumination separately contribute to the difficulty of the joint inference problem. Fitting a reflectance function from a single image is difficult even when the illumination is known because a single image only reveals a small fraction of the full reflectance function [7]. An inference approach must therefore be able to sensibly extrapolate this unseen information. The addition of unknown natural illumination further complicates the problem because of the sheer number of variables introduced. Many approaches assume that natural illumination can be modeled with a low-order parametric model like spherical harmonics or a small number of point light sources. These models fail to accurately predict scene irradiance when the reflectance model is not Lambertian. A full-color non-parametric illumination model representing the sphere of incident illumination is necessary to model real-world scenes.

Reflectance and illumination themselves are not the only problem—their interaction compounds the difficulty. Several ambiguities exist between reflectance and illumination caused by the image formation process that thwart inference algorithms. One is a bilinear ambiguity between the magnitude of the reflectance and the illumination. The consequence of the ambiguity is that multiplying the reflectance by a scale factor and dividing the illumination by that same scale factor will produce the same image as the unscaled reflectance and illumination. The problem is compounded by the fact that the bilinear ambiguity exists among each color channel independently and therefore gives rise to the color constancy problem. If we observe an object in an image, we cannot be sure whether the color of that object is due to the reflectance function, the incident illumination, or some combination of both. An ambiguity also exists between the specularity of the reflectance function and
the sharpness of the illumination map, as first noted by Ramamoorthi and Hanrahan [42]. This means that if we increase the specularity of the reflectance function and blur the illumination environment, we would produce the same image as the original reflectance and illumination. Consequently, a trivial solution to the reflectance and illumination inference problem is a perfect mirror reflectance function. Careful use of prior knowledge is required to resolve these ambiguities.

We utilize the DSBRDF model and priors described in Chapter 4. We model the illumination environment with a non-parametric representation of the incident illumination field. A non-parametric model allows for a great deal of expressibility that must be constrained in order to reduce the solution space. We use several priors to do this. First, we develop an entropy prior that models the entropy loss of the illumination due to its interaction with the object reflectance. We also utilize a natural image statistic prior that encourages a plausible natural illumination environment to be recovered.

We thoroughly evaluate the effectiveness of our Bayesian joint estimation with synthetic and real images. First, we demonstrate the ability of our model to successfully infer reflectance and illumination in a number of synthetic scenes. We quantitatively evaluate results on synthetic scenes by computing the log-space root-mean-square error between the ground truth and inferred reflectance and illumination. We qualitatively discuss results on real scenes and compare inferred illumination to the ground truth. In the end, we show that the key features of the model greatly ease this difficult inference problem.

7.1 Modeling Natural Illumination

Handling natural illumination is essential to enable radiometric scene property inference from real-world images. Real-world scenes cannot be adequately represented by simple models like a small set of point lights or spherical harmonics. We must consider that a
scene is illuminated by every visible direction in full color. Although this ostensibly burdens the inference problem, there is a great deal of latent structure in the natural illumination environment that we can exploit.

7.1.1 A Non-parametric Illumination Model

In the past there have been numerous approaches for modeling natural illumination environments. Barron and Malik use spherical harmonics as a linear basis [2]. Romeiro and Zickler model natural illumination as a linear combination of data-derived basis functions [45]. Rather than attempt to reduce the variability of the representation, we use a very expressive representation but control variability with carefully constructed priors. The priors will help overcome the ambiguities of the problem and encourage the recovery of accurate illumination maps. We represent natural illumination non-parametrically as a dense grid on the sphere which we call the illumination map \( L = \{ L_{\theta, \phi, \lambda} | \theta \in [0, \pi/2], \phi \in [0, 2\pi], \lambda \in \{ R, G, B \} \} \) for incident angle \( \{ \theta, \phi \} \) and color channel \( \lambda \). We can think of this representation as a wide-angle image of the surrounding environment. Because of the great expressiveness afforded by this model, we must apply intelligent priors to constrain the solution space and recover sensible estimates.

7.1.2 Illumination Priors

We know that the natural illumination environment can be interpreted as a wide-angle image of the scene surrounding the object. It follows, then, that we can place natural image statistics priors on the illumination environment to properly constrain it. Previous works on natural image statistics has found that the gradients of natural images form a distribution with a heavy tail [22]. We can therefore place a heavy-tailed prior distribution on the gradient magnitudes of the illumination map. We do this with a hyper-Laplacian
Figure 7.1: Gradient Magnitude Histogram of Natural Illumination. Subfigure 7.1a shows an example natural illumination map. Subfigure 7.1b shows the gradient magnitude histogram of this natural illumination map with a hyper-Laplacian distribution overlaid. We can see that the hyper-Laplacian distribution is a good model for the gradient magnitudes.

\[
p_1(L) \propto \prod_{\theta, \phi} \exp \left[ -b^{-1} \left( \sum_{\lambda} \frac{\partial L_{\theta, \phi, \lambda}}{\partial \theta}^2 + \frac{\partial L_{\theta, \phi, \lambda}}{\partial \phi}^2 \right)^\alpha \right], \tag{7.1}
\]

where \( \frac{\partial L_{\theta, \phi, \lambda}}{\partial \theta} \), \( \frac{\partial L_{\theta, \phi, \lambda}}{\partial \phi} \) is the partial derivative of the illumination map in the y and x direction, respectively, \( b \) is the scale parameter, and \( \alpha \) is the shape parameter. Here we set \( \alpha = 0.8 \).

Note this formulation of the image gradient magnitude takes the color of the gradient into account.

Figure 7.1 shows an example real-world natural illumination environment and its corresponding gradient magnitude histogram. We have fit a hyper-Laplacean distribution to this histogram to verify that it is indeed a good fit to the data.

This prior is important for two reasons. First, it helps promote a gradient distribution in the recovered illumination map that a natural image would have. It will also help conquer the ambiguity between the specularity of the reflectance and sharpness of the...
illumination map by deterring the optimization from choosing the trivial solution—a mirror-like reflectance function and a blurred illumination map. If the current estimate is the trivial solution, the soft gradients of the illumination map will cause the image-gradient prior to have very low probability.

7.1.3 The Entropy Increase by Reflection

Our goal is to properly constrain the space of possible illumination estimates that is made ambiguous during image formation. Recall that part of the reason why this ambiguity exists is that the material reflectance acts on the illumination as a bandpass filter. We empirically show that, due to the band-limited reflectance of incident irradiance, the entropy of the distribution of reflected radiance becomes higher than when there was no bandpass filtering (i.e., the reflectance has all frequencies—perfect mirror reflection).

Entropy has been studied in many areas of computer vision as a useful information theoretic metric to constrain estimation. Alldrin et al. [1] use the entropy of the distribution of albedo values of a textured surface to resolve the generalized bas-relief ambiguity in photometric stereo. They assume that the true distribution of albedo values is sparsely peaked and use entropy to guide the estimation to find the albedo values that conform to this property. Finlayson et al. [14] propose an approach to remove shadows by finding a direction in a 2D chromaticity feature space under which illumination is invariant. The correct invariant direction minimizes the entropy of the image formed by projecting the 2D feature points along it. As it so happens, entropy can also be used to order the set of potential illumination maps that we face during joint estimation.

Figure 7.2 demonstrates the effect of the reflectance on the entropy of the reflected radiance for a variety of materials. As our intuition suggests, the action of the BRDF as a bandpass filter causes a blurring of the illumination and thus a spreading of the histogram.
Figure 7.2: The BRDF always causes an increase in entropy of the reflected radiance. The most specular materials (e.g., nickel) cause the least increase while the most diffuse (e.g., green-latex) cause the greatest increase. Only a perfect mirror (e.g., chrome), will not increase entropy.

This, in turn, increases the entropy of the reflected radiance. We’d like to recover the true illumination environment and to do this we assume that the entropy increase in the observed image is due entirely to the BRDF. To this end, we constrain the illumination to have minimum entropy, so that the BRDF will be responsible for causing the increase in entropy of the outgoing radiance.

It is with this intuition that we formally derive our entropy prior. Entropy is defined as the expected value of the information context of a random variable. While typically defined for discrete random variables, we use a continuous definition which will allow us to take the derivative with respect to the illumination map later. The entropy of an image is an integral over the distribution of intensity values,

$$H = -\int p(x) \log p(x) dx,$$

(7.2)

where in our case $p(x)$ is the histogram of $L$. To ensure that entropy is differentiable with respect to the illumination map, we write the histogram using kernel density estimation
with a Gaussian kernel

\[ p(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - L_i)^2}{2\sigma^2} \right], \tag{7.3} \]

where \( N \) is the number of pixels in the illumination map, \( L_i \) chooses the \( i^{th} \) point, and \( \sigma^2 \) is the variance of the Gaussian kernel. Using this expression of entropy, we place an exponential prior on the illumination map with an exponent proportional to the entropy

\[ p_2(L) \propto \exp \left[ -H(L) \right]. \tag{7.4} \]

Multiplying the two illumination priors together, we obtain the complete illumination prior,

\[ p(L) \propto p_1(L)p_2(L). \tag{7.5} \]

### 7.1.4 Resolving the Color Ambiguity

One problem we have not addressed thus far is the color constancy problem between the reflectance and the illumination. As previously mentioned, if we are given an image of an object that appears red, we cannot know whether the red appearance is due to the reflectance or the illumination. This is made even more difficult by the fact that we are only dealing with a single object: if we observed that all the objects in a scene were red, we might assume that the redness is due to illumination. To handle the color ambiguity, we adopt the grey-world hypothesis \[5\] and assume that the lighting environment is, on average, uncolored. Consequently, we assume that the dominant coloration of the observed image is caused by reflectance.

We propose a simple way of exploiting this assumption. When performing joint estima-
Figure 7.3: Evaluation of synthetic experiments under natural illumination. The bottom plot shows the log-space RMSE for 100 MERL BRDFs under four natural illumination environments. Highlighted are eight results from that plot shown at the top. This plot demonstrates the ability of the model to successfully infer reflectance and illumination in a variety of illumination environments for many different materials.
Figure 7.4: Quantitative evaluation of natural illumination estimation versus BRDF specularity. This figure shows the RMS error of the natural illumination estimates in blue for 100 MERL BRDFs in four illumination environments. In red is the “specularity” of the BRDF, as measured by the $\gamma$ value of the most specular lobe. This figure illustrates the general trend that high specularity BRDFs enable more accurate illumination estimation.

In all the lighting environments and materials we tried this approach and it successfully deduced the material color from the scene color.

### 7.2 Inferring Reflectance and Natural Illumination

As in Chapter 5, we will minimize the negative log of the posterior with an alternating minimization approach. After initialization, we fix the reflectance and chromaticity estimates and estimate the illumination. Then, we fix the illumination estimate and estimate the reflectance and chromaticity. We alternate estimating reflectance and illumination in this fashion until convergence (determined by the difference in the negative log posterior between successive iterations).
Figure 7.5: Predicting the appearance of materials with recovered illumination. We use the recovered illumination map to predict the appearance of materials with lower frequency reflectance. The top row shows the ground truth and the bottom row shows the predicted appearances. The input image for each subfigure is the top-left image. These results demonstrate the ability to accurately predict object appearance with the recovered illumination map.

7.3 Experimental Results

We thoroughly evaluate the effectiveness of our method with a number of synthetic and real-world images. First, we explore full reflectance and natural illumination inference in synthetic scenes. Dealing with synthetic scenes allows us to ignore some confounding variables and gives us the opportunity to better analyze the results. Finally, we use our framework to estimate reflectance and natural illumination in the real world.

7.3.1 Synthetic Results

To evaluate our algorithm synthetically, we created 400 experiments using 100 MERL BRDFs in four different illumination environments. Each material comes from the MERL BRDF database [32]. Illumination environments are from Paul Debevec’s Light Probe Gallery [12]. We compare each recovered material and illumination estimate to the ground-truth values.

Each synthetic experiment uses a single 200 × 200 pixel input image and attempts to infer a 128 × 64 pixel illumination map. We found that the alternating minimization of
Figure 7.6: Predicting object appearance from different views. By using the recovered illumination map and reflectance estimate, we can show the object as it would appear from different viewing directions. The first row of each subfigure contains ground truth renderings for various BRDFs, using the measured BRDF values and true environment maps, and the second row contains the predicted appearance using the recovered reflectance and illumination from our method.

the negative log posterior typically takes 5 to 10 alternations to converge. In practice, the optimization concludes in about 15 minutes running on an NVIDIA Tesla K40m.

Like the single point-light experiments in chapter 5, we use log-space RMSE to quantitatively evaluate the results. However, determining the scale difference between the recovered and ground-truth BRDF is not straightforward now that the illumination is no longer a single point. Therefore, we will use the scale factor that minimizes the log-space RMSE,

$$\hat{E}_{\ell\text{-RMS}} = \min_{\alpha} \sqrt{\frac{1}{N} \sum_{\theta_h, \theta_d, \phi_d} \left( \log f(\theta_h, \theta_d, \phi_d) - \log \alpha \varphi(\theta_h, \theta_d) \right)^2},$$

(7.6)

Figure 7.3 shows the accuracy of the reflectance estimates compared to a baseline method and compared to the error of fitting the DSBRDF directly to measurements. For the baseline
Figure 7.7: Relighting comparisons. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

Method we simply run our method without priors. We can see visually that the specularity and color of the recovered BRDF is often very close to ground truth. This indicates that our model and priors are working as intended. Many of the recovered illumination environments are also very accurate and detailed—this is particularly true for scenes with specular BRDFs. For more diffuse BRDFs, many of the important features of the illumination environment are captured, including bright windows and the sky.

Figure 7.4 examines the relationship between the specularity of BRDFs and the accuracy of the estimated illumination. We can see that, in general, more specular BRDFs allow for
Figure 7.8: Relighting comparisons continued. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

more accurate natural illumination estimates. This makes sense, because less information is destroyed during the image formation process for specular BRDFs than diffuse. This relationship is especially prevalent in the “rnl” and “uffizi” lighting environments. Figure 7.3 also illustrates this effect.

Figure 7.5 demonstrates how we may use the recovered illumination map to accurately predict the appearance of other materials. The results show that, even for materials that
are not close to perfect mirror reflectance, we can use the recovered illumination to accurately compute the appearance of other materials, showing the ability of using an object of arbitrary material as a light probe.

Figure 7.6 shows one of our most important results: that we can accurately predict object appearance in the same scene from different viewing directions. The figure shows renderings of the ground truth BRDFs and illumination compared with our recovered reflectance and illumination from different viewing directions. The results show the ability to properly predict object appearance from any location in a scene. As mentioned, this ability could
Figure 7.10: Relighting comparisons continued. The first column in each subfigure show a photograph of an object under several different illumination environments. We recover the reflectance and illumination of the object and then use the recovered reflectance to relight the object with each illumination environment which is shown in the subsequent columns. The visual similarity of the relighting to the ground truth demonstrates the accuracy of our reflectance estimates.

benefit object recognition and tracking algorithms.

7.3.2 Real scenes

The most challenging scenes—and the ultimate goal of this paper—are ones in the real world. They may present significant global illumination effects, camera noise, and other unique problems. Despite these great difficulties, we show that our model is able to accurately capture reflectance and illumination from a single image.
Figure 7.11: Results on “lobby” and “spiralStairs” illumination environments. Each row shows the recovered reflectance and illumination of a different object with the ground-truth illumination for comparison. Note how features of the reflectance function are accurately captured: the “apple”, “bear”, “horse”, and “milk” objects are shiny and the recovered reflectance function has sharp specular highlights; the “tree” object has a softer glossy reflection that is captured in the reflectance estimate.

For the real-world experiments, we have captured a new database of high dynamic-range images of real objects with aligned ground-truth geometry taken under a variety of indoor and outdoor illumination environments that we introduce.

Figures 7.7 and 7.8 examine the accuracy of the reflectance estimates through a series of relightings. For each image in the data set, we use the reflectance recovered by our method to relight the object using the ground truth illumination of each environment. In this way, we can visually compare reflectance results to the ground-truth images. This figure shows many compelling results, including those of the “bear” object and the “tree” object. It can be seen that the method is able to recover reflectance and illumination in a wide variety of scenes and it can be effectively used to predict object appearance in novel environments.

Figures 7.11 and 7.12 show real-world results for each illumination environment. In each

\(^4\)Available online at http://www.cs.drexel.edu/~kon/natgeom
Table 7.12: Results on “garden”, “main”, and “picnic” illumination environments. Each row shows the recovered illumination of a different object with the ground-truth illumination for comparison. This illustrates that our algorithm is able to capture important features of the illumination environment. For example, in the “picnic” scene, the method captures the sun behind a building from the “apple” object.

Environment, we can see that our algorithm recovers plausible reflectance estimates through a cascaded rendering of spheres. For example, in Figure 7.11, note the diffuse highlight of the “tree” object. In addition, many important features of the natural illumination environments are recovered. For example, in Figure 7.12 (b), many illumination estimates (especially from the horse object) capture the position and detail of the skylight that is casting much of the light in the scene. Also note the “milk” object in each figure, which does not contain a full hemisphere of surface normals. In these cases, unseen portions of the lighting environment are assumed to be black. Overall, the objects that allow the best recovery of the illumination environment seem to be the “apple”, “bear”, and “horse”. This is likely because the surface normals of these objects span the entire hemisphere facing the environment.
Figure 7.13: Comparison to Romeiro and Zickler [45]. We compare our results qualitatively on a data set of four real spheres created by [45]. The top row shows the input image, the second row shows the predicted appearance from a novel viewpoint using the BRDF recovered using [45], the third row shows the ground truth appearance of the object from the novel viewpoint, the fourth row shows the predicted image from the novel viewpoint using the recovered reflectance function using our method and the ground truth illumination map, and the fifth row shows our result after manual color correction. As shown, our method in general recovers a more accurate BRDF.

camera, and the objects are mostly concave so they do not cause significant inter-reflection.

Figure 7.13 shows a direct comparison to the method of Romeiro and Zickler [45] on the real-world dataset captured by the authors. As shown our method can more accurately capture the frequency characteristics of the reflectance function. In particular, we correctly estimate the soft specular highlight of the gray plastic sphere in the fourth column whereas the method of [45] recovers a more specular BRDF. We believe this is because the joint
inference of reflectance and illumination has a better opportunity to reconstruct the true reflectance than marginalization because the number of possible illumination environments is vast. Representing the distribution and marginalizing over all these possibilities accurately is very difficult—directly representing and solving for the illumination allows us to avoid this problem. Instead, we can exploit specific illumination characteristics that would lead to the most plausible illumination environment.

7.4 Conclusion

In this chapter, we presented a reflectance and natural illumination recovery method that uses flexible reflectance and illumination models in order to handle real-world scenes. We introduced strong priors that keep inference tractable and enable us to overcome the ambiguities due to the image formation process. We used synthetic examples to further analyze the performance of our model under a variety of scenarios. Finally, we demonstrated the performance of our algorithm on real-world scenes.
Until this chapter, we have examined the inference of reflectance of multi-material scenes under point-light illumination and the inference of reflectance of uniform objects under natural illumination separately. To complete this work, we must unify these approaches into a single radiometric scene decomposition method. This is no simple task, as allowing for multiple reflectance functions under natural illumination causes a significant increase in variability and ambiguity of the problem. Additionally, we will relax our assumptions about known geometry by using commodity depth sensors to capture approximate geometry of the scene from multiple views, which will increase the practical applicability of this work.

In this chapter, our goal is to decompose a scene radiometrically: to infer the spatially-varying reflectance of the surfaces, to infer the natural illumination of the scene, and to refine the geometry using the initial rough geometry and color observations given to us through a set of RGB-D images from multiple views. We perform this decomposition on complex scenes, with objects interreflecting, shadowing, and occluding one another, which greatly enhances the difficulty of the problem. We first use the Kinect Fusion algorithm [24, 35] to integrate a series of RGB-D images into a single triangle mesh and simultaneously compute extrinsic camera parameters that is input to our algorithm. As high dynamic-range (HDR) cameras become more ubiquitous, we assume that the RGB images given to us are HDR. Our method then segments the geometry into regions of distinct reflectance, infers the reflectance of those regions, infers the natural illumination of the scene, estimates spatially-varying diffuse texture of the geometry, and refines the initial input geometry.
Recovering these radiometric scene properties in real-world scenes is difficult. As we have discussed, real-world scenes have complex reflectance functions (often featuring off-specular peaks, retroreflection, and subsurface scattering effects) that can vary greatly across a surface and natural illumination causes surfaces to be illuminated from arbitrary directions in the scene. We must now handle both spatially-varying reflectance and natural illumination simultaneously to fully solve the radiometric scene decomposition problem.

As in previous chapters, we will assume that there are a small number of fundamental materials in a scene. In chapter 6, we proposed a spatial segmentation model that is used to infer the reflectance function of each pixel in the image. In this chapter, we are expanding the problem to multiple views of the scene. This means that we must modify our spatial segmentation model to exist on the surface of the geometry rather than existing on the pixels of the image. We also allow the diffuse texture of the scene to vary with high frequency to enable the recovery of surfaces with texture.

We note that the materials in the scene convey a varying amount of information about the illumination environment. For example, a mirror ball will give very accurate information about the illumination environment whereas a Lambertian ball will give very little. We can exploit this observation when estimating the illumination environment by preferring pixels from more specular materials.

In previous chapters, we used 3D laser-stripe range scanners to acquire the geometry of objects in a scene. Acquiring geometry with these scanners is expensive and time consuming. Recently, commodity RGB-D sensors, such as the Microsoft Kinect, have enabled the cheap and ready capture of geometry through a simple image acquisition. RGB-D sensors are appealing for research in photometry because of the strong link between light and geometry.

Depth data from RGB-D sensors, however, is noisy. This is a real problem for recovering
reflectance and illumination because even a small change in the surface geometry can greatly alter the appearance of an object, particularly for specular surfaces. In order to leverage the fast and cheap capture of geometry from RGB-D sensors, we must address the geometric uncertainty and develop a method to tackle it.

In this chapter, we develop a geometry refinement technique for enhancing the rough input geometry. The depth values from RGB-D sensors are typically noisy and the limited spatial resolution of many RGB-D sensors means that fine geometric detail is difficult to capture even when multiple depth images are fused. Additionally, precise surface normals are necessary to accurately estimate the reflectance behavior of specular objects. To manage this problem, we incorporate a procedure to modify the geometry to match the scene appearance.

One major problem that we have yet to address is indirect illumination. The problem of estimating complex reflectance functions under natural illumination in complex scenes is impeded by the radiometric interaction between the objects in the scene. When objects are close to one another, they can cause shadowing, interreflection, and other indirect illumination effects. Properly accounting for this is crucial to accurately estimating reflectance and illumination in real scenes. Without taking indirect illumination into account, a shadowed region could be improperly inferred as a darker material. Furthermore, a mirror-like reflectance function is primarily identified by the fact that it reflects other objects in the scene—we need an understanding of indirect illumination to leverage this information. We incorporate indirect illumination directly into the image formation likelihood to allow our method to seamlessly handle it.

Our key contribution is a probabilistic framework for inferring the rich radiometric scene information (i.e., the complex reflectance, natural illumination, and geometry) contained

---

CHAPTER 8: SPATIALLY-VARYING REFLECTANCE, NATURAL ILLUMINATION, AND REFINED GEOMETRY FROM RGB-D
within images. This is accomplished by several individual contributions: a) a spatially-varying reflectance model that allows for complex, real-world reflectance functions and textured surfaces by separating the scene into a small number of fundamental materials while allowing diffuse texture to vary within material regions; b) an illumination estimation procedure that takes advantage of the frequency characteristics of the different materials in the scene to improve the inferred illumination; c) a geometry refinement scheme that leverages the rich information encoded in scene appearance to refine the input geometric surface given by an RGB-D sensor; and d) an image formation likelihood that accounts for shadowing, inter-reflection, and other indirect illumination effects that are pervasive in real-world images.

To evaluate our method, we demonstrate our radiometric scene decomposition algorithm on a novel synthetic data set and a novel real-world data set. Our synthetic data set is created by combining a set of synthetic shapes, reflectance functions from the MERL database of BRDFs [32], and measured illumination environments [12]. Our real-world data set is created using a range of scene illumination (e.g., indoor and outdoor) with a variety of objects that span gamut of real-world materials in their reflectance (e.g., very specular, glossy, matte), texture (e.g., high-frequency texture, smoothly varying texture, no texture), and geometry (e.g., fine detail, convex objects, concave objects). We show that the proposed method successfully separates regions with different reflectance, accurately estimates the reflectance parameters and texture, estimates the illumination environment, and refines geometry from a small number of input images.

8.1 Radiometric Scene Decomposition

Our goal is to infer the spatially-varying reflectance, infer the natural illumination, and refine the geometry—in other words, radiometrically decompose a scene—from a set of RGB-D.
D images. To do this, we construct a Bayesian framework where we model the spatially-varying reflectance \( R \), natural illumination \( L \), and geometry \( G \) as random variables. We write the observations (i.e., the HDR pixel values) as \( I \) and use Bayes’ rule (with the assumption that reflectance, illumination, and geometry are independent) to write the posterior probability,

\[
p(R, L, G|I) = p(I|R, L, G) p(R) p(L) p(G).
\]  

(8.1)

This allows us to specify prior distributions on the reflectance, illumination, and geometry that capture the real-world variation of the variables.

We perform inference in this model by computing a maximum a posteriori (MAP) estimate. In practice, we alternate between maximizing the posterior with respect to the reflectance, illumination, and geometry.

### 8.2 Modeling Indirect Illumination

A major problem for reflectance estimation algorithms in the real world is indirect illumination. In this work, we consider indirect illumination to be light that arrives at a surface from other surfaces in the scene as opposed to light that arrives from the infinitely-distant illumination environment. Methods that ignore indirect illumination will incorrectly attribute interreflection or shadows with changes in the reflectance or geometry of a surface. This can cause important information about the illumination to be lost. For example, consider that we know the rough geometry and understand how shadows form—we can use shadowed regions to help us infer the main illuminant directions. Without this understanding of light transport, we become limited in the information we can extract from a scene.

To handle indirect illumination, we simply compute the predicted irradiance \( \hat{I}(p; R, L, G) \)
at pixel $p$, using an unbiased rendering algorithm. Formally, we write the predicted pixel value,

$$I_\lambda(p; R, L, G) = \int_v E_\lambda(t(e, v; G), -\omega_i; R, L, G) dv,$$

where $t(e, v; G)$ is the closest point on the surface of the geometry $G$ on the ray from eye position $e$ along viewing direction $v$, $E_\lambda(x, \omega_o; R, L, G)$ is the radiance from surface point $x$ to the camera, and $v$ is a set of directions passing through pixel $p$. The radiance from world point $x$ in direction $\omega_o$, $E_\lambda(x, \omega_o; R, L, G)$, is written,

$$E_\lambda(x, \omega_o; R, L, G) = \int_\Omega f_\lambda(x, \omega_i, \omega_o; R) L(\omega_i) v(x, \omega_i; G) d\omega_i + \int_\Omega f_\lambda(x, \omega_i, \omega_o; R) E_\lambda(x', -\omega_i; R, L, G) \tilde{v}(x, \omega_i; G) d\omega_i,$$

where $v$ is a visibility function that determines if ray beginning at $x$ in direction $\omega_i$ intersects the geometry, and $f_\lambda(x, \omega_i, \omega_o; R)$ is the BRDF at surface point $x$ with reflectance parameters $R$. This is a rewriting of the rendering equation [25] into a direct and indirect lighting component. We solve the integral using Monte Carlo Integration (i.e., using a path tracing algorithm). Importantly, we can take the derivative of these integrals with respect to the reflectance and illumination parameters in order to maximize the posterior using gradient-based algorithms. Appendix A derives the gradient of the error function with respect to the reflectance and illumination parameters and describes an efficient algorithm for computing them.
Figure 8.1: Example geometric bases. This figure visualizes the geometric bases \( g \) created by computing the eigenvectors of a similarity matrix of the triangles in the scene. By taking a linear combination of these bases, we can represent a large variety of segmentations of the geometry surface.

We can now write the likelihood function using a log-Laplace distribution centered on the predicted irradiance,

\[
p(I|R, L, G) \propto \prod_{p, \lambda} \exp \left[ - \frac{\log I_{p, \lambda} - \log \hat{I}_\lambda(p; R, L, G)}{b} \right]. \tag{8.4}
\]

Taking the logarithm of the predicted irradiance is important as it makes the likelihood function independent of the scale of the intensity values.

### 8.3 Spatial Segmentation

Thanks to the DSBRDF model, we can model real-world materials accurately and compactly. Real-world scenes, however, generally contain more than one reflectance function. In the most extreme case, a scene could have a unique BRDF at each surface point. The
right spatially-varying reflectance model will allow for the most expressibility (i.e., how many scenes it can represent) while limiting the dimensionality of the model for ease of inference.

We observe that there are typically only a handful of different base reflectance functions in a scene although the diffuse color of a surface may vary with high frequency. We can use this observation to develop a spatial segmentation model that dictates the reflectance function present at each surface point out of a small set of reflectance functions. Then, we can allow the diffuse color of the surface to vary within reflectance segments according to well-established priors on albedo. This scheme will allow us to model a large variety of scenes with a compact parameterization.

We construct our spatial segmentation model by creating an indicator layer, $S$, across the surface of the scene that dictates the reflectance function of each surface point. That is, $S_x$ indicates, from $1 \ldots K$, the reflectance function at surface point $x$. Rather than modeling the indicator layer directly, we can construct a set of hidden layers, one for each material, and then use the hidden layer whose value is greatest to determine the base material at that spatial point. More formally,

$$S_x = \arg \max_k M_{x,k}. \quad (8.5)$$

This formulation allows us to model the hidden material layers $M$ rather than directly modeling the indicator layer $S$. The main advantage of this approach is that we can constrain the hidden layers $M$ more easily than the indicator layers.

This model gives great expressive power at the cost of increased dimensionality. We alleviate this problem by expressing the hidden layers as a linear combination of geometric
basis functions,

\[ M_{x,k} = \sum_{n=1}^{N} g_{x}^{(n)} a_{k}^{(n)}, \quad (8.6) \]

where \( g \) are the \( N \) geometric basis functions and \( a \) are the coefficients. Figure 8.1 gives a visual description of the geometric basis system. Estimating the spatial scene segmentation is now simply a matter of estimating the geometric basis coefficients \( a \).

We want to ensure that the material regions are largely contiguous as they often are in real life. Therefore, we want geometric basis functions that vary smoothly across the surface. For this, we turn to the largest eigenvectors of the similarity matrix of the triangles in a scene.

The \( k \) largest eigenvectors of a matrix of a similarity matrix between triangles will give us a set of smoothly varying basis functions across the geometry of a scene. Furthermore, we want the bases to allow changes in material where we might expect two different materials to meet (e.g., in regions with sudden sharp curvature or regions with sudden changes in color). To satisfy these requirements, we can modify our distance metric between triangles to take into account the angle between adjacent triangles and the color of adjacent triangles. To compute the color of each triangle, we simply average the input pixel values of each triangle and use nearest neighbor to fill in triangles with no image pixels. We compute the distance between two adjacent triangles by taking a weighted sum of the geodesic distance between triangle centroids, the angle between the triangles, and the difference in color (in LAB color space) between the triangles. A distance matrix is computed using all-pairs shortest path. The similarity matrix is then computed,

\[ Z_{i,j} = \exp \left[ -\frac{d_{i,j}^2}{\sigma^2} \right], \quad (8.7) \]
This scheme will construct a set of geometric bases that prefers material region boundaries along geometric and photometric seams.

We must take care to initialize the material segments in a way that gets us as close to the correct solution as possible to minimize the time spent searching for the optimal solution. To accomplish this, we oversegment the input geometry into small clusters that share proximity, similar surface normals, and similar color in the input image. This is performed by spectral clustering on the similarity matrix described above. This scheme gives us a good starting point for the rest of the optimization.

One problem we may encounter in this formulation is that two different material regions may, in fact, be the same material. In the proposed formulation, there is no encouragement to consolidate the material regions and so using two material regions to represent one material will have at least the same likelihood as using one material region. To solve this problem, we must provide external motivation for similar regions to merge.

Our proposed solution is to manually attempt to merge neighboring material regions. If, after merging, the likelihood function is higher than the previous likelihood function (plus some small epsilon) then the merge is accepted. This step encourages the method to use the smallest number of material regions possible to accurately model a scene.

8.4 Spatially-varying Texture

We note that, while one material region may have a relatively constant specular behavior along a large surface area, the diffuse texture can change within material regions with high frequency. For example, a wooden desk will a generally uniform gloss from the finish although the grain of the wood will cause the diffuse color of the desk to vary. We must be able to model this variation in texture to fully capture real-world scenes.

To tackle this problem, we estimate a high-resolution texture map of the diffuse lobe
chromaticity of the DSBRDF model. Because the DSBRDF lobes are always ordered by their gamma curves, the chromaticity of the first lobe contains the diffuse color of the BRDF. We drop the dependency on color channel ($\lambda$) here for brevity and let $\tilde{c} = c_1$ be the diffuse texture that is accessed by a surface point $u$. This color value, $\tilde{c}_u$, is allowed to vary within material regions.

A problem with representing scenes so flexibly is that the diffuse texture can overfit the input images by representing specular highlights or fine changes in surface geometry. We must carefully constrain the texture values to prevent this.

We adopt two strategies found in prior work on estimating albedo. Previous work has found that the global distribution of albedo values has low entropy [2]. Therefore, we place an entropy prior on the diffuse texture. The entropy prior has the form,

$$p(\tilde{c}) = \exp[-H(\tilde{c})], \quad (8.8)$$

where $H$ is the entropy of texture map $\tilde{c}$. We compute the entropy using the continuous form of entropy and model the histogram of values using kernel density estimation so that we can take derivatives for gradient-based optimizers.

We also apply a strong spatial smoothness prior on the texture map,

$$p(\tilde{c}) \propto \prod_u \prod_v \mathcal{N} \left( \tilde{c}_u \left| \tilde{c}_v, \frac{1}{\exp \left( -\frac{||u-v||^2}{2\sigma^2} \right)} \right. \right), \quad (8.9)$$

where $\tilde{c}_u$ is the texture map, and $u, v \in \mathbb{R}^2$ are texture coordinates. This spatial smoothness prior enforces similar diffuse values strongly when the texels are close and decays for texels far apart. This is a strong spatial smoothness prior that helps prevent overfitting.
8.5 Natural Illumination

As in Chapter 7, we use a spherical panorama to represent the natural illumination of the scene and we use the natural image statistic prior and our novel entropy prior on the illumination. We also introduce a novel chromaticity prior that prefers unsaturated colors in the illumination map with a Dirichlet distribution,

\[ p(L) \propto \prod_{\theta, \phi, \lambda} \left( \frac{L_{\theta, \phi, \lambda}}{\sum_{\lambda'} L_{\theta, \phi, \lambda'}} \right)^{\alpha-1}, \]  

(8.10)

where \((\theta, \phi)\) is an illumination map pixel location and \(\lambda\) is a color channel. This prior helps enforce the grey-world assumption, in which we assume that, on average, the illumination is uncolored.

When initializing the illumination, our primary goal is to simply capture the scale of the scene illumination. To do this, we estimate a one-pixel illumination map first. We then initialize the full-sized illumination map with the value of this single pixel. This method gives us a good starting point for the general intensity of the scene that will speed up convergence of the method later on.

8.6 Reflectance-aware Illumination Estimation

It is important to consider how different objects in the scene impart different amounts of information about the illumination environment. For example, a perfect mirror will reflect all incoming light and preserve the information about the scene illumination. On the other hand, a Lambertian surface acts as a low-pass filter over the illumination environment, removing all of the high-frequency illumination information. We can take advantage of this fact to improve the illumination inference.

We improve our use of the image information by introducing a weight \(w_p\) in the error
term for each pixel \( p \) that is equal to the average of the \( \gamma \) curve of the most specular reflectance lobe of all surfaces within pixel \( p \). As we have noted, the \( \gamma \) value of the DSBRDF roughly corresponds to the specularity of the reflectance function. Therefore, pixels in the input image with more specularity will carry more weight in the error function. This scheme allows us to use the reflectance estimates to improve the illumination estimate.

### 8.7 Optimizing Rough Geometry

In this work, we are given a series of RGB-D images to infer radiometric properties of a scene. We then utilize a depth image fusion algorithm \([24, 35]\) to create a single geometric model of the scene. This gives us a good initial starting point, but this starting geometry is only approximate.

Depth image fusion algorithms improve the quality of the raw depth images by combining the information from several depth maps, but they are not without noise. There are two main sources of error when using a series of registered and combined RGB-D images: details finer than the resolution of the depth images are lost and large-scale low-frequency stitching errors that arise while fusing the data into a single geometric model. We use a targeted approach to deal with this noise.

To fix the large-scale low-frequency errors, we must modify the vertices of the geometry. Modifying all the vertices is a difficult because of the high dimensionality of the problem. Furthermore, the solution space must be constrained is some way so that the geometry does not self-intersect or otherwise violate physical constraints. We attempt to correct the large-scale low-frequency error by modeling the mesh vertices using a set of smoothly varying bases. We can actually reuse the geometric bases from the scene segmentation for this purpose.

We write the vertices as an initial vertex (i.e., the input geometry) \( V^{(0)} \) plus a linear
Figure 8.2: Evaluation of the importance of considering global light transport. Left: An image of a simple scene consisting of a sphere and a quad. Middle: Our algorithm understands interreflection and therefore correctly estimates the reflectance of the quad. Right: When interreflection is not modeled, the algorithm cannot correctly infer reflectance in this scene.

\[ V_x = V_x^{(0)} + \sum_{j=1}^{J} w_j g_j^{(j)} N_x, \]  

(8.11)

where \( x \) is a point on the surface, \( w \) is a set of coefficients that controls the influence of the geometric bases, and \( N_x \) is the surface normal at \( x \). This formulation allows the inference algorithm to “shrink” or “expand” pieces of the geometry so that they better match the input images.

8.8 Experimental Results

In this section, we describe the application of our method to synthetic and real scenes. First, we evaluate our method on a novel synthetic dataset so that we can quantitatively compare to ground truth. Next, we demonstrate our method on real scenes.

Synthetic data gives us a chance to evaluate estimates of our method directly to ground truth. We have created a dataset of synthetic scenes to evaluate our algorithm. Each scene consists of several synthetic shapes in complex configurations to imitate the challenges of real-world scenes. The scenes are rendered from a small set of unique views using PBRT.
Figure 8.3: Three example scenes of our synthetic dataset. The first four columns show the input images. The fifth column shows the illumination map used to render the images. The sixth column shows the ground truth geometry used to render the images. The seventh column shows the geometry after we artificially perturb it by adding random noise to the vertices and then smoothing. This process is repeated several times to make the geometric noise more low-frequency.

[40] and those images are used as input to our algorithm. This allows us to synthetically test our method on a variety of shapes and reflectance functions.

We use the NVIDIA OptiX ray tracing engine for both rendering the scene given the estimated reflectance, illumination, segmentation, texture, and geometry and for computing the gradients of the log posterior with respect to the reflectance, illumination, texture, segmentation, and geometry. Each of the following experiments consists of input data from three to four images with roughly 200 × 200 pixels. We found that each experiment took between a few days and a week to converge on an NVIDIA Tesla K40m. This can vary depending on the number of input images, the size of the images, and the number of assumed fundamental materials. This long running time is largely due to the computation of the gradient of the log posterior that can be time-consuming because of the large amount of variables to estimate (e.g., reflectance parameters for several regions, the pixels of the illumination map, the segmentation parameters, the texture, and the geometry).
Figure 8.4: Two Example Geometry Refinement Results. In this figure we show two results of geometry refinement on a synthetic scene. The first column shows the input image, the second column shows the predicted appearance of the scene before geometry estimation, the third column shows the predicted appearance of the scene after geometry estimation, the fourth column shows a difference image between the first column and the second column, and the fifth column shows a difference image between the first column and the third column. As shown, the geometry refinement process improves the rough initial geometry.

8.8.1 Importance of Interreflection

Figure 8.2 shows an experiment that demonstrates the importance of accounting for indirect illumination. The scene contains a specular quad that is reflecting light from an adjacent...
Environment 1

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>0.966</td>
<td>1.973</td>
<td>3.001</td>
</tr>
<tr>
<td>B2</td>
<td>0.954</td>
<td>1.970</td>
<td>2.004</td>
</tr>
<tr>
<td>B3</td>
<td>1.434</td>
<td>2.307</td>
<td>1.931</td>
</tr>
</tbody>
</table>

Environment 2

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>1.554</td>
<td>2.236</td>
<td>2.119</td>
</tr>
<tr>
<td>B2</td>
<td>1.069</td>
<td>2.062</td>
<td>1.865</td>
</tr>
<tr>
<td>B3</td>
<td>1.337</td>
<td>2.590</td>
<td>5.538</td>
</tr>
</tbody>
</table>

Environment 3

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>1.091</td>
<td>2.011</td>
<td>2.563</td>
</tr>
<tr>
<td>B2</td>
<td>1.260</td>
<td>1.812</td>
<td>1.697</td>
</tr>
<tr>
<td>B3</td>
<td>1.310</td>
<td>2.152</td>
<td>2.324</td>
</tr>
</tbody>
</table>

(a) Reflectance Error

Environment 1

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>Error</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>0.023</td>
<td>0.022</td>
<td>0.022</td>
<td>-32%</td>
<td>-36%</td>
</tr>
<tr>
<td>B2</td>
<td>0.020</td>
<td>0.019</td>
<td>0.019</td>
<td>-26%</td>
<td>-28%</td>
</tr>
<tr>
<td>B3</td>
<td>0.024</td>
<td>0.024</td>
<td>0.026</td>
<td>-30%</td>
<td>-25%</td>
</tr>
</tbody>
</table>

Environment 2

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>Error</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>0.024</td>
<td>0.025</td>
<td>0.023</td>
<td>-31%</td>
<td>-27%</td>
</tr>
<tr>
<td>B2</td>
<td>0.019</td>
<td>0.018</td>
<td>0.020</td>
<td>-31%</td>
<td>-32%</td>
</tr>
<tr>
<td>B3</td>
<td>0.024</td>
<td>0.024</td>
<td>0.031</td>
<td>-32%</td>
<td>-11%</td>
</tr>
</tbody>
</table>

Environment 3

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>Error</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>0.024</td>
<td>0.025</td>
<td>0.023</td>
<td>-31%</td>
<td>-27%</td>
</tr>
<tr>
<td>B2</td>
<td>0.019</td>
<td>0.018</td>
<td>0.020</td>
<td>-31%</td>
<td>-32%</td>
</tr>
<tr>
<td>B3</td>
<td>0.024</td>
<td>0.024</td>
<td>0.031</td>
<td>-32%</td>
<td>-11%</td>
</tr>
</tbody>
</table>

(b) Geometry Error

Figure 8.5: Quantitative evaluation of novel synthetic dataset. We have constructed a set of synthetic scenes that consist of three different “blob” meshes rendered with three different reflectance functions in three different illumination environments rendered from four viewpoints. We run our radiometric scene decomposition algorithm on each scene. Subfigure 8.5a on the left shows a table of log-space root-mean-square error of the inferred reflectance for each scene and on the right shows the three example scenes rendered with the estimated reflectance, illumination, and geometry from a novel view compared to ground truth. Subfigure 8.5b on the left shows a table of the error of the refined mesh for each scene and on the right shows the difference image between the scene rendered with the estimated reflectance, illumination, and geometry and the ground truth before and after geometry estimation. We can see that the geometry estimation greatly reduces the errors on the border of the geometry.

sphere. Our method accounts for indirect illumination and therefore correctly estimates the reflectance of the quad. When the indirect illumination computation is disabled, the quad is estimated as being diffuse instead. This experiment demonstrates the importance of considering indirect illumination for reflectance and illumination inference.
Figure 8.6: Results on Blob Experiments Under Three Illumination Environments. For each reflectance function and mesh, we show a novel view of the scene and use our inferred reflectance, illumination, and geometry to predict the appearance of the novel view. We also quantitatively evaluate the reflectance estimates by computing the log-space root-mean-square error compared to the measured BRDF. We also quantitatively evaluate the geometry by computing the distance between each vertex in the estimated mesh to the closest point on the ground-truth geometry and average all distances. We also show the percentage the geometry error is reduced by our method in parentheses (a negative number indicates a reduction in error). The general low-frequency quality of the illumination environment eases the accurate inference of reflectance and illumination in these scenes.

8.8.2 Evaluation of Geometry Refinement on Synthetic Blob Dataset

In this section, we show detailed qualitative and quantitative results on a set of three different meshes rendered with three different MERL BRDFs under three different illumination environments (27 total scenes). In these experiments, we artificially perturb the geometry of each mesh by adding noise and then smoothing the vertex positions. We then use our method to estimate the reflectance and illumination and refine the geometry.
Figure 8.7: Results on a synthetic scene. Subfigure 8.7a shows the input images. Subfigure 8.7b shows comparisons to ground truth. The first column shows the estimated appearance using the inferred reflectance, illumination, and segmentation of the scene, the inferred illumination, and the estimated segmentation map. The second column shows the input appearance and ground truth appearance from a novel view, ground truth illumination, and ground truth segmentation. Subfigure 8.7c shows the inferred reflectance functions visualized as a series of spheres rendered with varying point-light directions.

Figure 8.3 shows 3 of the 27 synthetic scenes. For each experiment, we use the four input images and the artificially-perturbed geometry as input to our method. As shown,
each of the meshes, reflectance functions, and illumination environments are distinct and therefore provide a good range of inputs to evaluate our method.

Figure 8.5 shows detailed quantitative results of these scenes. Figures 8.4 and 8.6 show detailed qualitative results. To evaluate our results, we give a visual comparison of a ground truth novel view compared to a rendering of that view using the inferred reflectance, illumination, and geometry of our method. To quantitatively evaluate the reflectance, we compute the log-space root-mean-square error between the estimated reflectance and the
Figure 8.9: Results on a real scene. Subfigure 8.9a shows the three input images. Subfigure 8.10b shows intermediate results of the algorithm: after reflectance and illumination estimation, after texture estimation, and after geometry estimation. It’s important to note that our method correctly infers the specular highlights rather than using the diffuse texture to model them. Subfigure 8.9c shows the predicted appearance from a novel view and geometry, illumination, and segmentation estimates. The first column shows the estimated appearance using the inferred reflectance, illumination, and segmentation of the scene, the refined geometry, the inferred illumination, and the inferred segmentation map. The second column shows the ground truth appearance from a novel view, the input geometry, the ground truth illumination. Subfigure 8.9d shows the inferred reflectance functions visualized as a series of spheres rendered with varying point-light directions. Our method correctly recognizes the bright specular highlight on the table and uses other highlights in the scene to recover a detailed illumination map.
ground-truth reflectance,

\[ E_{\ell-RMS} = \sqrt{\frac{\sum_{\theta_h, \theta_d, \phi_d} (\log f(\theta_h, \theta_d, \phi_d) - \log \hat{f}(\theta_h, \theta_d))^2}{N}}, \]  

(8.12)

where \( f \) is the ground-truth BRDF, \( \hat{f} \) is the BRDF to be evaluated, and \( N \) is the number of samples of the ground-truth BRDF. To quantitatively evaluate the geometry, we compute

---

**Figure 8.10:** Results on another real scene. Results are presented as in Figure 8.9.
the distance between each vertex in the estimated mesh to the closest point on the ground-truth geometry and average all distances. We also show percentage the error was reduced by our geometry refinement compared to the initial geometry error.

From the qualitative results, we can see that our method is able to capture reflectance well, especially when the true reflectance of the scene does not feature sharp specular high-

Figure 8.11: Results on another real scene. Results are presented as in Figure 8.9.
lights or the illumination environment does not have many high-frequency illuminants or the geometry is smooth. When the object is highly specular and the illumination environment has high-frequency detail, the initial error in the geometry causes the reflectance and illumination inference to be trapped in a local minimum. In a majority of the experi-
ments, the geometry error is reduced by a large margin—on average, by 25%. This indicates that our algorithm is using the rich radiometric information encoded in the appearance to successfully refine geometry.

8.8.3 Decomposition on Synthetic Scenes

Figure 8.7 shows results on an example synthetic scene. This scene consists of several teapots in a room with a textured floor. This tests the ability of our algorithm to infer reflectance and illumination from relatively simple scenes. The results show that we infer plausible reflectance functions for each of the teapots and an accurate illumination map. Our method is also able to capture the texture of the floor.

Figure 8.8 shows results on a final synthetic scene. This scene contains a single object that has many different materials in a complex natural illumination environment. Note that inaccuracies on the top of the car and on the front windshield are because of the limited data of those regions in the input images. The results show that we are able to successfully estimate plausible reflectance functions and infer the salient features of the illumination map.

8.8.4 Decomposition on Real Scenes

Although synthetic scenes are a good test bed for our method, our ultimate goal is inference in real-world scenes with real-world conditions. We have collected a small dataset of N real world scenes. In each scene we capture the geometry of the scene using the Microsoft Kinect sensor and acquire high dynamic-range (HDR) images of the scene from several different viewing directions. We use the Kinect Fusion algorithm [24, 35] to combine multiple depth images into a triangle mesh that is used as the input geometry. We also capture the ground truth illumination using a mirror ball placed in the center of the scene.
Figure 8.9 shows results on a real scene. Note that this scene would be challenging if the reflectance was assumed to be a diffuse reflectance model as the light reflecting off the table opposite the viewer would be estimated erroneously as a change in diffuse texture. Our framework is able to correctly deduce that this is due to specular reflectance while also correctly estimating the texture of the table. We use our inferred reflectance, illumination, and geometry to predict the appearance of the scene from a novel view that looks very similar to the ground truth. The estimated segmentation map is correct except for the table being represented by two separate materials because the segmentation map has no incentive to collapse two similar materials when it can represent it equally well as two materials. In this scene we are also able to capture the texture detail on both the table and the coffee mug which would not be possible without explicitly modeling texture.

Figure 8.10 shows results on another real scene. This scene is more difficult than the last as it contains increased diffuse texture and objects with finer geometric detail. Despite this difficulty, our method is still able to reasonably segment the scene although the table is split into multiple different regions. Note that the shadowed regions in front of the penguin and salt shakers and the specular highlights on the table and wooden bowl are estimated correctly and not as being part of the diffuse texture. Our method is able to distinguish these regions because we consider complex reflectance functions and illumination and model the indirect illumination effects between objects. The solid white regions of the image exist because geometry has not been captured in those regions. We are also able to capture the texture on the penguin, salt shakers, table, and wooden bowl. The texture is faint in some regions, which is likely due to the strong spatial smoothness prior we place on the texture map. Note that we compare to a novel view of the scene where many of the surfaces are not seen at all in the input images (e.g., the back of the penguin and the salt shakers). Because
of this, we don’t recover fine texture details in those areas but we are able to extrapolate the reflectance of those unseen regions because of the spatial segmentation model that leverages the geometric bases.

Figure 8.11 shows results on another real scene. This scene contains an object with a metallic-like reflectance function. This type of complex reflectance function would be difficult to model accurately with a combination of simple models such as Lambertian plus Torrance-Sparrow. Additionally, there is a great deal of interreflection of other objects in the piggy bank object. Despite this, we successfully infer the reflectance of the piggy bank object and the main light sources in the illumination environment allowing us to accurately predict the appearance of the scene from a novel view. Again, we compare to a novel view where some of the surfaces are not seen in any input image yet we are able to correctly extrapolate the extents of the material regions because of the geometric bases. Note that the initial geometry captured by the Kinect has significant inaccuracies (e.g., there are missing portions of the ear of the piggy bank) that make it difficult for our method to refine the geometry because the difference is so great. The segmentation map correctly separates the objects although it uses several materials to compose the table.

Figure 8.11 contains the most geometrically complex arrangement of objects in our dataset. Our method correctly infers the illumination as being a cluster of lights located behind the viewer although finer details are not captured because there are no highly-specular materials in the scene. We also recover plausible reflectance functions, such as the dull specular highlight on the black lamp, that allow us to predict the appearance of the scene from a novel view. Fine texture detail, such as the text on the book, is captured although there is some noise present in the diffuse albedo. This is mostly in spots where the surface is view by few or no input images, and may also be due to small errors in the
intrinsc or extrinsic camera parameters.

8.9 Conclusion

In this final chapter, we have presented a method for decomposing a scene into its radiometric elements (reflectance, illumination, and geometry) given rough geometry and images of that scene from several angles. We demonstrated that our model is capable of modeling the large variation of real-world scenes but remains tractable during inference through our priors. We showed that modeling complex reflectance behavior and natural illumination is important for extracting the most information possible from RGB-D images. We developed a spatial segmentation model that models large contiguous material regions and allows the diffuse texture to vary across the surface. We developed a geometric refinement scheme to optimize the initial geometry provided by an RGB-D sensor. We utilize an unbiased renderer to represent the image formation process, allowing us to model indirect illumination effects that exist in the real world. We combined these contributions using a Bayesian formulation that enables full, real-world radiometric scene decomposition.
Chapter 9: Conclusion

Appearance is the product of incident light, geometry, and reflectance. Decomposing a scene into these radiometric components is an important task for understanding images in the real world. In this thesis, we have created a method for accurate inference of reflectance in single and multi-material scenes, under point-light and natural illumination given known or approximately-known geometry from one or more images. We have accomplished this through an important set of contributions.

First, we developed a reflectance model that is capable of representing the complex reflectance behavior that exists in the real world. We did this by thinking about the BRDF as a probability distribution of exitant light rays and by modeling that probability distribution with a novel hemispherical exponential power distribution. We fit our novel DSBRDF model to a set of measured BRDFs and used the database to derive a set of reflectance bases that encode the primary modes of variation of real-world BRDFs. This enabled us to achieve a more compact modeling of reflectance than a mixture of Lambertian and Cook-Torrance model while realizing superior expressiveness. We used our reflectance bases to construct a novel reflectance prior that succinctly encodes the distribution of real-world BRDFs. We showed that the DSBRDF can overcome ambiguities in the inverse rendering problem because of the compact but expressive representation and our data-driven reflectance priors.

We then tackled the problem of estimating reflectance in scenes with more than one unique reflectance function. We leveraged the observation that scenes typically have a small number of “fundamental” reflectance functions. This observation was used to construct a
spatial segmentation model that dictates, for each pixel, the fundamental material present at that pixel using a set of latent MRF layers. We placed a smoothness prior distribution on the latent layers to encourage the recovery of reflectance regions like those in the real world. Our carefully constructed spatial segmentation model enabled us to tractably solve the combinatorial problem of determining the spatial extent of each reflectance function.

We also tackled the difficult problem of inferring reflectance under natural illumination environments. We utilized a non-parametric natural illumination model alongside strong priors to enable the tractable inference of reflectance and illumination in the real world. We noticed that a panoramic illumination environment is a natural image that we can place natural image statistic priors on. We also explored how the reflectance function will increase the entropy of the reflected radiance, and how the illumination must have low entropy to compensate for this increase. These priors made it possible for us to recover rich natural illumination environments from only a single image and with known geometry.

Finally, we developed a method for inferring complex reflectance, inferring natural illumination, and refining geometry in multimaterial scenes from RGB-D data. To accomplish this, we first extended our spatial segmentation model to the 3D geometry and created a compact parameterization of it using a set of geometric bases. Next, we proposed a texture estimation method that allows us to recover the fine diffuse texture present on many surfaces (e.g., wood grain). We then created a technique for refining the initial geometry acquired through RGB-D sensors by accounting for the low-frequency stitching errors that are created through the RGB-D fusion process. Lastly, we proposed a way to handle indirect illumination, shadowing, and interreflection effects that occur between objects in real-world scenes by directly simulating it. We showed that modeling this indirect illumination is critical for correctly estimating reflectance and illumination in a generative framework. We
then demonstrated that our method successfully recovers complex reflectance, natural illumination, and refines geometry in scenes from RGB-D.

9.1 Limitations

This thesis is an important step towards a method for the complete radiometric understanding of scenes. There are, however, some limitations to the proposed method.

First, we have made some simplifying assumptions about the nature of reflectance. To ease the computation burden, we assume that surface reflectance can be modeled with the Bidirectional Reflectance Distribution Function (BRDF). This necessarily eliminates our ability to handle any transmissive surfaces, like glass or clear plastic. We also assume that the light can be modeled with geometric optics. This makes it impossible to simulate the quantum effects of diffusion or interference. Although this limits the applicability of the method somewhat, these assumptions hold true in the majority of real-world scenes.

We also note that the method is unable to handle atypical scenes that display high-frequency changes in fundamental reflectance behavior. For example, a paint embedded with metal flakes would have constant changes in reflectance across its surface. The geometric bases used to model the spatial segmentation, being smoothly varying, cannot model this high frequency change in reflectance.

Another limitation is the assumption that the only sources of light emission come from an infinitely-distant source. This is applicable to most scenes depending on the scale of the scene. For example, if the scene rests entirely upon a small table, we can assume that the ceiling lights or windows are infinitely distant. This assumption can be violated, however, when the scene itself contains a source of light. For example, if a lamp is placed on the table and turned on, our proposed method will not understand that part of the scene itself is emitting light. Instead, it would likely assume that the bulb of the lamp was reflected
more light than it receives. Fortunately, the proposed framework could easily incorporate the possibility of objects in the scene emitting their own light by introducing an emissivity attribute over the geometry of the scene.

The geometry estimation method has some limitations as well. Our approach to geometry estimation is to slightly expand or contract regions of the geometry to best match the input images. The goal of this approach is to resolve some of the low-frequency stitching errors that arise from depth map fusion algorithms. Because of these assumptions, our method cannot complex structures that were not captured in the original geometry which can frequently happen if the depth sensors return erroneous data (this can happen depending on the infrared reflectance properties of the surface). A more complex geometric refinement model is necessary to overcome this problem.

9.2 Future Work

As RGB-D sensors grow more ubiquitous and mobile hardware becomes more powerful, we believe that the work in this thesis will become more useful in time. A popular area of mobile development is augmented reality, where virtual objects are added to real image or video. To realistically add digital objects, it is necessary to properly illuminate and occlude the digital objects based on the true illumination environment and scene geometry. After applying our method, inserting virtual objects becomes simple because we have radiometrically decomposed the scene into reflectance and illumination.

Virtual tourism is an exciting area that allows users to explore real-world places from the comfort of their home. Providing an immersive virtual tourism experience relies on accurately decomposing a scene radiometrically so that it can be re-rendered for a user to explore virtually. To accomplish this task, we could extend the scale of the method proposed in this thesis to handle many input images and very large-scale scenes. Currently,
the computational complexity of the method in this thesis prohibits the use of large-scale scenes like those that would be typical for virtual tourism.

We would also like to see the proposed method applied in the robotics community. In the future, robots will require an understanding of the world not only through objects but also through material information. That is because interacting with objects requires a sense of the tactile properties of the object (e.g., softness, hardness, or fragility). It’s possible that reflectance information could help computers make inferences about the physical properties of materials.

Finally, we believe that the radiometric scene decomposition method proposed in this thesis should be incorporated into the problem of complete scene understanding (i.e., the understanding of all properties of a scene—radiometric information, object type, number, and identity, location information, etc.). This is because of correlations between the types of objects one might find in a scene, the reflectance of those objects, the setting of the scene, and how the scene is illuminated. For example, there is some pattern in the types of objects one finds outside (e.g., trees and buildings are more likely than kitchen items). There is likely also a pattern to the illumination environment one would find outside (e.g., it is likely illuminated by the sun and the sky). By attempting to infer radiometric properties of a scene in addition to the types or identities of objects, we could exploit these different relationships through sophisticated priors.
Bibliography


Appendix A: Important Gradients for Optimization

Throughout this thesis, we have relied on computing a maximum a posteriori estimate of our Bayesian framework to infer reflectance and illumination. In practice, this often means minimizing the negative log posterior using a gradient-based optimizer. In order for these optimizers to work, we must supply a function that computes the gradient of the error function for the current reflectance, illumination, and geometry estimate. This function is used by gradient-based optimizers to quickly minimize the error function. In this chapter, we provide gradients of the error function with respect to the reflectance and illumination parameters and an algorithm for computing them.

A.1 Gradient of the Negative Log Likelihood

In practice, we minimize the negative log posterior, which is a sum of the negative log likelihood and the negative log of the prior distributions. In this work, we write the negative log likelihood as a sum of absolute values,

\[-\log p(I|R, L, G) = \sum_p \left| I_p - \hat{I}(p; R, L, G) \right|,\]

where \(p\) is an image pixel, \(I_p\) is the value of pixel \(p\) in the input image, \(\hat{I}\) is a function that predicts the pixel value at pixel \(p\) given the current parameters of the reflectance \(R\), illumination \(L\), and geometry \(G\). We write the predicted pixel value,

\[\hat{I}_\lambda(p; R, L, G) = \int \mathcal{E}_\lambda(t(e, v; G), -\omega_i; R, L, G)dv,\]
where \( t(e, v; G) \) is the closest point on the surface of the geometry \( G \) on the ray from eye position \( e \) along viewing direction \( v \), \( E_\lambda(x, \omega_o; R, L, G) \) is the radiance from surface point \( x \) to the camera, and \( v \) is a set of directions passing through pixel \( p \). This is simply an integration over the pixel \( p \). The radiance from world point \( x \) in direction \( \omega_o \), \( E_\lambda(x, \omega_o; R, L, G) \), is written,

\[
E_\lambda(x, \omega_o; R, L, G) = \int_\Omega f_\lambda(\omega_i, \omega_o; R, x)L(\omega_i)v(x, \omega_i; G)d\omega_i + \int_\Omega f_\lambda(\omega_i, \omega_o; R, x)E_\lambda(x', -\omega_i; R, L, G)\bar{v}(x, \omega_i; G)d\omega_i,
\]

where \( v \) is a visibility function that determines if ray beginning at \( x \) in direction \( \omega_i \) intersects the geometry, and \( f_\lambda(\omega_i, \omega_o; R, x) \) is the BRDF at surface point \( x \) with reflectance parameters \( R \). This is a rewriting of the rendering equation [25] into a direct and indirect lighting component. Note that here we assume that the foreshortening term \( \cos \theta_i \) is absorbed by the BRDF \( f_\lambda \) for simplicity.

Crucially, we rewrite the irradiance \( E \) with an expression that easily translates into an iterative path tracing algorithm,

\[
E(x, \omega_o; R, L, G) = \sum_i \prod_j f(\omega_i^{(j)}, \omega_o^{(j)}; x_j)D(x, \omega_o; R, L, G),
\]

where

\[
D(x, \omega_o; R, L, G) = \int_\Omega f_\lambda(\omega_i, \omega_o; R, x)L_i(\omega_i)v(x, \omega_i; G)d\omega_i,
\]

is the direct lighting integral, and \( \omega_i^{(j)}, \omega_o^{(j)}, x_j \) are the incident, exitant, and collision point

---

**Appendix A: Important Gradients for Optimization**
of the $j^{th}$ intersection point on the light path. This better represents the steps in an
erative path tracing algorithm: we first shoot a light ray to the surface, keeping track of
an attenuation parameter which is the product of all BRDFs seen so far, and then we solve
for the direct lighting component which is the integral, before continuing the path tracing
by shooting a ray to the next surface. This form will make it simpler to take gradients with
respect to the variables $R$, $L$, and $G$ and incorporate it into an iterative path tracing-like
algorithm to compute the gradients.

We can write the partial derivative of the negative log likelihood,

$$\frac{\partial}{\partial R} - \log p(I|R, L, G) = - \sum_p \left( \frac{I_p - I\lambda(p; R, L, G)}{|I_p - I\lambda(p; R, L, G)|} \right) \frac{\partial I\lambda(p; R, L, G)}{\partial R},$$  \hspace{1cm} (A.6)

where $R$ represents all reflectance parameters (whatever the model may be). We then write
the derivative of the predicted pixel value $\hat{I}$,

$$\frac{\partial I\lambda(p; R, L, G)}{\partial R} = \int_v \frac{\partial E\lambda(t(e, v; G), -\omega_i; R, L, G)}{\partial R} dv,$$  \hspace{1cm} (A.7)

which is simply an integration over the pixel $p$.

We can now write the derivative of the radiance $E$,

$$\frac{\partial E(x, \omega_o; R, L, G)}{\partial R} = \sum_{i=0}^\infty \left( \frac{\partial \prod_j f(\omega_i^{(j)}, \omega_o^{(j)}; x_j)}{\partial R} D(x, \omega_o; R, L, G) + \prod_j f(\omega_i^{(j)}, \omega_o^{(j)}; x_j) \frac{\partial D(x, \omega_o; R, L, G)}{\partial R} \right).$$  \hspace{1cm} (A.8)

Appendix A: Important Gradients for Optimization
It’s important to write the derivative of the product of the BRDFs as a product of two factors,

\[
\frac{\partial \prod_j f \left( \omega_i^{(j)}, \omega_o^{(j)}; x_j \right)}{\partial R} = \left( \prod_j f \left( \omega_i^{(j)}, \omega_o^{(j)}; x_j \right) \right) \left( \sum_j \frac{\partial f \left( \omega_i^{(j)}, \omega_o^{(j)}; x_j \right)}{\partial R} f \left( \omega_i^{(j)}, \omega_o^{(j)}; x_j \right) \right). \tag{A.9}
\]

This rearrangement allows us to easily see how to compute the derivative in a path tracing algorithm: we can iteratively compute the new sum term, which is the derivative of the attenuation factor, and the new integral term can be computed just like the direct lighting integral.

Algorithm A.1 gives the steps of this efficient algorithm to compute the gradient of reflectance parameters considering indirect illumination. Note that it has the same asymptotic complexity as path tracing. We can use the same algorithm for taking the derivative with respect to the surface normals of the surface.

Computing the gradient of the error function with respect to the illumination parameters is simpler than the reflectance function. This is because the illumination only appears inside the direct lighting integral. We can therefore write the gradient of the error function with respect to the illumination parameters,

\[
\frac{\partial E(\mathbf{x}, \omega_0; \mathbf{R}, \mathbf{L}, \mathbf{G})}{\partial \mathbf{L}} = \sum_{i=0}^{\infty} \prod_j f \left( \omega_i^{(j)}, \omega_o^{(j)}; x_j \right) \int \omega_i f \left( \omega_i, \omega_o^{(i)}; x_i \right) \frac{\partial L(\omega_i)}{\partial \mathbf{L}} \tilde{v}(\mathbf{x}_i, \omega_i) d\omega_i. \tag{A.10}
\]

Numerically computing the gradient requires Monte Carlo integration of the direct lighting integral and the gradient of the direct lighting integral. Multiple importance sampling is
Algorithm 1: Compute Predicted Pixel Irradiance Gradient with Respect to Reflectance

attenuation[1 × 3] ← 1
∂attenuation/∂R[|R| × 3] ← 0
radiance[1 × 3] ← 0
∂radiance/∂R[|R| × 3] ← 0

let r be a ray from the camera location through the pixel being shaded
while true do
    let x be the intersection point when ray r is traced into the scene
    if no surface is hit then
        break
    directlighting[1 × 3] ← 0
    ∂directlighting/∂R[|R| × 3] ← 0
    for 1 ... nsamples do
        s ← random ray direction
        if ray s hits the illumination environment then
            directlighting ← directlighting + f(r, s; x) * L(s)/p(s)
            ∂directlighting/∂R ← ∂directlighting/∂R + ∂f(r, s; x)/∂R * L(s)/p(s)
        radiance ← radiance + attenuation * directlighting
        ∂radiance/∂R ← ∂radiance/∂R + attenuation * (∂attenuation/∂R * directlighting + ∂directlighting/∂R)
        r' ← random ray direction
        attenuation ← attenuation * f(r', r; x)
        ∂attenuation/∂R ← ∂attenuation/∂R + f(r', r; x)/∂R f(r', r; x)
        r ← r'

Figure A.1: Compute Predicted Pixel Irradiance Gradient with Respect to Reflectance. This algorithm is a modification of the path tracing algorithm to keep track of the gradient with respect to reflectance.

typically used to compute the direct lighting integral by separating sampling from a BRDF proposal distribution and an illumination proposal distribution and combining the samples. We use the same technique for computing the gradient of the direct lighting integral. When computing the gradient of the direct lighting integral with respect to the BRDF, we use the same BRDF proposal distribution. When computing the gradient of the direct lighting integral with respect to the illumination, the illumination proposal distribution is changed to only sample rays in the direction of the illumination map pixel $L_{θ,φ}$ whose derivative is being taken.

Appendix A: Important Gradients for Optimization
A.2 Gradient of the DSBRDF Model

To compute the gradient of the negative log likelihood with respect to the reflectance parameters, we must be able to take the derivative of the BRDF value with respect to the reflectance parameters. We can write the gradient of the DSBRDF model with respect to the DSBRDF coefficients,

$$\frac{\partial f_\lambda(\theta_h, \theta_d)}{\partial \Psi} = \sum_r c_{r,\lambda} \exp \left[ \kappa_r(\theta_d) \cos \gamma_r(\theta_d)(\theta_h) \right] \cos \gamma_r(\theta_d)(\theta_h) \left( \frac{\partial \kappa_r(\theta_d)}{\partial \Psi} + \kappa_r(\theta_d) \frac{\partial \gamma_r(\theta_d)}{\partial \Psi} \log \cos \theta_h \right).$$

The derivatives of $\kappa$ and $\gamma$ are written,

$$\frac{\partial \kappa_r(\theta_d)}{\partial \Psi} = \kappa_r(\theta_d) b(\theta_d; \kappa, r),$$

$$\frac{\partial \gamma_r(\theta_d)}{\partial \Psi} = \gamma_r(\theta_d) b(\theta_d; \gamma, r),$$

where $b$ are the data-derived DSBRDF bases.
Vita

Stephen Lombardi received his Bachelor of Science in Computer Science from The College of New Jersey in 2009 and his Masters of Science in Computer Science from Drexel University in 2012.

Selected Publications

- Stephen Lombardi, Ko Nishino, Yasushi Makihara, Yasushi Yagi, “Two-Point Gait: Decoupling Gait from Body Shape,” 14th International Conference on Computer Vision, December 2013