Wide-Baseline Light Fields – Imaging and Applications

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Michael De Zeeuw

B.S., Electrical and Computer Engineering, Calvin College

Carnegie Mellon University

Pittsburgh, PA

May 2024

Acknowledgments

This thesis is the culmination of years of work, throughout which I have been blessed to be surrounded by a network of support, both personal and professional.

First off, I would like to thank my advisor, Prof. Aswin Sankaranarayanan. Back in 2017, I was an undergraduate student at Calvin College looking for a research internship to start looking ahead to my goals of graduate school. Thanks to a Calvin alumnus who is faculty at Carnegie Mellon, Prof. Shawn Blanton, I was connected to CMU and found Aswin's lab doing research that greatly interested me. Though he did not have a history of taking on undergraduate research interns from other institutions, Aswin agreed to have me for the summer, and this has led to a rewarding research experience for me as part of his lab. His keen eye for novel ideas and clever solutions has guided me through my years of research, and his evident aim for both research success and personal well-being is greatly appreciated. Thanks for everything.

Second, I would like to thank my committee: Vijayakumar Bhagavatula, Srinivasa Narasimhan, and Wolfgang Heidrich. Their keen insight during my prospectus of this thesis has led to a far better result, and I appreciate their time and thought that has contributed to this work.

Next, I would like to thank all members of the Image Science Lab, past and present, for being a source of both inspiration and support throughout my time in the lab. I am constantly amazed by the work that is being done by my peers, and they have inspired me to work harder and think creatively to match.

This dissertation was supported by the National Science Foundation (grant numbers 1801382 and 1652569), the National Geospatial Agency's Academic Research Program (award no. HM0476-1-22-0004), the Sony Faculty Innovation Award, and the ARCS Foundation (Pittsburgh Chapter Award). I am grateful for their support of my doctoral work.

I would also like to thank the community of friends that I have found here in Pittsburgh for their continued support over the years. To my small group, I could not have asked for a better group of friends that have constantly supported and encouraged me when things were difficult and celebrated with me when things went well.

To my family, thank you for your unending support, both through this Ph.D. and in the years that led me here. My parents and siblings have always had my back and have been instrumental in getting me to this point.

Finally, I could not have done any of this without the support and care of my wife, Cailin. She has seen me through the thick and thin of every step of the way. And for our daughter, I do this all for you.

Abstract

The world around us is filled with complex visual phenomena that we may wish to understand or recreate. Iridescence, intricate geometry, and complex spatial texture are all examples of such phenomena that can be challenging to both capture and represent. In order to characterize them, we need to consider how light interacts with different spatial points over a broad range of incoming and outgoing directions. Light field cameras provide a framework for capturing this spatial and angular information in a single image exposure; however, traditional hand-held light field cameras only observe a small fraction of the cone of light emitted by a given scene point. As a consequence, the study of interesting angular effects like iridescence are beyond the scope of such cameras.

The aim of this thesis is to present a novel light field imaging device capable of measuring over a wide baseline and to explore the space of applications that would significantly benefit from such light field data. We motivate why and how wide-baseline light fields (WBLFs) open up many exciting new capabilities in light field processing, 3D shape reconstruction, and iridescent reflectance capture.

The core contribution of this thesis is the imaging system design for a WBLF camera. We achieve a wide-baseline by imaging the scene indirectly through an ellipsoidal mirror. The combination of a light field camera and the ellipsoidal mirror provides rich measurements in space and angle. We further analyze the captured light fields to understand the system's capabilities, including considerations of resolution, angular range, coverage, and depth of field. Additionally, we develop a novel calibration procedure to account for the unique challenges present in such an imaging system.

We show this WBLF camera in action through a set of different applications that benefit from our data. First, we develop a suite of geometric processing algorithms to provide a parallel to the techniques employed in traditional light field imaging, including viewpoint synthesis, refocusing, and shape reconstruction. In particular, our shape reconstruction technique leverages the unique benefits of the WBLF camera to jointly estimate 3D shape and surface normals. Second, we develop a network similar to state-of-the-art 3D scene representations to train over WBLF data and produce 3D shape and reflectance reconstructions for small objects with varied appearance. Finally, we show that our system is capable of measuring the high dimensionality of the spatially-varying bidirectional reflectance distribution function (SVBRDF) efficiently. We additionally develop a neural SVBRDF representation to render the captured data over a broad range of novel viewpoints, illuminations, and geometries.

In total, these contributions provide a significant advance to light field imaging by establishing the foundation for wide-baseline light field imaging and its applications.

Contents

List of Figures

List of Tables

1 Introduction

As our eyes observe the world around us, we see a significant amount of visual complexity and diversity. From a bevy of vibrant colors to ever-changing visual appearances as we move through the world, the light that we observe contributes to our understanding and admiration of the things that we see. The nature of this interaction with light for many things in this world is often highly complex and can reveal interesting characteristics of the objects involved. What if we could design a camera that could effectively capture this complexity to add to our understanding of the world?

Everything we see is constantly interacting with light in interesting ways; light traversing a scene reaches a point in the scene and is then redistributed according to that point's material properties and geometric information. The angular cone of light that is emitted by that point often captures rich insights into the nature of this interaction with light. From a fairly mundane encoding of the surface's reflectance, be it Lambertian or specular, to more exotic phenomena like iridescence caused by diffraction, thin-film interference, or structural coloration, measuring the radiance of light over a large solid angle—perhaps, even a hemisphere—can play a pivotal role in understanding shape and reflectance.

Standard 2D cameras only capture one perspective of each scene point's reflectance, which does not allow for a full understanding of the scene. Recently, light field cameras have been developed to sacrifice some amount of spatial resolution in exchange for sampling in the angular domain. However, such designs only measure a small portion of the cone of light emitted by a scene point. For example, a camera observing a scene point that is a meter away from a lens with a light gathering diameter of 5 cm measures a small cone that has an angular spread of $\sim 3^\circ$. Even using a powerful microscope objective with a numerical aperture of, say, 0.5 only allows us to measure a cone of light with a spread of 60°. A multi-camera light stage, or perhaps a kaleidoscope, can be used to surround the object of interest to capture a larger cone of angles; yet, the number of measurements per scene point scales linearly in the number of cameras/viewpoints, which results in a sparse measurements of light rays over the angular cone of interest. Hence, current systems are woefully inadequate when it comes to providing a dense

measurements of the wide angular cone of light emitted by a scene.

This thesis proposes a novel design for the acquisition of wide-baseline light fields, i.e., a system that measures a large fraction of the cone of light emitted by scene points. We achieve this by using a standard light field camera that observes the scene of interest, not directly, but *through an ellipsoidal mirror*. This design relies on the observation that the ellipsoidal mirror morphs the wide cone of light from one of its foci to a much smaller cone at the other focus point; hence, placing the scene at one focus and the camera at the other allows us to acquire a wide-baseline light field. We show that the design is, in principle, capable of obtaining a larger cone of light than prior designs, often close to a hemisphere for many scene points. The measurement of wide-baseline light fields can significantly enhance conventional uses of such signals while enabling a bevy of new capabilities.

1.1 Thesis Contributions

The aim of this thesis is to present a novel light field imaging device capable of measuring over a wide baseline and to explore the space of applications that would significantly benefit from such light field data. Most literature on light field capture and processing is focused on narrow-baseline imaging systems, as these systems present the most common form of single-shot light field capture. However, this thesis presents motivation for why and how wide-baseline light fields open up many exciting new capabilities in light field processing, 3D shape reconstruction, and reflectance estimation.

In particular, this thesis makes the following contributions, highlighted in Figure 1.1:

- **Wide-Baseline Light Field Camera Design (Chapter 3).** We propose a design for a catadioptric device using a light field camera and an ellipsoidal mirror for capturing wide-baseline light fields. We analyze our device to consider the sampling, resolution, and coverage of our captured light fields. We also consider design variations that extend the capabilities of our system, including additions for illumination control and extended angular range. The use of an ellipsoidal mirror provides unique challenges for calibrating our system, so we develop a suite of novel calibration procedures in Chapter 4 to achieve our desired results.

- **Geometric Light Field Processing from WBLF Data (Chapter 5).** We demonstrate that our captured light fields can be processed using conventional light field processing techniques to refocus the scene at different depths, reconstruct the shape of objects, and generate synthetic images from novel viewpoints. Additionally, the larger angular cone can dramatically enhance these capabilities by providing extremely shallow depth of fields and the ability to image the backside of an opaque ob-

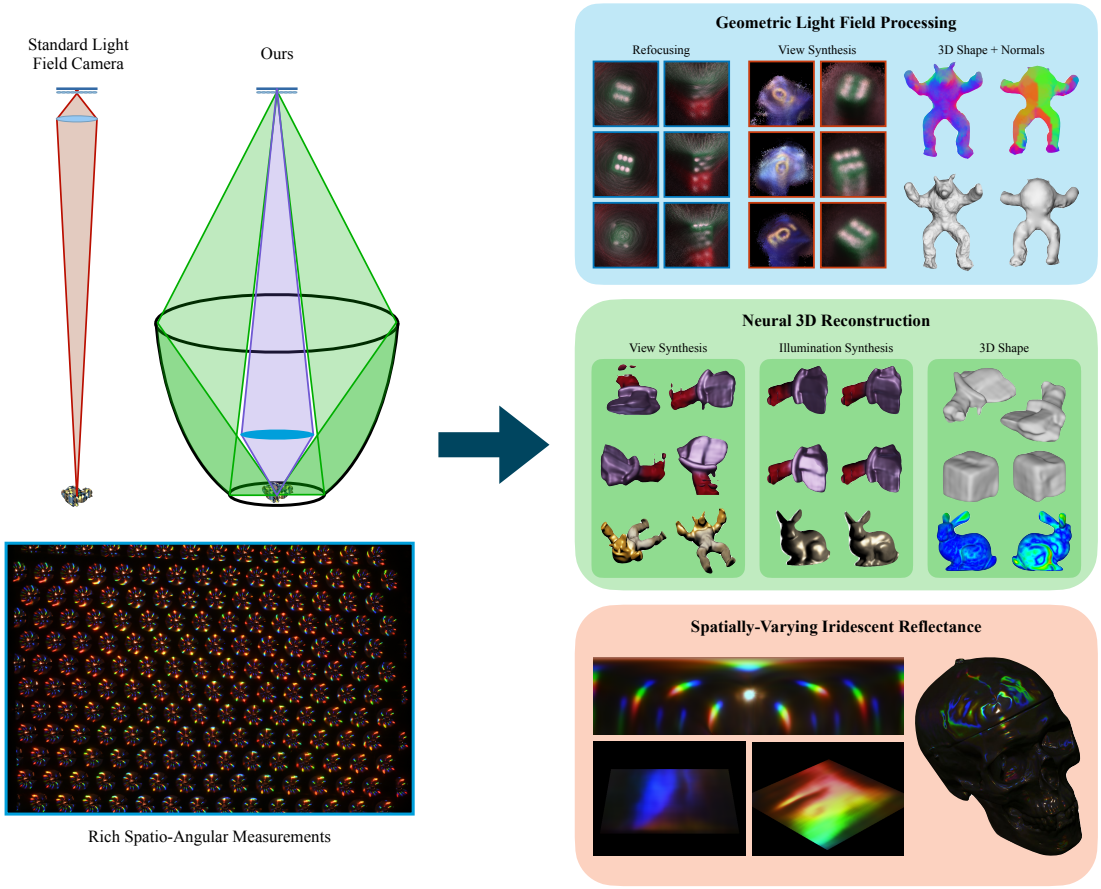

Figure 1.1: **Summary of thesis contributions.** We propose an imaging design for a wide-baseline light field camera and show that it is capable of measuring a much larger cone of light compared to conventional light field imaging techniques. This imaging system provides rich measurements in both space and angle. We apply this camera system for three main contributions. First, we parallel traditional light field processing techniques by developing a suite of geometric processing algorithms on our wide-baseline light field data, including refocusing, viewpoint synthesis, and 3D shape and normal estimation. Second, we extend the 3D reconstruction capabilities of our system by employing an implicit neural scene representation; this improves our 3D shape estimation, especially for real scenes, and enables rendering from both novel viewpoints and novel illumination conditions. Finally, we apply the wide-baseline light field camera to capturing dense measurements of the spatially-varying bidirectional reflectance distribution function (SVBRDF); paired with a neural reflectance representation, we show that we are able to reconstruct spatially-varying iridescent materials, which exhibit high-frequency reflectance details in both spatial and angular domains.

ject. Furthermore, by using visibility as a physical cue, we can estimate the surface normals of scene points; such a capability, especially in the absence of any active illumination or strong assumptions on the scene illumination, is unique to our setup.

- **Neural 3D Reconstruction from WBLF Data (Chapter 6).** We observe that state-of-the-art neural scene representations benefit from our captured data. These benefits include more continuous sampling relative to the sparse 2D input images used along with single-shot capture of 3D geometry and reflectance. We implement an implicit neural scene representation with added illumination control to produce high-quality 3D reconstructions of geometry, novel viewpoint synthesis, and novel illumination synthesis for both synthetic and real scenes.

- **Reconstructing Spatially-Varying Iridescent Reflectance from WBLF Data (Chapter 7).** Our wide-baseline light field camera provides dense measurements in space and angle over a broad angular range. By further adding dense illumination control, our setup is capable of efficient measuring across the full gamut of the spatially-varying bidirectional reflectance distribution function (SVBRDF). In particular, our measurement capabilities over a broad range of angles, including grazing angles to the scene surface, enables us to capture high-frequency angular effects like iridescence. Combined with the spatial measurements that our system provides, we are able to capture spatially-varying iridescent reflectance phenomena, which may be caused by diffraction or structural coloration. The high data output and lack of fully expressive iridescent models leads us to further develop a neural SVBRDF representation, which we train over our wide-baseline measurements. We show that we are able to reconstruct complex iridescent materials that exhibit high-frequency effects in both spatial and angular domains.

1.2 Thesis Layout

The remainder of the thesis is organized as follows.

Chapter 2 outlines the fundamentals of light field imaging, including prior techniques for capturing such information. We motivate why we aim to extend the effective baseline of light field capture with our imaging setup.

Chapter 3 details the optical design of the wide-baseline light field camera. We describe the optical qualities that enable this design and consider the capabilities of our imaging system. We further consider additions to this design that were made to support additional applications of our camera. *Chapter 4* establishes the required calibration techniques for our imaging system.

Chapters 5-7 showcase our three primary applications of the wide-baseline light field camera: geometric light field processing, neural 3D reconstruction, and spatially-varying iridescent reflectance.

Chapter 8 concludes this thesis, highlights the impact of this work on the field of light field imaging, and notes avenues for future work in this space.

2 Wide-Baseline Light Field Imaging

In this thesis we aim to measure the light that is propagating from a scene over a broad range of outgoing angles. These broad angles provide rich insight into the characteristics of the scene, including geometry and reflectance information. In this chapter, we consider the basics of representing and capturing this information.

2.1 Light Fields

Light fields are a useful structure for defining how light is propagating through space. [Gershun, 1939] define the light field as a mapping of the geometry of the space of 3D rays to the radiometric properties of light present along a given ray. The properties measured can be any quantity, but a typical usage of the light field defines this as the radiance of light L propagating along the ray. The geometry of the 3D rays is most simply defined by the plenoptic function, defined by [Gershun, 1939] as $L(x, y, z, \theta, \phi)$, a 5D function defining rays originating at any 3D point (x, y, z) and propagating in spherical direction (θ, ϕ). This 5D function can extend to higher dimensions by considering time, wavelength, and polarization. Most simply, we extend to include wavelength by considering the intensity of light within red, green, and blue channels that align with the spectral response of the imaging system.

If we assume that the light field is propagating through empty space and that the intensity of light along a ray does not experience any changes while propagating through this space, then we can simplify the 5D plenoptic function to four dimensions. To limit the degrees of freedom of the light field function to these four dimensions, we have several options for parameterization. [Levoy and Hanrahan, 1996] define a 4D parameterization of the light field as the propagation of light rays between two parallel planes, as shown on the left in Figure 2.1. This parameterization allows for simple resampling for novel renderings and aligns well with light field imaging systems that measure the incoming light field along a plane. However, this system is less well-suited for light fields propagating over a broader angular

range, as the spatial extent required on the two planes would increase dramatically as we move toward grazing angles. An alternative approach is to parameterize the light field as two intersections with a sphere of known radius and position in 3D space [Camahort *et al.*, 1998], illustrated in the center of Figure 2.1. Each ray propagating within the sphere will intersect with it at two locations; representing these locations by the spherical angles (θ, ϕ) creates four dimensions across which to consider the light field. This parameterization considers the entire light field for bounded scenes, including at grazing angles; however, it aligns less closely with more conventional image projection operations. Finally, we can simply consider the 5D plenoptic function at points along a 2D plane; this produces 2D intersection coordinates (x, y) and 2D spherical angles for the direction (θ, ϕ). This is shown in the figure on the right. We generally consider our imaging system in the context of this final parameterization. This allows us to consider how a light field at one plane (e.g. the object plane) propagates to the light field at another plane (e.g. the sensor plane). Separating the representation into a spatial location and an angular component makes analysis of spatial and angular resolution simpler and more intuitive.

2.1.1 Light Field Imaging

Since light field information is present in the full range of light ray directions present in the scene, measuring the full light field is a daunting task. Most light field imaging systems therefore capture just a portion of the full light field; most commonly, this involves capturing a narrower range of light field angles or sparsely measuring the light field with discrete cameras. Several techniques exist for light field imaging, including microlens array-based cameras, multi-camera arrays, and other multi-view imaging techniques. Figure 2.2 summarizes the differences between these techniques and our proposed imaging system.

Microlens array-based light field cameras [Lippmann, 1908, Ng *et al.*, 2005] provide a compact optical design for resolving the scene focused on the image plane of a lens into spatial and angular dimensions. Light fields acquired from such devices have been studied extensively for refocusing [Ng, 2005] as well as estimation of shape and reflectance [Wang *et al.*, 2016]; the interested reader is referred to [Levoy, 2006] for a survey on light fields, and [Georgiev *et al.*, 2006] for a study of space-angle tradeoffs with such designs. The angular range of the light field that these devices are capable of capturing is limited by the aperture of the main objective lens. A standard aperture of 5 cm imaging a scene from a meter away captures a cone with an angular range of only $3°$.

Light fields have also been studied in the context of microscopy [Levoy *et al.*, 2006, Lin *et al.*, 2015]. Unlike traditional photography, microscopes do gather a larger cone of light from scene points; crucially,

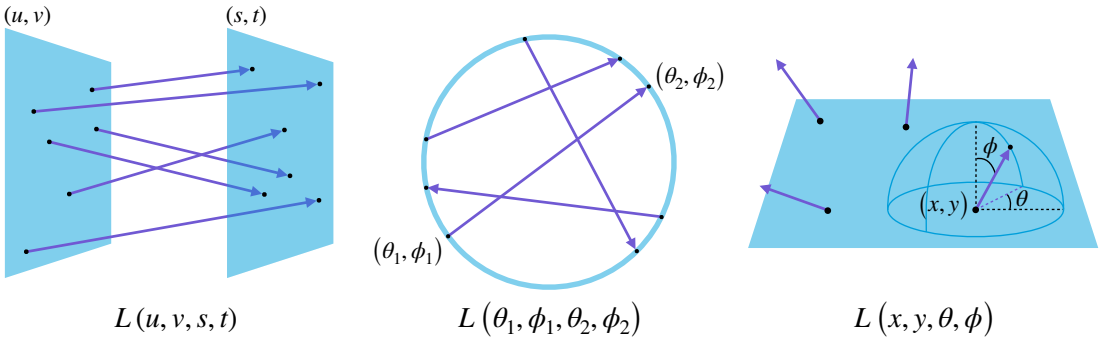

Figure 2.1: **Different parameterizations of the light field.** Light fields characterize the propagation of a field of light rays through space. This propagation can be parameterized in different ways based on the application. First, we show a classic two-plane parameterization of the light field (left). This representation denotes light rays propagating from a point (u, v) on one plane to a point (s, t) on another, resulting in a light field defined by $L(u, v, s, t)$. This system performs well for light field rays that all generally propagate in the same direction. Second, we show a spherical parameterization (center). In this case, light fields are recorded based on the spherical coordinates of the two intersections of a ray with a sphere, resulting in a light field defined by $L(\theta_1, \phi_1, \theta_2, \phi_2)$. This system does well to represent light fields within an enclosed space, defined by the sphere. Finally, we show a point-angle parameterization (right). Here, light field rays are denoted with an origin (x, y) and direction (θ, ϕ), resulting in a light field defined by $L(x, y, \theta, \phi)$. This system enables consideration of angle and space separately for light fields where the plane of consideration is of importance. It also represents a broad angular range of light rays easily.

the large *spatial magnification* of the microscope results in commensurate compression of the angular cone—a property that is shared with our design. However, even these devices are limited in angular range. Assuming a numerical aperture of 0.5 for a powerful microscope objective, the light field microscope still would only cover an angular range of 60°. Attempting to expand this angular range further would require objective lenses that have impractically large numerical apertures.

Camera arrays [Wilburn *et al.*, 2005] provide a different approach to acquire light fields; their large baseline provides a synthetic aperture that excels in focusing through complex occluders [Vaish *et al.*, 2006]. Closely related to camera arrays are light stages [Debevec *et al.*, 2000], used extensively for shape and reflectance acquisition, and kaleidoscopes, used for acquiring light fields [Manakov *et al.*, 2013] as well as 3D shape [Ahn *et al.*, 2021, Lanman *et al.*, 2009, Xu *et al.*, 2018]. All of these techniques produce multiple view points—real or virtual—that fully surround an object; yet, each view point only observes

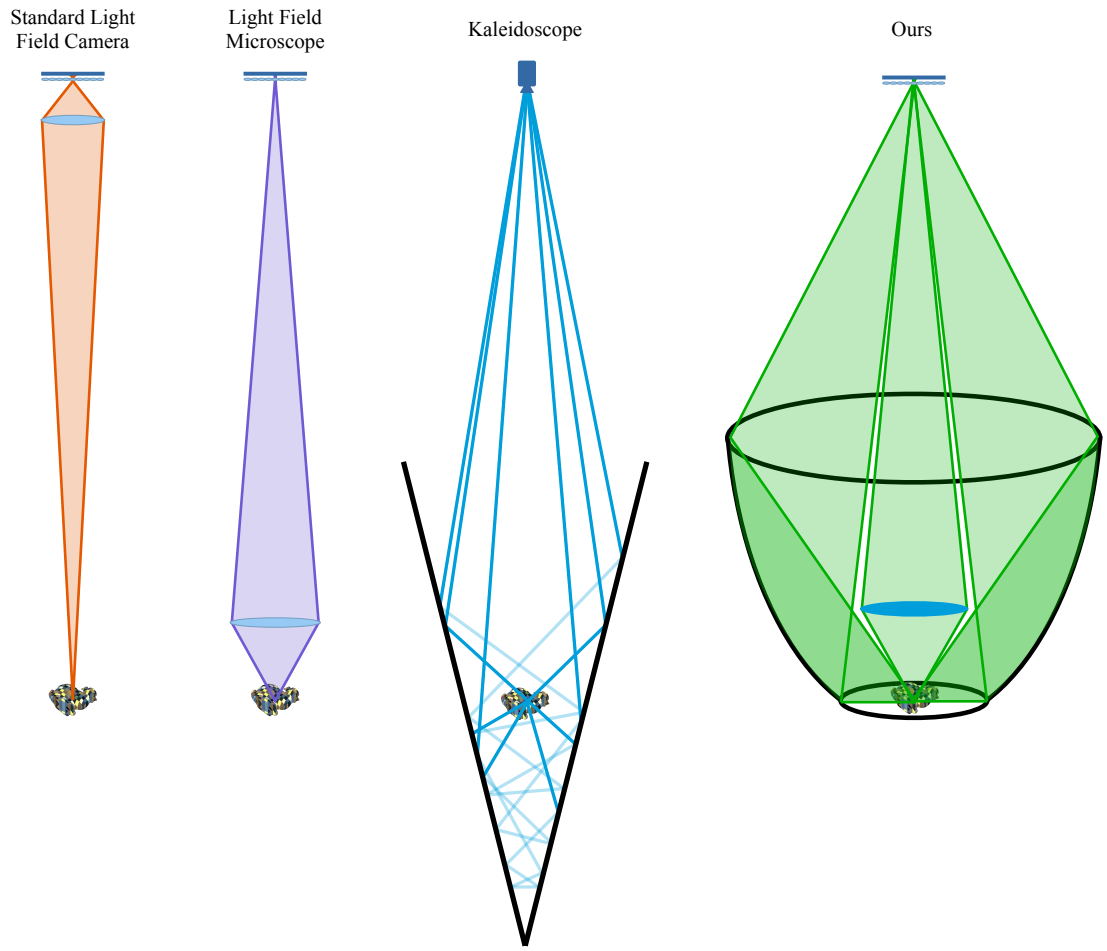

Figure 2.2: **Comparison of different light field imaging techniques.** We show the ray mapping for several common light field imaging techniques along with our proposed imaging system, which will be discussed further in Chapter 3. First is a traditional microlens array-based light field camera, which measures only a very narrow cone of the object light field. Second is a light field microscope, which captures a much larger cone due to the large spatial magnification of the system. However, these devices miss out on any grazing angles of the light field. Third is a kaleidoscope, which utilizes a configuration of planar mirrors to produce a wide range of synthetic viewpoints of the object. These devices do cover a broad angular range, but they do so sparsely and discretely, leaving large gaps in the angular extent due to self-occlusion and mirror geometry. Finally, we show our imaging design, which utilizes an ellipsoidal mirror to continuously map a large angular range of the light field to our sensor, which is a microlens-based light field camera design.

each scene point a single time, producing just one additional measurement direction. Thus, we obtain a very sparse aliased sampling of the cone of light from any scene point. Additionally, kaleidoscope designs that produce a large number of virtual view points, many of these views are severely occluded by the mirror geometry or the object itself. This further limits the number of times a particular scene point is observed.

Closely related to the proposed imaging of wide-baseline light fields is a body of work that aims to acquire wide *field of view* (FoV) light fields. [Taguchi *et al.*, 2010b] image an array of spherical mirrors with a conventional camera; here, each spherical mirror provides a wide FoV image with a slightly different perspective, thereby mimicking a camera array. Moving the camera axially in front of a symmetric mirror has also been shown [Taguchi *et al.*, 2010a] to produce wider FoV devices. Similar designs for expanding the FoV, using refractive optics, were considered in [Dansereau *et al.*, 2017]. While these works are conceptually similar to ours, enhancing the FoV of the camera is significantly different from capturing a larger cone of light from scene points. Expanding the FoV is equivalent to expanding the angular range of the measured light field *at the sensor plane*. This does not expand the angular range of the measured light field at the scene.

2.1.2 Wide-Baseline Light Field Imaging

This thesis proposes a device capable of measuring light fields over a broad angular range, as shown in Figure 2.2. Through the use of an ellipsoidal mirror, we are able to continuously map a broad cone of the light field from the object to a narrower, measurable cone at the sensor. We describe this design is more detail in Chapter 3.

2.2 Catadioptric Imaging

Since our imaging design utilizes an ellipsoidal mirror to achieve the desired angular range, we consider other prior uses for such imaging optics. Our proposed approach falls under the class of catadioptric cameras, where mirrors are used to augment the refractive optics found in consumer devices. Baker and Nayar [Baker and Nayar, 1998] study the family of mirror shapes that can alter the field of view of a perspective camera, and ellipsoidal mirrors are among the shapes that they analyze; our work can be interpreted as a specialized application of these ideas, when a scene and a light-field camera are placed in the two foci of such a mirror.

The large light collection area of mirrors has been used extensively to study reflectance properties of materials, using ellipsoidal [Mukaigawa *et al.*, 2007, 2009] as well as parabolic mirrors [Dana and Wang,

2004, Ghosh *et al.*, 2007, Zhang *et al.*, 2015]. In particular, our use of ellipsoidal mirrors is motivated from Ghosh et al. [Ghosh *et al.*, 2007] and Mukaigawa et al. [Mukaigawa *et al.*, 2007, 2009], where the BRDF of a material sample is measured over a large angle. All of these works apply the reflective properties of curved mirror surfaces with focal points. For light originating or passing through one of these focal points (ellipsoidal geometries have two, parabolic geometries have one), the reflection off the mirror behaves in a highly controlled manner (ellipsoidal mirrors focus this light to the other focal point, parabolic mirrors parallelize this light). This leads these works to solely focus on the reflectance of a single scene point. Considering these works in terms of the light field $L(x, y, \theta, \phi)$, these devices densely measure a broad range of θ and ϕ but measure from just a single point (x, y).

Moving outside the focal point loses the controlled reflective mapping, but these light rays still propagate similarly to the others; the geometry is smooth in both cases, so small changes away from the focal point will have small changes in the resulting mapping. Our work combines the angular range provided by these mirrored devices with the spatio-angular measurements provided by a light field camera. By placing a light field camera within our system, we consider light field rays beyond just the focal point of the scene. This enables us to retain the angular measurements that capture reflectance information while additionally considering multiple views of different spatial points in the scene. We can apply this additional information to reconstruct the scene geometry and consider spatially-varying reflectance. The next chapter details our optical design that utilizes both an ellipsoidal mirror and a light field camera.

3 Wide-Baseline Light Field Camera

In this chapter, we introduce a novel optical design for capturing wide-baseline light fields.

As we discussed in Chapter 2, conventional designs for capturing light fields lack full coverage of the hemisphere of observation directions surrounding a scene. Whether this is due to the limited cone of light captured by a light field camera or due to the sparse measurements from a multi-camera or kaleidoscopic imaging setup, this proves detrimental to a variety of light field processing tasks. Our proposed light field imaging setup is capable of capturing wide-baseline light fields in a single exposure. We achieve this by using a standard light field camera that observes the scene of interest, not directly, but *through an ellipsoidal mirror.* This design relies on the observation that the ellipsoidal mirror morphs the wide cone of light from one of its foci to a much smaller cone at the other focus point; hence, placing the scene at one focus and the camera at the other allows us to acquire a wide-baseline light field. We show that the design is, in principle, capable of obtaining a larger cone of light than prior designs, often close to a hemisphere for many scene points.

Furthermore, we analyze the properties of our wide-baseline light field camera, including resolution, depth of field, coverage, and more. In this, we consider the unique ways that the ellipsoidal mapping impacts the measured light field. This impact provides particular benefits that we can utilize and limitations that we aim to mitigate. To overcome some limitations, we introduce variations to the optical design, including illumination control and additional imaging optics. The design considerations for these variations will be discussed here, with the applications discussed in Chapters 5 - 7 using different variations based on the target applications.

Contributions. In total, this chapter introduces a new methodology for capturing wide-baseline light fields, making the following key contributions.

- *Wide-baseline light field camera design.* We design a catadioptric device using a light field camera and an ellipsoidal mirror for capturing wide-baseline light fields.

- *Wide-baseline light field analysis.* We derive notions of resolution, depth of field, coverage, and more analytical tools to better understand the optical mapping of the ellipsoidal mirror.
- *Variations for illumination control and extended angular range.* We develop variations to the optical design to better fit the imaging setup to particular applications.

3.1 Imaging Overview

We now describe an overview of our imaging system.

3.1.1 Imaging Through an Ellipsoidal Mirror

The key optical element of our imaging design is an ellipsoidal mirror, as shown in Figure 3.1. The geometry of an ellipsoid is driven by the radii in each of the three coordinate directions. Since we aim to apply the ellipsoidal to mapping the light field from a scene to a sensor, we limit our consideration to the subset of ellipsoids known as spheroids. In this case, two radii are set to be equivalent, meaning our 3D geometry is formed by revolving a 2D ellipse around its major axis. The geometry is therefore defined by the major and minor radii of the 2D ellipse, which we have denoted as a and b, respectively. Together, they define the overall shape of the ellipsoid along with the location of two focal points, separated by a distance $c = 2\sqrt{a^2 - b^2}$.

Ellipsoids have several unique properties, most notably that any ray originating at one focal point that is reflected off the inner surface of the ellipsoid will end up reaching the other focal point. This makes the ellipsoid very useful as imaging optics, as it acts very similarly to a very large aperture lens between these two points. This property has led to the ellipsoid being popularly employed in reflectance measurement devices, as outlined in Chapter 2.

We apply this property of the ellipsoid to the mapping of the light field from a scene at one focal point to a sensor at the other. The center illustration in Figure 3.1 shows how this mapping occurs for rays leaving the lower focal point. Note in particular that light rays that exit the lower focal point at broad angles (more perpendicular to the primary optical axis between the focal points, shown in green) are mapped to far narrower angles at the upper focal point and vice versa. For the full broad cone of light leaving the lower focal point, a far narrower cone reaches the upper focal point. This is important, as most imaging and sensor optics operate best when imaging narrower cones. Thus, the broad cone leaving the scene, which could not be measured by any one camera with refractive optics, is mapped to a cone that can be measured by such means. This is the core of our imaging technique and lends

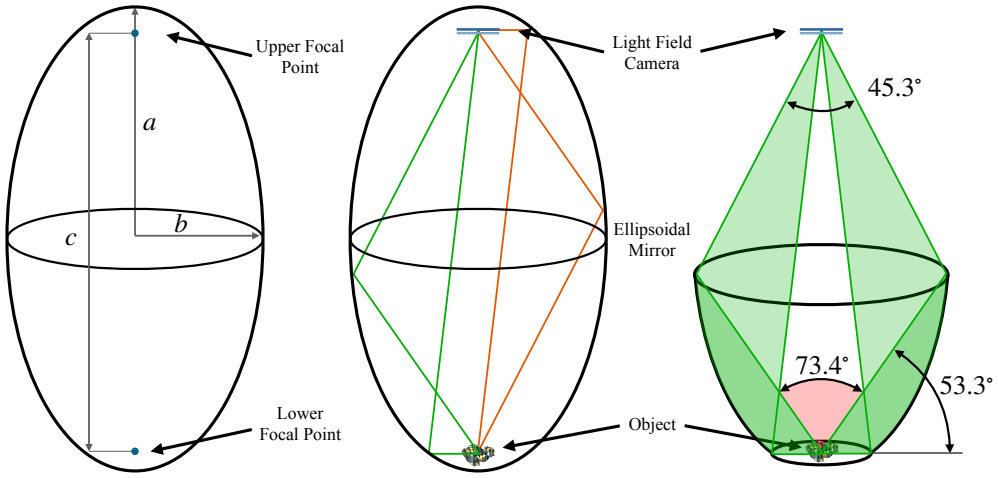

Figure 3.1: **Imaging setup.** Illustration of our imaging setup using an ellipsoidal mirror and a light field camera. The left hand side shows the overall geometry of the ellipsoidal mirror. Our setup uses an ellipsoid with major and minor radii of $a = 287.87$mm and $b = 135.47$mm respectively. The two focal points of the ellipsoid are spaced $c = 508.0$mm apart. Ellipsoids map rays originating at one focal point to the other point through reflection with the interior surface, as shown in the center. Through the ellipsoid, wide angles at the object plane are mapped to narrow angles at the sensor plane while narrow angles at the object plane are mapped to wide angles. The right hand side shows the setup with a section of the ellipsoidal mirror that maps to the measurable cone of light for the light field camera. The corresponding light field mapping is shown in green, which illustrates the mapping of a wide-baseline light field (53.3° over the horizon ×180° in azimuth) to a measurable light field (45.3° cone in diameter). The red cone in the center corresponds to the missing cone centered on the major axis of the ellipsoid.

the "wide-baseline" descriptor to our camera design, since the measurement of these light rays would otherwise require a light field imaging setup with a large baseline between multiple sensors. Conversely, the narrower central cone of light leaving the scene, once reflected off the mirror, will reach the sensor at extreme angles, as shown in orange in the figure. We discuss the impact of these rays in future sections.

The specific geometry of our imaging setup is largely driven by the market availability of image-quality ellipsoidal mirrors. Among the limited available options, we prioritized mirrors that could reflect to the widest possible cone of light measurable at the sensor to limit the size of the unmeasurable cone, with a second priority of having a large enough size to sample a larger spatial extent. The considerations led us to an ellipsoidal mirror with the geometry defined in Figure 3.1. As seen in the figure and discussed previously, the ellipsoidal mirror will reflect a wide cone of light from the scene to a narrow cone of light

at the sensor. Since light rays that reflect off the upper region of the ellipsoid will be at wide angles on the sensor, they cannot be measured. So, our final imaging setup uses just a section of the ellipsoid instead. This results in the mapping shown on the right in Figure 3.1, where a cone of the light field that spans a 53.3° angle above the horizon and 180° in azimuth gets mapped to a 45.3° cone at the sensor, which is able to be measured using standard imaging optics.

We also consider this mapping in terms of solid angle to provide a sense for the density at which the broader cone is being mapped into the narrower cone. The cone of light leaving the lower focal point as defined in Figure 3.1 covers a solid angle of 5.04 sr, while the cone that reaches the sensor covers a solid angle of 0.44 sr. This leads to a measurement density 11.56× greater than the measured light field, which allows us to measure a broader cone in a single exposure but imposes limitations on angular resolution. We discuss these limitations further in Section 3.2.

3.1.2 Light Field Camera

The previous section describes the mapping through the ellipsoid that occurs between the two focal points. While this provides a useful baseline for analysis, it does not consider the mapping of the entire light field. As we consider scene points further from the focal point, the mapped light field will be dispersed farther from the sensor focal point as well, and the focus-to-focus mapping will break down. However, the geometry of the ellipsoid will still focus these rays *towards* the sensor focal point, even if not to an exact point. This allows us to measure these various rays through the use of a light field camera placed at this focal point. Per the discussion in Chapter 2, we utilize the light field design of a microlens array placed in front of an imaging sensor. This sacrifices spatial resolution at the sensor plane for measured angular resolution at the same plane; we consider the impacts on the corresponding object plane light field measurements in Section 3.2.

The choice of design parameters of this microlens-based light field camera is driven by the need to measure the 45.3° cone of light propagating to the sensor based on the mirror geometry; this requires a microlens array with an f-number close to one. Based on this target and available optical elements, we choose an array with 2.2mm lenslet diameters and a 3mm focal length. The large diameter of these lenslets introduces a significant amount of spatial blurring. Further, due to the non-uniform mapping of rays through the ellipsoidal mirror, this blurring crosses different ranges of space and angle at the scene based on the angle and location at the sensor. This is not ideal for resolving this blur or applying these measurements to multiple applications. Thus, to limit spatial blur, we mount a pinhole array aligned with the center of each lenslet on top of the microlens array; a detailed justification for the use of the

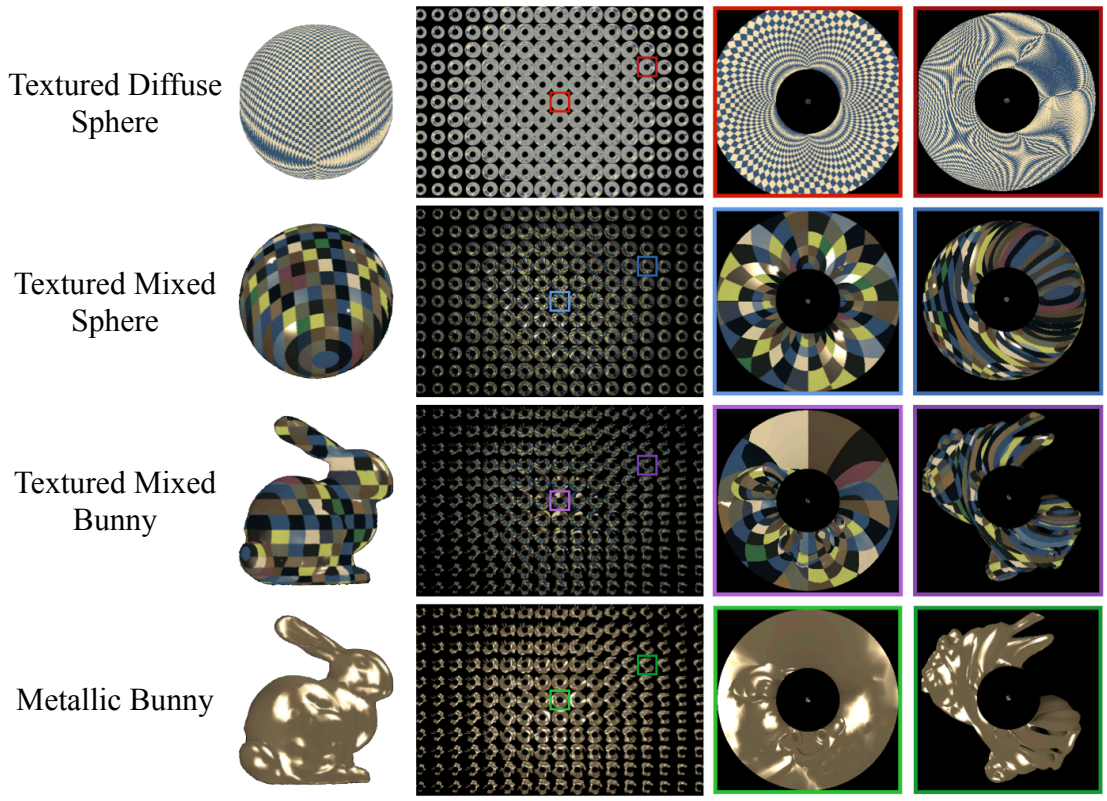

Figure 3.2: **Simulated imaging.** Example images rendered from a simulation of our imaging setup, including a close-up of two sub-aperture views for each scene. The black disk at the center of each sub-aperture is a consequence of the missing cone, highlighted in Figure 3.1.

pinhole array is presented in Section 3.2.1.

3.1.3 Full Optical System

By placing this light field camera at the upper focal point of the previously described ellipsoidal mirror section, we are able to capture wide-baseline light fields from the scene. Figure 3.2 shows simulated data of the captured light field images. Under each lenset, we observe a different subset of the light field via a different view of the ellipsoidal mirror below. Different textural elements from the surface of the object are transformed into these different views, revealing the spatial measurements of this camera. Various angular effects of the scene reflectance are observed as well, revealing the angular measurement capabilities. We thus observe that our optical design is capable of measuring light field information over a broad range of angles; further analysis of the capabilities of this camera is shown in the next section.

3.2 Analysis

The use of the ellipsoidal mirror in our setup maps the wide-baseline light field at the scene to a measurable light field captured by our sensor. This section details the specific characterizations of this mapping and analyzes the impact on the coverage, resolution, angular range, and depth of field of our camera.

3.2.1 2D Propagation Analysis

We begin with an analysis of a 2D version of our setup. Since our ellipsoidal mirror is rotationally symmetric along the optical axis, this 2D approximation provides a convenient foundation for our understanding. The 2D equivalent of our setup is a mirrored section of a 2D ellipse mapping rays from the scene to the sensor. We derive a closed form equation to describe the propagation of a light ray at the object plane to a light ray at the sensor plane.

We start with the equation of an ellipse centered at the origin $(x_c, y_c) = (0, 0)$, with a being the major axis, b being the minor axis, and the major axis being oriented along the y-axis of the coordinate system. We use simple ray tracing to propagate an object-plane light ray (x_1, θ_1) to its corresponding sensor-plane light ray (x_2, θ_2). In all of the following equations, $y_1 = -254$mm is the height of the object focus and $y_2 = 254$mm is the height of the sensor focus. The ellipse equation is

$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1. \tag{3.1}$$

We then use a standard time-parameterized ray equation on both x and y:

$$x_t = x_1 + t\cos\theta_1, \quad y_t = y_1 + t\sin\theta_1 \tag{3.2}$$

We use the subscript t to denote the location (x_t, y_t) as well as the time t for an intersection with the ellipsoidal mirror. This is found by plugging the above equations into the equation for the ellipse to get

$$\frac{(x_1 + t\cos\theta_1)^2}{b^2} + \frac{(y_1 + t\sin\theta_1)^2}{a^2} = 1 \tag{3.3}$$

$$\begin{aligned} a^2x_1^2 + 2a^2x_1t\cos\theta_1 + a^2t^2\cos^2\theta_1 \\ +b^2y_1^2 + 2b^2y_1t\sin\theta_1 + b^2t^2\sin^2\theta_1 = a^2b^2 \end{aligned} \tag{3.4}$$

$$\begin{aligned} t^2\left(a^2\cos^2\theta_1 + b^2\sin^2\theta_1\right) \\ +t\left(2a^2x_1\cos\theta_1 + 2b^2y_1\sin\theta_1\right) \\ +\left(a^2x_1^2 + b^2y_1^2 - a^2b^2\right) = 0. \end{aligned} \tag{3.5}$$

Equation 3.5 takes the form of a quadratic function on t, so we can solve the quadratic equation

$$\widehat{t} = \frac{-v \pm \sqrt{v^2 - 4uw}}{2u}, \tag{3.6}$$

$$u = a^2\cos^2\theta_1 + b^2\sin^2\theta_1, \tag{3.7}$$

$$v = 2a^2x_1\cos\theta_1 + 2b^2y_1\sin\theta_1, \tag{3.8}$$

$$w = a^2x_1^2 + b^2y_1^2 - a^2b^2. \tag{3.9}$$

We write out the two terms in the square root as

$$\begin{aligned} v^2 =& 4a^4x_1^2\cos^2\theta_1 + 8a^2b^2x_1y_1\cos\theta_1\sin\theta_1 \\ &+ 4b^4y_1^2\sin^2\theta_1 \end{aligned} \tag{3.10}$$

$$\begin{aligned} 4uw =& -4a^4x_1^2\cos^2\theta_1 - 4a^2b^2x_1^2\sin^2\theta_1 \\ &- 4a^2b^2y_1^2\cos^2\theta_1 - 4b^4y_1^2\sin^2\theta_1 \\ &+ 4a^4b^2\cos^2\theta_1 + 4a^2b^4\sin^2\theta_1. \end{aligned} \tag{3.11}$$

Combining these cancels several terms, leaving

$$\begin{aligned} v^2 + 4uw =& 8a^2b^2x_1y_1\cos\theta_1\sin\theta_1 - 4a^2b^2x_1^2\sin^2\theta_1 \\ &- 4a^2b^2y_1^2\cos^2\theta_1 + 4a^4b^2\cos^2\theta_1 + 4a^2b^4\sin^2\theta_1, \end{aligned} \tag{3.12}$$

which simplifies to

$$v^2 + 4uw = 4a^2b^2\left(\widehat{\alpha}(\theta_1) + \widehat{\beta}(x_1, \theta_1)\right) \tag{3.13}$$

$$\widehat{\alpha}(\theta_1) = a^2\cos^2\theta_1 + b^2\sin^2\theta_1 \tag{3.14}$$

$$\widehat{\beta}(x_1, \theta_1) = 2x_1y_1\cos\theta_1\sin\theta_1 - x_1^2\sin\theta_1 - y_1^2\cos\theta_1. \tag{3.15}$$

We plug this back into Equation 3.6 to sove for $\widehat{t}$, using $w = \widehat{\alpha}(\theta_1)$ and defining $\widehat{\gamma}(x_1, \theta_1)$:

$$\widehat{t}(x_1, \theta_1) = \frac{\widehat{\gamma}(x_1, \theta_1) \pm ab\sqrt{\widehat{\alpha}(\theta_1) + \widehat{\beta}(x_1, \theta_1)}}{\widehat{\alpha}(\theta_1)} \tag{3.16}$$

$$\widehat{\gamma}(x_1, \theta_1) = -a^2x_1\cos\theta_1 - b^2y_1\sin\theta_1. \tag{3.17}$$

This $\widehat{t}$ value provides the time for the ray propagation to the intersection point with the mirror. Propagating from this point to the sensor is simply using the ray propagation equations again:

$$x_2 = x_t + t_2\cos\theta_2, \quad y_2 = y_t + t_2\sin\theta_2, \tag{3.18}$$

where θ_2 is the angle after reflecting off of the mirror. Since y_2 is a fixed height at the location of the sensor, we can solve for the propagation time t_2:

$$t_2 = \frac{y_2 - y_t}{\sin\theta_2}, \tag{3.19}$$

which leads to the following solution for the sensor-plane intersection location x_2:

$$\begin{aligned} x_2 &= x_t + \frac{y_2 - y_t}{\sin\theta_2}\cos\theta_2 \\ &= x_1 + \widehat{t}(x_1, \theta_1)\cos\theta_1 + \frac{y_2 - (y_1 + \widehat{t}(x_1, \theta_1)\sin\theta_1)}{\tan\theta_2}. \end{aligned} \tag{3.20}$$

Finally, we need to define the new ray angle θ_2. We utilize the equation for the angle ϕ of the tangent of the ellipse at point (x_t, y_t):

$$\phi = \tan^{-1}\left(\frac{y_t b^2}{x_t a^2}\right). \tag{3.21}$$

The equation for the reflection angle θ_2 is therefore

$$\begin{aligned}\theta_2 &= \pi - \theta_1 + 2\phi \\ &= \pi - \theta_1 + 2\tan^{-1}\left(\frac{b^2(y_1 + \widehat{t}(x_1, \theta_1)\sin\theta_1)}{a^2(x_1 + \widehat{t}(x_1, \theta_1)\cos\theta_1)}\right).\end{aligned} \tag{3.22}$$

In total, we have derived an expression for the measured light ray (x_2, θ_2) as a function of the object-space light ray (x_1, θ_1) and the mirror geometry. We utilize this expression to determine coverage and resolution properties in 2D space.

Propagated Light Field

We first consider the portion of the outgoing light field from the scene that is propagated to our imaging sensor. Figure 3.3 shows a visualization of this propagation for both a full ellipse and a section of the ellipse, which corresponds to the section of the ellipsoid used in our imaging setup. The object-plane light field is color coded, with brightness corresponding to spatial location and color corresponding to angle. The plot shows how those color-coded light rays propagate to the sensor plane. The measurable region, assuming a 1 inch sensor and a microlens array capable of imaging a 45° cone of light, is shown on the plot. This highlights the key benefit of the ellipsoidal mirror, which maps the grazing angles (close to 0° [red] and 180° [blue]) to the measurable region at the expense of other angles.

Spatial and Angular Resolution

The limitations of resolution on the sensor propagate to the measured light field in particular ways. We perform gradient analysis on the propagated 2D light field formulation defined previously to quantify how changes at the sensor plane (corresponding to the achievable spatial and angular resolution) propagate back to the object plane. Based on the expression in (3.20, 3.22), we derive the changes in x_2 and θ_2 based on changes in x_1 and θ_1 as:

$$\begin{bmatrix} \frac{\Delta x_1}{2} \\ \frac{\Delta \theta_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial \theta_1} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial \theta_1} \end{bmatrix}^{-1} \begin{bmatrix} \Delta x_2 \\ \Delta \theta_2 \end{bmatrix} \tag{3.23}$$

We can also fix $x_1 = 0$ and vary y_1 in our equations for x_2 and θ_2. We then take the derivatives with respect to y_1 instead:

$$\begin{bmatrix} \frac{\Delta y_1}{2} \\ \frac{\Delta \theta_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial \theta_1} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial \theta_1} \end{bmatrix}^{-1} \begin{bmatrix} \Delta x_2 \\ \Delta \theta_2 \end{bmatrix} \tag{3.24}$$

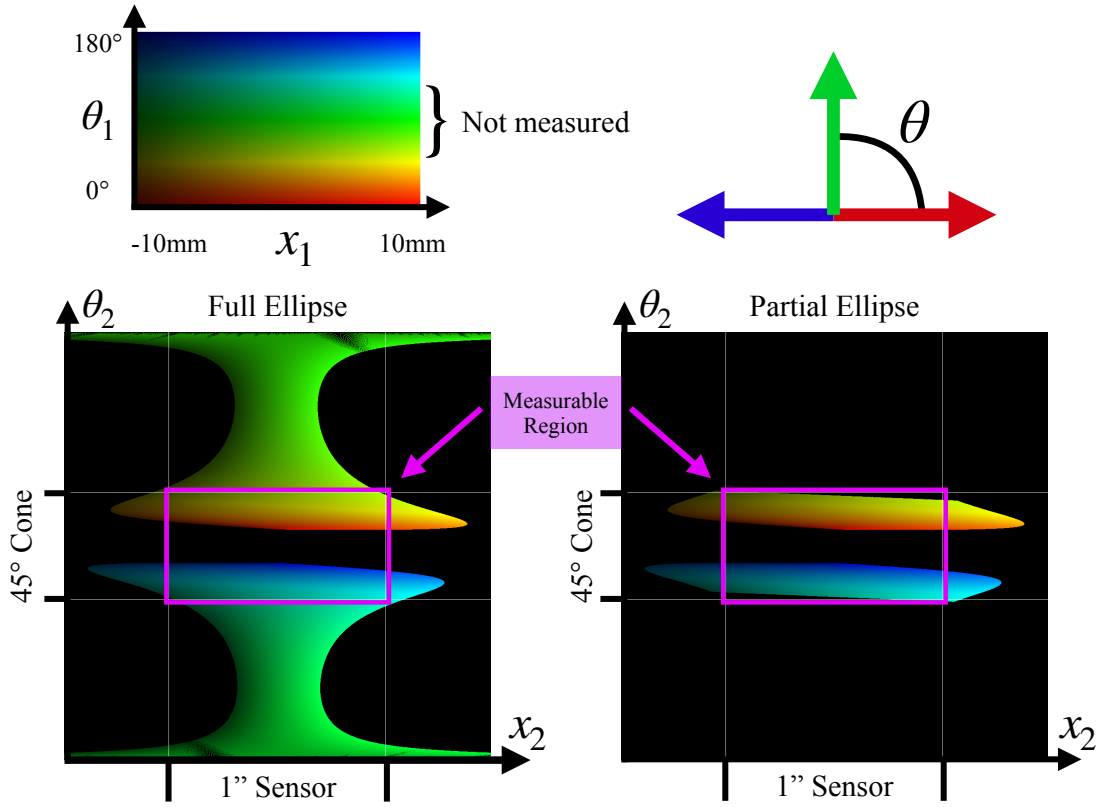

Figure 3.3: **Light field propagation.** Plot of the light field propagation from the object plane to the measurement plane for a complete (lower left) and partial (lower right) ellipse. Rays at the object plane are color coded according the the plot in the top left, and the corresponding mapping to the measurement plane is the location of that color on the lower plots. The measurable region, assuming a 1" sensor and a 45° measurable cone, is marked on the axes and highlighted. The actual section of the ellipse used in our imaging system omits primarily the unmeasured light rays, resulting in the mapping in the bottom right. This section preserves the wide angular range that we are looking for at the expense of the central cone of light most commonly captured in other setups.

To evaluate these expressions, we define Δx_2 as the size of the aperture used over the microlens array, for this indicates the size of one spatial "pixel". We test this value over multiple possible aperture sizes, which would be implemented in practice as an aperture array. We approximate $\Delta\theta_2$ as the pixel pitch of the sensor divided by the focal length of the microlens: $\Delta\theta_2 \approx \frac{5.94\mu\text{m}}{3\text{mm}} = 0.00198 = 0.113°$. Using these values and computing the partial derivatives of f_1 and f_2 based on their definitions in (3.20, 3.22), we can analyze the resulting object-plane resolution for various portions of our object-plane light field.

Figure 3.4 plots different object-plane resolutions for three aperture sizes defined in the figure. We consider the spatial resolution both in terms of y_1 and x_1. The plots show the theoretical resolution for the entire 180° range of outgoing angles propagated through a complete ellipse. Our imaging setup only captures outgoing angles up to about 50°, which corresponds to approximately $0° - 50°$ and $130° - 180°$ on the plots; this does not include angles close to 90° that observe worse spatial resolution. These regions are highlighted in the figure.

Focusing on these regions of interest, we can see that using no aperture on the microlens array results in very poor resolution in both space and angle. Adding a 200μm aperture improves the angular resolution to between 0.1° and 0.5°, the typical horizontal spatial resolution $|\Delta x_1|$ to less than 100μm, and the vertical spatial resolution $|\Delta y_1|$ to $\sim$ 10 to 80 µm. Adding a 10μm aperture maintains a similar angular resolution and improves the spatial resolution further, but this comes with the risk of significant diffraction blur that is not modeled by our ray tracing. In this case, we see the typical horizontal spatial resolution $|\Delta x_1| < 10\mu$m and the vertical spatial resolution $|\Delta y_1| \sim 1$ to 8 µm. Note that the grazing angles in the center row of plots (near 0° and 180°) exhibit significant worsening of spatial resolution. This is due to the fact that at these angles, small changes in the incoming ray will cause large changes in the intersection of this ray with the object plane, which is defined oriented upwards toward the other focal point. In other words, by fixing y_1 and only considering resolution in terms of x_1, our spatial resolution definition inherently punishes grazing rays. Our analysis in terms of y_1 with a fixed x_1 does not have these significant variations and is thus a more suitable metric for the resolution of our system.

Choice of Aperture

Based on our resolution analysis, reducing the aperture size greatly improves spatial resolution and has minimal impact on angular resolution beyond the 200μm aperture size. These theoretical limits provide an understanding of the impact of the aperture choice; however, other considerations, including diffraction effects and achievable calibration accuracy, suggest that the smallest apertures will not introduce practical improvements. At a high level, we would consider a theoretical spatial resolution less than 100μm to be sufficient, which is achieved with the 200μm aperture. We consider this choice empirically as well by simulating spatial textures with or without an aperture. Figure 3.5 shows a simulated image capture of a diffuse textured sphere with no aperture and with a 200μm aperture. The added aperture prevents the blurring of textural elements in the image capture that are needed for various processing techniques. We implement the chosen aperture as a pinhole array aligned with the centers of each lenslet.

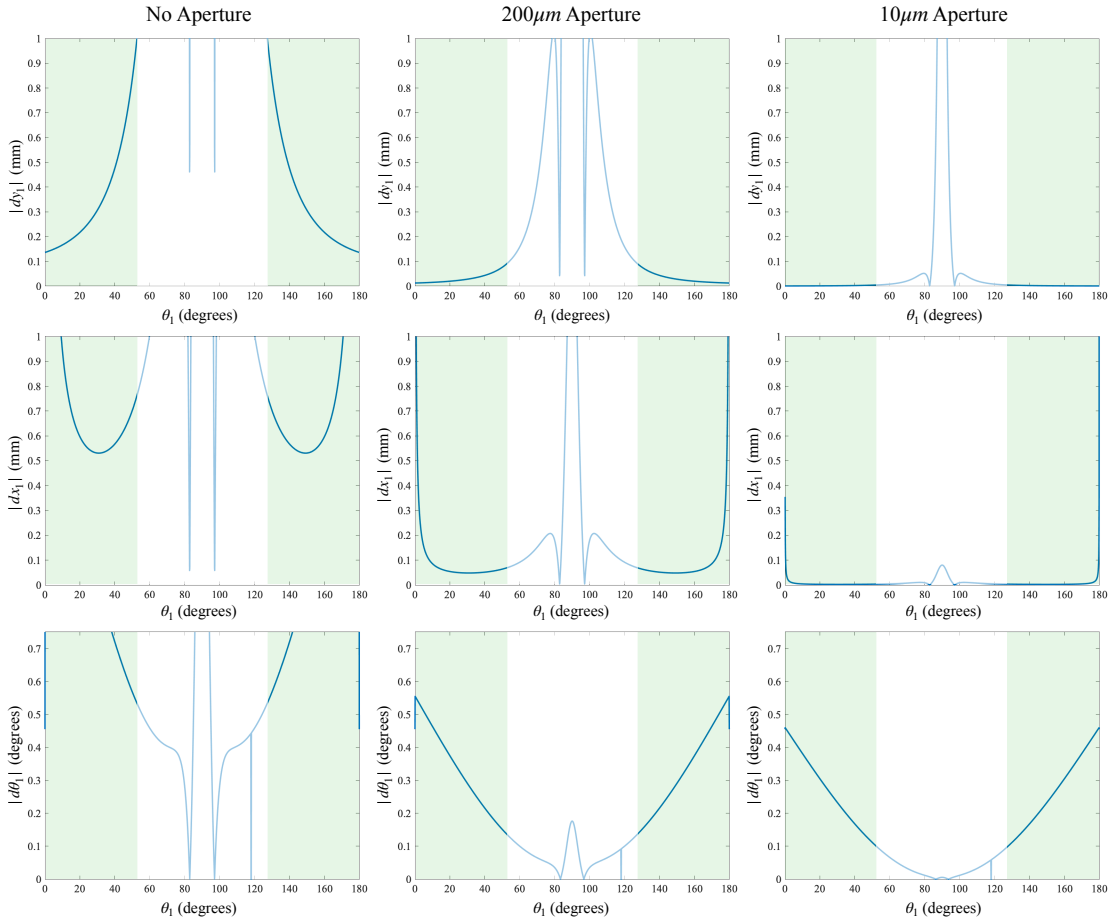

Figure 3.4: **Spatial resolution.** Plots of the object-plane angular and spatial resolution at $(x_1, y_1) = (0\text{mm}, 0\text{mm})$ as a function of θ_1. The resolution is strongly dependent on the spatial blur resulting from the chosen aperture size Δx_2, where different sizes are represented in the different columns. The top row plots the absolute value of the object-plane vertical spatial resolution $|\Delta y_1|$, the middle row plots the absolute value of the object-plane horizontal spatial resolution $|\Delta x_1|$, and the bottom row plots the absolute value of the object-plane angular resolution $|\Delta\theta_1|$. In our formulation, the smaller the resolution the better. We observe that the spatial resolution depends on the plane of consideration; for our wide angular range, $|\Delta y_1|$ aligns more with the rays that we are measuring.

2D Depth of Field Analysis

Continuing our 2D analysis, we consider the depth of field of our imaging system. For conventional camera systems, the depth of field is determined by the aperture and focal length of the main lens. For light field cameras, the measurement of different angles allows for refocusing and rendering of

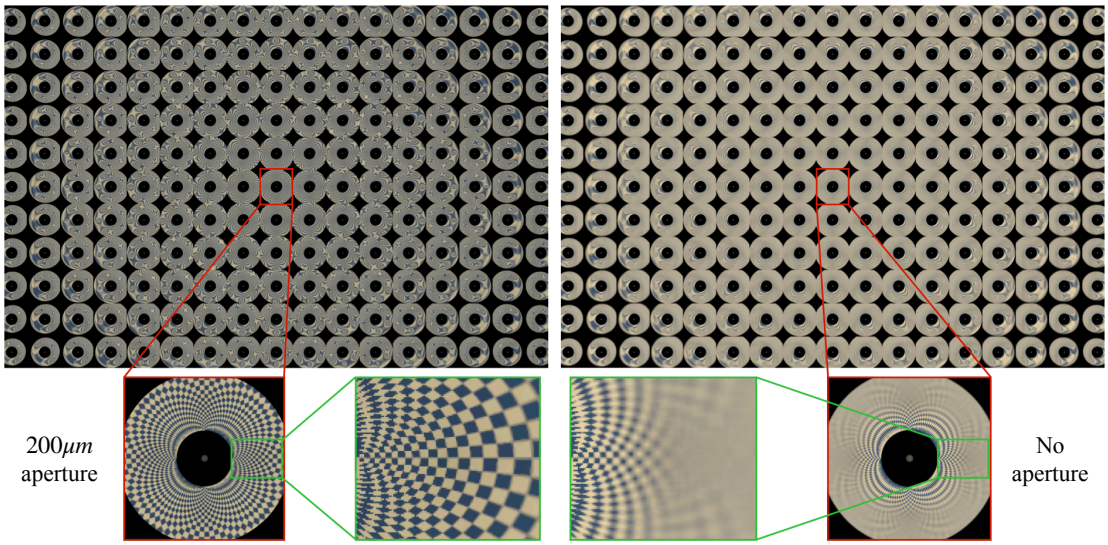

Figure 3.5: **Aperture study.** Simulated image captures with no aperture (right) and with a 200μm aperture (left). Adding the aperture preserves textural details in the image capture, as evident in the magnified areas shown.

images from different aperture sizes in post-processing. To the extreme, these light field cameras can be processed to produce all-in-focus images, with extremely large effective depths of field. While this can be useful for producing aesthetic images, practical considerations often lend preference to narrow depths of field; consider applying a focus stack, readily obtained from light field data, to solve depth from focus/defocus. In this case, narrower depths of field lead to better depth resolution. To this end, traditional single-shot light field cameras are again limited by the maximum aperture of the main lens.

Our imaging system captures wide-baseline light field images, which capture an angular cone ranging up to the full hemisphere of observation directions. Theoretically, this gives us an infinitely small depth of field. We illustrate this through a simple 2D example shown in Figure 3.6. In this example, we consider the imaging of a textured surface. In particular, we consider the measurement of the light field passing through a point of interest at distance d from the surface, shown in red. If we are capable of measuring a light field with cone width θ, we define s as the size of a virtual sensor positioned f away from the point that would measure this cone. For our imaging setup, we aim to have θ approach 180°, while other light field cameras would have θ be much smaller. The defocus observed as we move off the surface is dependent on the size of a surface texture element T, which is mapped to an area on the sensor of size b_T. From similar triangles, we can define the relationship between these variables in the following:

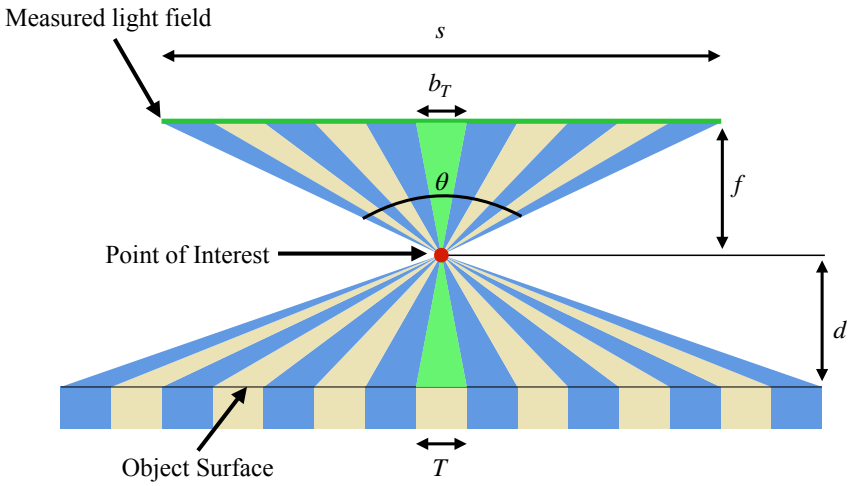

Figure 3.6: **Depth of field.** Illustration of the derivation of depth of field as a function of measured angular range. We test a point (in red) some distance d above the surface of the scene to see if it resolves in focus. If we are able to measure a light field with angular range θ, then we can produce a virtual sensor plane of size s at distance f from the point. With this geometry, a texture element of size T on the surface will project to size b_T on the virtual sensor. When this projected texture fills the virtual sensor, we will still resolve the point of interest without defocus blur.

$$s = 2f \tan\left(\frac{\theta}{2}\right), \quad \frac{d}{T} = \frac{f}{b_T}. \tag{3.25}$$

We can see that as d decreases with all the other variables fixed, the projected texture size b_T would increase. The point of interest would be considered in focus when all of the measurements on our virtual sensor correspond to the same point, which occurs when $b_T = s$. If this occurs at a distance $d > 0$ from the surface, then we know that refocusing from our light field data to the patterned surface will be in focus for all points located $\pm d$ from the surface. At the in focus point when $b_T = s$, the angle subtended by a texture element of size T on the test point d units away is equal to θ, the angular span of the light field. We plug this into the similar triangle expression in Equation 3.25 to find the maximum distance d at which the algorithm would consider a point to be in focus. The depth of field is twice this distance, meaning

$$d_{max} = \frac{T}{2\tan\left(\frac{\theta}{2}\right)}, \quad dof = 2d_{max} = \frac{T}{\tan\left(\frac{\theta}{2}\right)}. \tag{3.26}$$

As θ approaches a full hemisphere of 180°, d_{max} approaches 0. So, if we are able to measure a full

hemisphere of light above every surface point, there are no theoretical limits to the minimum depth of field with our imaging system. Instead, the achievable depth of field is limited only by the span of the angular cone that we measure (which changes at different spatial locations) and the quality of the measured rays, which are constrained by our imaging resolution along with the limitations of the optics used in our hardware setup. We consider these factors further in Section 3.2.2.

3.2.2 3D Ray Propagation Analysis

The 2D analysis performed in the previous section provides a useful baseline of understanding for the optical mapping through the ellipsoidal mirror and establishes several theoretical system capabilities. We extend this analysis to 3D; rather than derive the ray propagation equations, we simulate our optical design and visualize the captured light field in different ways. We first consider the distribution of light rays captured by the light field camera. To visualize this, we place a color pattern of size 20 mm $\times$ 20 mm at the focal point of the ellipsoidal mirror. Figure 3.7 shows a visualization of the resulting light field as observed by the camera, including the spatial measurements of the color pattern as well as colormapped visualization of the azimuth and elevation angles θ and ϕ of the observed rays at the object plane. We observe that each spatial point is measured by multiple pixels, with lenslets closer to the center of the image (therefore also closer to the upper focal point) converging closer to the lower focal point (center of the blue region). We note the large range of angles observed as well, from $0 < \theta < 2\pi$ in azimuth and $\sim 0.19\pi < \phi < \pi/2$ in elevation, which paired with the spatial mapping indicates a large span of measurements of the object-space light field. These simulated measurements align with the known geometry of the ellipsoid and the corresponding size of the outgoing cone of light.

Additionally, we consider our measurements on a point-by-point basis. Figure 3.8 visualizes the continuous region of rays that reach our sensor from a scene point at the focal point along with three points shifted by 5mm away from the focal point along each coordinate axis. This continuous mapping is a key benefit of the ellipsoidal mirror. The darker patches within each measured region correspond to the actual measurements once the space-angle tradeoff is taken into account, where the aperture mask creates sparser measurements in space at the sensor plane. This corresponds to measuring over small discrete regions which effectively provides samples over the entire visible region. We additionally observe that while the measurement distribution changes (and reduces in size) as we move away from the focal point, we are still capturing a broad set of angles from each point. This illustrates the nature of the breakdown of focus-to-focus mapping as we move away from the focal point The light field camera design, particularly the lenslet spacing and focal length, dictates how this distribution is actually

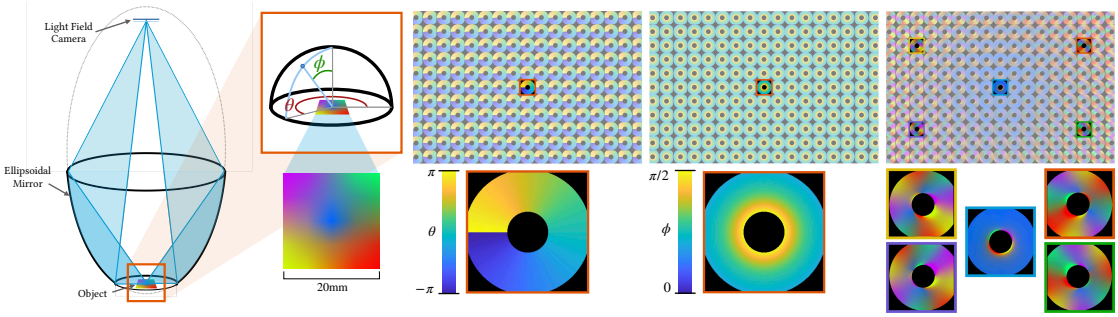

Figure 3.7: **Visualization of captured light field.** To showcase the space-angle coverage of our setup, we place a color-coded plane at the object focal point and capture a wide-baseline light field image. The far right column shows this captured image, which highlights the spatial extent of our measurements. The middle two columns visualize the angular information encoded in the captured light fields; we color-code the captured object-space azimuth angle θ and elevation angle ϕ according to the convention defined in the diagram. Note that we closely consider only one lenslet view each for θ and ϕ; the angular measurements are nearly uniform across different lenslets with only minor variations that are indistinguishable in this visualization.

measured.

Considerations of the angular range measured by our setup here are a useful framework, but the ellipsoidal mirror does not map each region of the measured cone to the sensor equally. So, we also consider visualizing our coverage via histograms of the captured spatial and angular information. Figure 3.9 shows a 2D histogram of spatial measurements at the object plane along with histograms for the azimuth and elevation angles θ and ϕ, respectively. The spatial histogram is shown in log2 scale and covers a 20 mm $\times$ 20 mm area. We observe that we capture a large number of measurements across this entire region, with more measurements being taken near the focal point in the center. This fall off in observations will dictate the choice in scene size for the various applications in Chapters 5 - 7.

The angular histograms illustrate the non-uniform mapping of the ellipsoidal mirror. In elevation angle, we capture relatively fewer grazing angles than we do higher angles. This is due to the mirror geometry, which is ultimately mapping these measurements to a disc on the sensor, as observed in Figure 3.2. The grazing angles get mapped to the center of this disc, while the higher angles get mapped to the outer region of the disc. This creates a larger area of pixels that are measuring the higher angles than the grazing angles, though we still capture many measurements of each. Our measurements in azimuth are generally uniform, as the ellipsoidal mirror is rotationally symmetric. The only variation is introduced by the sensor size, which is rectangular. This produces peaks when aligned to the corners

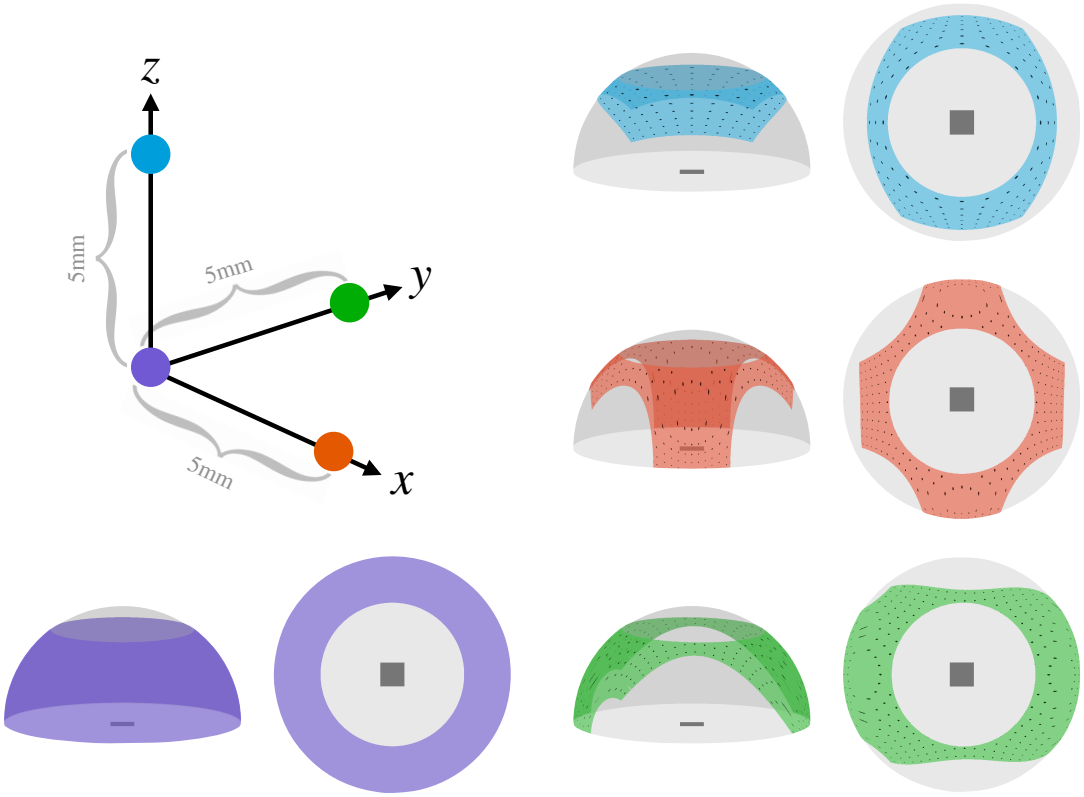

Figure 3.8: **Point-wise coverage.** Given a full hemisphere of possible sample directions from a single point, we visualize the subset of this space that can actually be sampled using the wide-baseline light field camera. We analyze this region for four points: one at the object focal point and three points shifted by 5 mm away from the focal point along each coordinate axis, illustrated on the left. Accounting for the space-angle tradeoff of the light field camera, the dark points within each region specify the actual measured rays. Note that there are no dark points for the point at the object focal point; here, the entire cone is mapped to a single point on the sensor and is able to be measured completely through one lenslet.

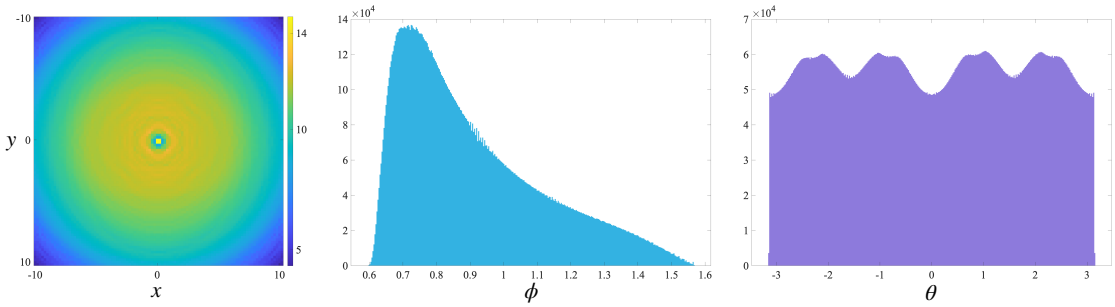

Figure 3.9: **Measurement coverage histograms.** We consider the coverage of space and angle that our imaging system provides. We consider the density of measurements for different spatial locations over a 20 mm × 20 mm area on the left. We note that due to the breakdown of focus-to-focus mapping, we observe a fall-off in measurement density as we move away from the focal point. This limits the measurement range of our system. We further plot the density of measurements over the range of angular measurements, defined by the same spherical angle notation used previously. In the center, we plot the measurement density over elevation angles ϕ; note that the non-uniform mapping of the ellipsoidal mirror leads to sparser measurements of grazing angles relative to the rest ($\phi \rightarrow \frac{\pi}{2}$). On the right, we plot the measurement density over azimuth angles θ; here we observe generally uniform measurements due to the rotational symmetry of the mirror. The only variation is introduced by our rectangular imaging sensor, which breaks the rotational symmetry.

and valleys when aligned with the sides, with different valley depths based on the different dimensions of the rectangular sensor.

Comparison to Common Light Field Setups

It is illustrative to compare the measurements of our setup with those achieved by other light field imaging techniques. Figure 3.10 visualizes a setup where two light field cameras are placed 12.5cm from the sample with a light collection area of 5cm; these numbers approximately correspond to the shortest focusing distance and entry aperture of the Lytro Illum. We can immediately observe the difference in angular coverage between the two setups; for example, the proposed imaging setup captures light rays along the horizon that the light field cameras miss. We observe on the right the effect of this difference in coverage. We illustrate the coverage of measurements of a surface point on a simulated textured sphere. Our imaging system is capable of diverse measurements that include the reflectance of the scene point (red glossy material with two observed specular highlights) as well as measurements that are occluded by the rest of the measured object (lower corners of the plot). The light field cameras capture far narrower

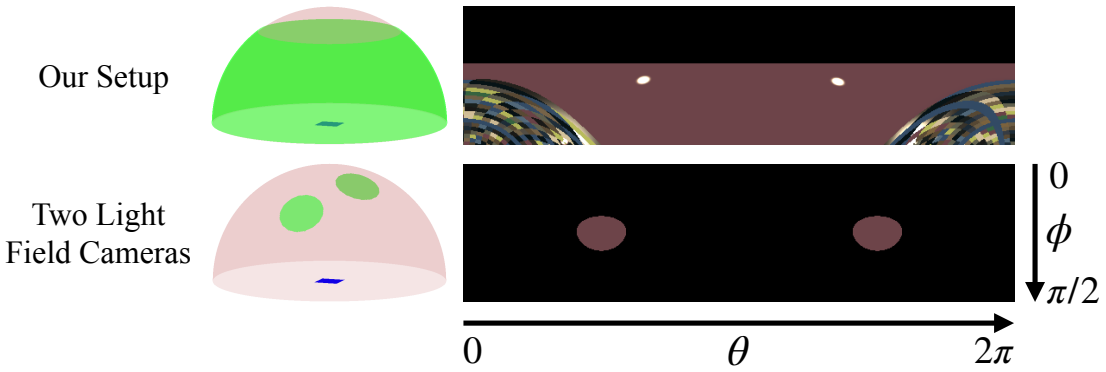

Figure 3.10: **Light field measurement comparison.** Comparison of the light field measurements of our setup versus that of a pair of light field cameras. The left shows a plot of the coverage of the hemisphere of outgoing rays from the central scene point for each configuration. The right shows this coverage visualized on a 2D plot of the spherical angles (azimuth θ and elevation ϕ). In this case, we show the measurements of a textured sphere in simulation with a surface point aligned with the central measurement point of our system. Note that our system provides rich measurements of this scene, with many observations of the reflectance of the scene point (red) as well as observations occluded by the rest of the sphere (texture in the bottom corners of the top plot). We utilize this rich information in particular in Chapter 5. The two light field cameras (bottom) capture far narrower cones that miss both reflectance information (no specular highlights) and geometric information (no occluded observations).

cones of the light field that miss large amounts of this information. This difference in coverage affects subsequent processing of the light field for a variety of applications, which we explore in Chapters 5-7.

Depth of Field Computation

We can extend the depth resolution analysis from Section 3.2.1 to account for the pixel pitch and non-uniform measurements of our camera. In that section, we derive a theoretical depth of field approaching 0 for a device capable of measuring the full hemisphere of light above each surface point. In practice, due the blurring caused by our pixel pitch and aperture along with the sparse and non-uniform measurements inherent with our imaging setup, we do not achieve this theoretical limit. To compute a more practical depth of field, we simulate a plane with a checkerboard pattern that is tilted slightly relative to the main axis of the ellipsoid. We then render an image perpendicular to the main axis using the light rays measured from our setup. The resulting rendering is shown on the top of Figure 3.11.

To compute the depth of field, we define the region of this image that is in focus, shown by the two red lines in the figure. This width corresponds to the horizontal extent of the intersection of the

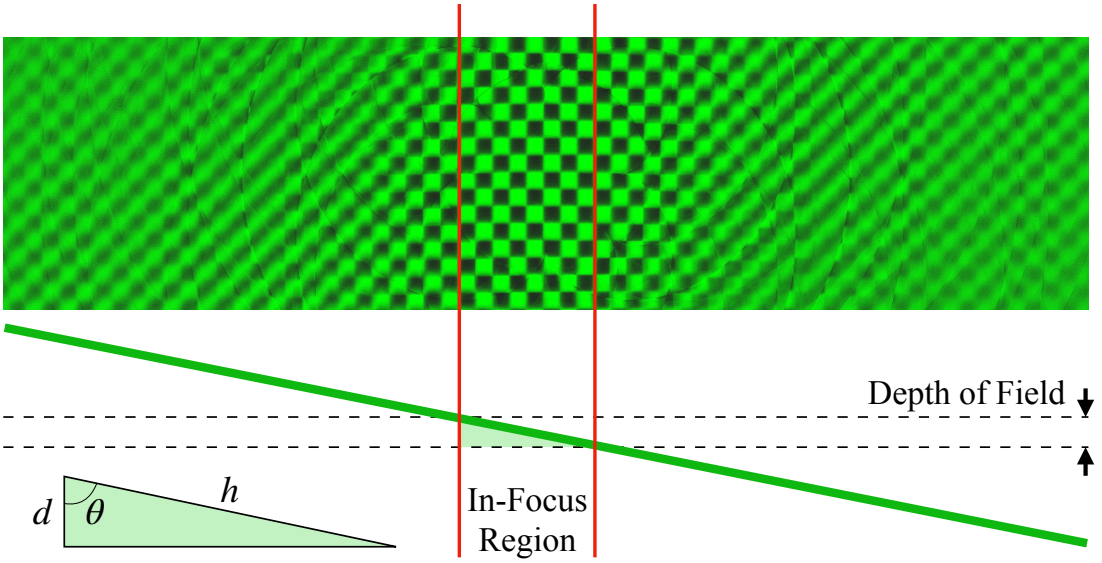

Figure 3.11: **Depth of field computation.** Illustration of the depth of field computation process. A rendering of a tilted plane is depicted on top, with the geometry of the tilted plane shown below. The triangle formed by intersection of the tilted plane with the depth of field is shown on the bottom left.

tilted plane with the depth of field, as shown in the bottom of Figure 3.11. This intersection forms the triangle shown in the bottom left of the figure. Since the in-focus region covers 8 squares, we compute the hypotenuse of this intersection triangle as 8 times the width of one square: $h = 8w$, with $w = 125\mu$m. Using this value, along with the angle $\theta = 88.85°$ describing the tilt of the plane, we compute the depth of field d in Equation 3.27.

$$d = h \cos\theta = 20\mu\text{m} \tag{3.27}$$

So, our setup has a simulated depth of field of 20μm. For our camera prototype, this may be further impacted by imperfections in the optical setup, which also impact the sharpness of in-focus regions.

The Missing Central Cone

A drawback of using an ellipsoidal mirror to remap wide-baseline light fields is that the span of angles that we measure is not continuous. Specifically, the imaging setup does not measure the cone of light from the scene that is centered around the major axis of the ellipsoid. This missing cone, marked in Figure 3.1, results in the dark central disk in the sub-aperture images in Figure 3.2. The missing central cone affects the performance of the shape estimation and rendering algorithms that we describe later.

Perhaps ironically, this is the very cone of light that is (easily) captured by a traditional light field camera via the use of its objective lens. However, since we observe both the gap in observed rays near the scene and the gap in pixels on our light field camera, we can consider applying additional optics within this region to help fill in the missing central cone. We consider a design for this in Section 3.4.

3.3 Implementation

We aim to produce the capabilities of the previously described imaging system in a lab prototype. In this section we describe the details of this prototype.

3.3.1 Lab Prototype

To implement this imaging setup, we use components based on the outline of the imaging setup defined earlier. For the light field camera, we use a Nikon Z6 camera with a hexagonally-packed microlens array mounted one focal length away from the sensor. Each lenslet has a diameter of 2.2mm and a focal length of 3mm. We additionally mount an aperture array above the microlens array, where each 200μm pinhole of the aperture aligns with the center of each lenslet. This camera as a whole is mounted above the object at the approximate location of the second focal point of the ellipsoidal mirror. The section of ellipsoidal mirror is mounted around the object, which is placed at the approximate location of the first focal point. The full setup is shown in Figure 3.12. For illumination, we create a small spotlight to illuminate the object without adding specular artifacts by reflecting off the mirror. With this base design, we do not calibrate or control the location of this light source. Section 3.4 discusses several variations on this optical design that include calibrated illumination. The figure also shows an additional camera and projector added to the setup to aid in calibration, which is discussed in Chapter 4. Further details about each component of this prototype are outlined in Table 31.

3.3.2 Light field camera construction.

A key element of the optical design is the light field camera. As noted previously, we modify a Nikon Z6 camera with a mounted microlens array and aperture mask to create this camera. We briefly discuss the modification procedure below.

We first open up the Z6 to gain access to the optics directly in front of the sensor. The lowpass IR/UV cut filter and dust reduction system are removed from in front of the camera sensor to reduce the thickness of the glass in front of the sensor. This allows the microlens array to be placed directly onto this glass at the proper focusing distance from the sensor. We 3D print a small spacer to place between

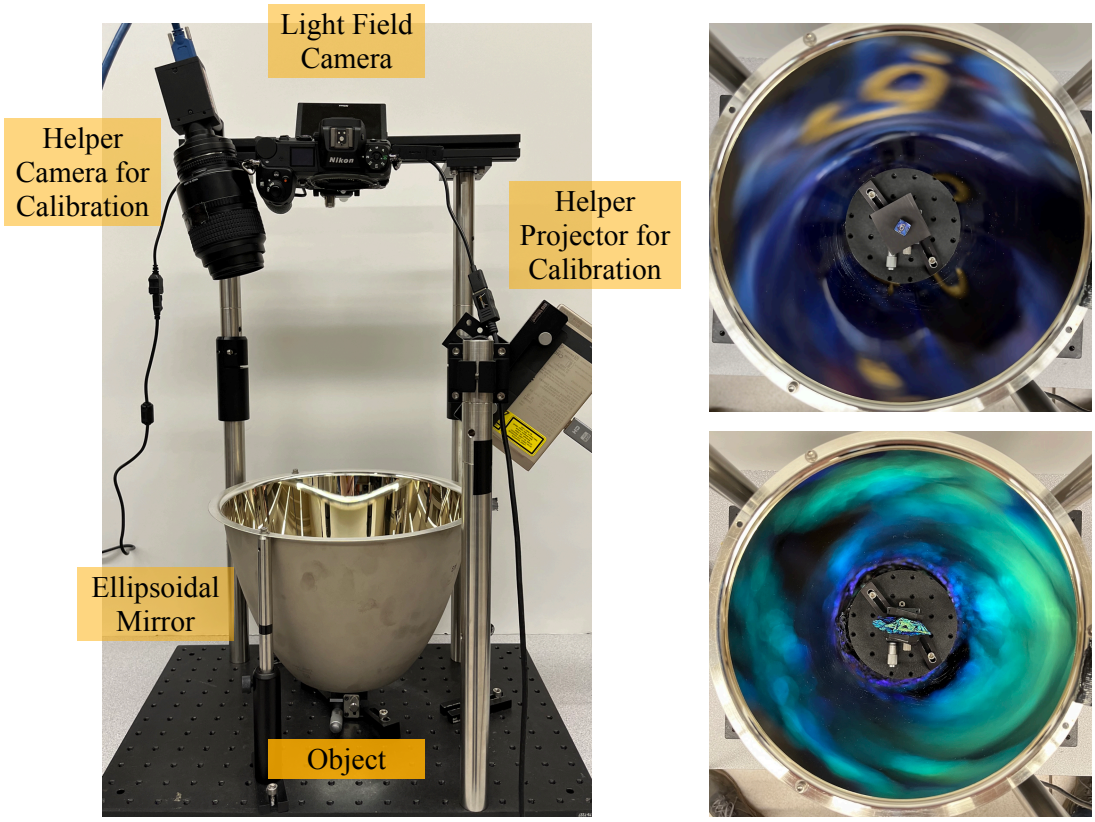

Figure 3.12: **Prototype.** Images of the real imaging setup. On the left is the entire camera, with the ellipsoidal mirror section below, the light field camera above in the center, and a helper camera and projector mounted around the apparatus for calibration. On the right are two views from the camera toward the mirror for two different objects. Each object is set on top of a vertical translation stage at the focal point. A list of components is provided in Table 31.

the sensor and the microlens array to ensure this focusing distance is achieved. The aperture mask, which aligns with the hexagonal packing of the microlens array, is placed directly onto the microlens array. This optical stack is held in place by the lens mounting ring on the Z6 and an additional 3D printed frame. The mounting ring presses the entire assembly stack (spacer - microlens array - aperture mask - frame) and holds it in place.

This mounting keeps the microlens array and aperture mask placement secure while limiting occlusions of the imaging sensor. Early results from this prototype utilized an alternative to the 3D printed frame, where several small, slightly compressible spacers were placed between the edges of the aperture mask and the frame of the mounting on the front of the camera. When the lens mounting ring is

Table 31: Components for the Light Field Camera

Component	Details
Light Field Camera	See individual components below.
Camera	Nikon Z6 mirrorless camera.
Microlens array	Hexagonally-packed microlens array from Fresnel Technologies with 2.2mm diameter lenses with a 3mm focal length (item #300).
Aperture mask	Custom-printed photomask. The aperture mask was designed to align each aperture opening with the center of each lenslet in the microlens array. Each opening was a square with a side of 200μm.
Helper Camera	Grasshopper with Nikkor 105mm macro lens
Helper Projector	Sony laser projector (MP-CL1A)
Ellipsoidal Mirror	Modified Optiforms off-the-shelf elliptical reflector E509F. The mirror was modified to have the bottom cutoff align with the focal point.

tightened back into place, these spacers hold the microlens array and aperture mask together against the glass in front of the image sensor, producing the desired optical stack with some occluded regions. Figure 3.13 shows captured light field images from this previous mounting scheme and the current 3D printed technique. The prior mounting was used in several results in this thesis, which can be noted by the occlusions present in the light field photographs.

3.4 Design Variations

The previous sections detail an imaging system capable of measuring wide-baseline light fields. This design was the implementation used in Chapter 5. However, several key limitations of this design, including the lack of illumination control and the missing central cone, limit the possibilities when moving to additional applications such as neural 3D reconstruction (Chapter 6) or iridescent reflectance capture (Chapter 7). This section outlines several modified designs

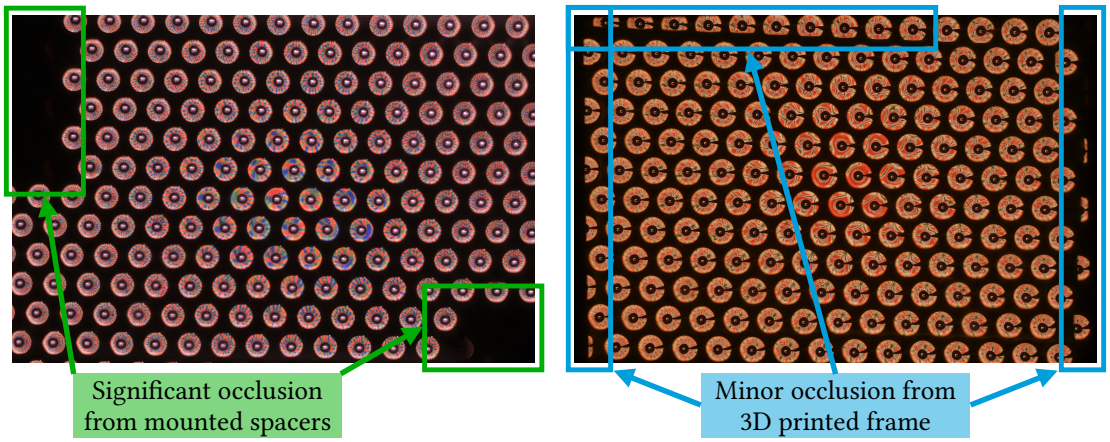

Figure 3.13: **Comparison of microlens mounting technique.** Light field images captured using older compressible spacer mounting technique (left) and more recent 3D printed mounting frame (right). We note the difference in the occlusion of measured light field rays.

3.4.1 Sparse Illumination Control

The base WBLF camera design does not provide for calibrated illumination; images are collected in uncalibrated illumination conditions. However, to avoid caustics caused by the illumination reflecting off the mirror, we use a collimated LED light source to illuminate the scene. We extend this idea to illumination control by mounting three calibrated LEDs around the upper rim of the ellipsoid, as shown in Figure 3.14. The light sources are calibrated via the helper camera, discussed further in Chapter 4. The other components of the prototype are identical to the prior setup, though we remove the helper projector in favor of alternate calibration techniques, also highlighted in Chapter 4. We utilize this system for sparse calibrated illumination for our work in Chapter 6.

3.4.2 Dense Illumination Control

Uncalibrated or sparsely-calibrated illumination is well suited for the primarily geometric applications in Chapters 5 and 6. However, our reflectance reconstruction work in Chapter 7 aims to capture high-frequency reflectance information in both space and angle by densely measuring the spatially-varying bidirectional reflectance distribution function (SVBRDF). This requires far more dense illumination control, including providing illumination to angles that are covered by the ellipsoidal mirror. We therefore introduce an illumination arm within the mirror itself, sacrificing some occluded rays for our target illumination control.

We mount a small strip of 12 LEDs over the object inside of the mirror, as shown in Figure 3.15. These

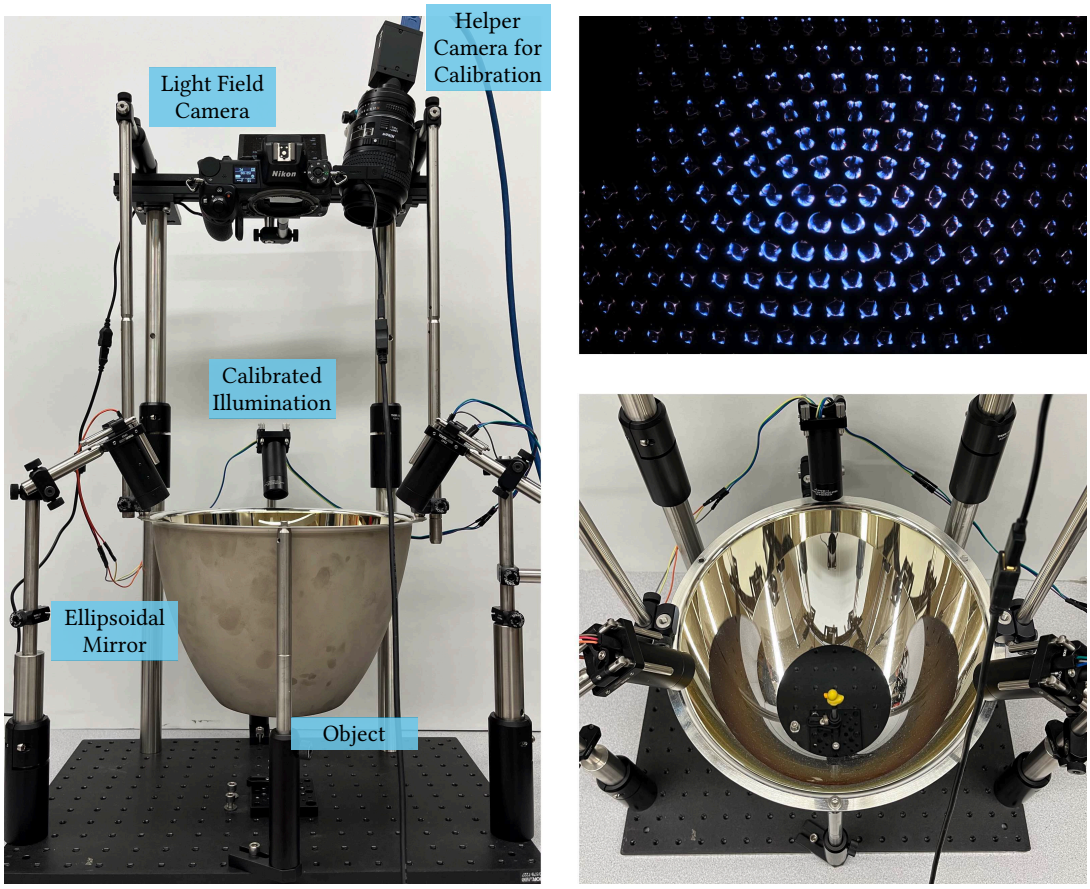

Figure 3.14: **Prototype 2.0.** Optical prototype of the wide-baseline light field camera. The ellipsoidal mirror is positioned with the object at the lower focal point and the light field camera at the upper focal point. Calibrated illumination and a calibrated helper camera provide the additional information needed in this technique. On the right are an example capture of a blue resistor and a close-up image showing the mirror in more detail.

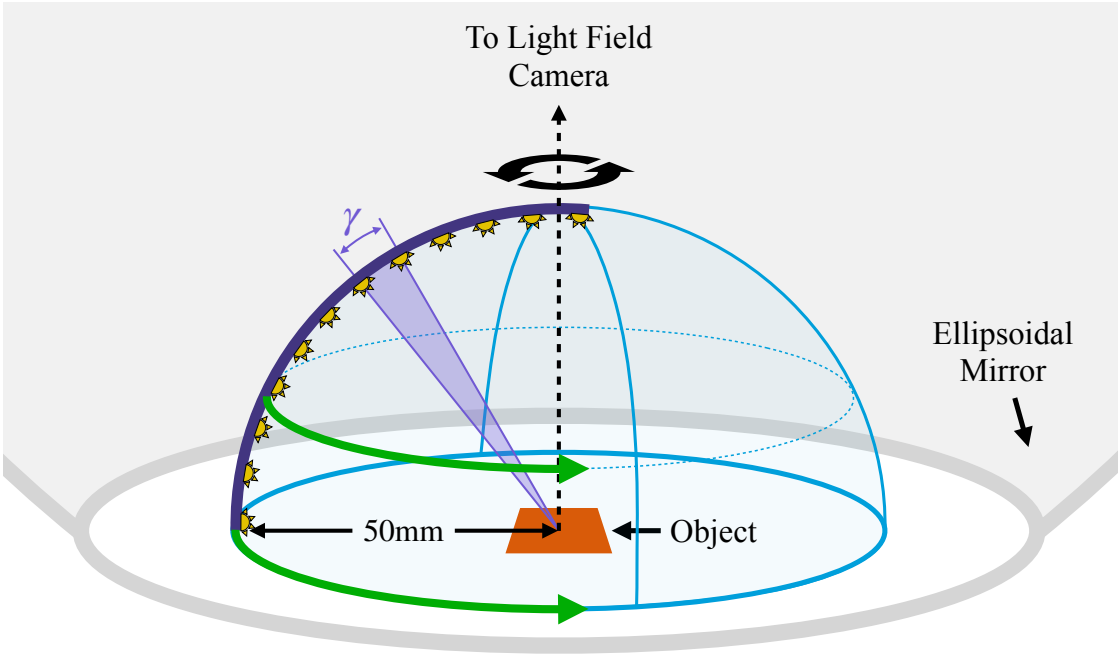

Figure 3.15: **Dense illumination control.** We mount an LED strip over the object within the ellipsoidal mirror to provide dense illumination control. This provides discrete illumination control in elevation with spacing $\gamma \approx 8^\circ$ and continuous illumination control in azimuth via a rotation stage. The LEDs are 50 mm away from the sample, so they must be calibrated as nearby point light sources.

LEDs span a circular arc 50 mm away from the sample, which gives us $\gamma \approx 8^\circ$ spacing in elevation. The illumination arm is mounted on a rotation stage to provide full illumination control in azimuth revolving around the primary optical axis. This does come with the cost of a small amount of occlusion, observed in Figure 3.16. This primarily limits us for retroreflective materials, where the observation directions of interest co-align with the illumination direction.

Contrary to typical light stage designs for illumination control, our LED mounting is forced to be very close to the object by the mirror geometry. This forces us to consider each light source as a nearby point light source as opposed to a directional light source. We account for this through additional calibration, where each point light source location is triangulated by a set of measurements. This procedure is described more fully in Chapter 4.

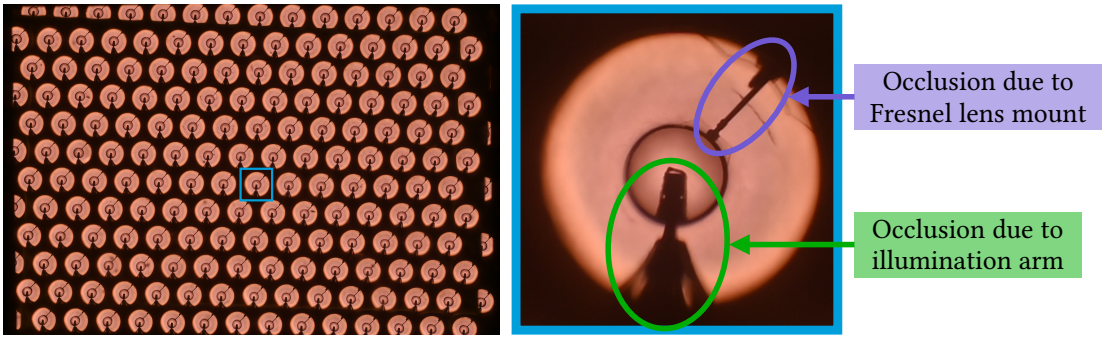

Figure 3.16: **Lens and illumination occlusion.** The Fresnel lens mounting and the illumination arm both introduce occlusion in our measurements. The occlusion from the illumination arm varies as the arm is moved through its range, while the Fresnel lens mounting is fixed. However, this occlusion earns us dense illumination control and many measurements of the central cone.

3.4.3 Extending the Measurement Cone

The analysis in Section 3.2 highlighted the missing central cone of the measured light field as a key limitation of this design. Not only are these rays not being measured, but there is an empty region within each lenslet view of the light field camera corresponding to this missing region. We can use these blank pixels to help fill in the missing cone by mounting a lens in the space between the mirror and the camera. In particular, using the lens placement detailed in Figure 3.17, we can fill a large portion of the missing cone for the focal point at the cost of a small amount of occlusion due to the lens mounting. This occlusion along with the now-utilized central pixels of each lenslet view are shown in Figure 3.16.

Since the light field camera employs an aperture mask of a pinhole array to prevent blurring, we can adjust the focal length and placement of this lens to adjust the space-angle tradeoff for these rays. The pinhole array design takes what would otherwise be defocus blur and turns it into spatial sampling, so we can increase the defocus to broaden the area that we can measure. The placement options discussed here are visualized in Figure 3.17.

We first choose a starting placement and focal length based on the widest gap in the existing measured light rays. This gap occurs at height 73.21 mm above the lower focal point of the ellipsoid, with the total gap between focal points being $h = 508.0$ mm. Using the focal length equation, with $u = 73.21$ mm and $v = 508.0 - 73.21 = 434.79$ mm, we find

$$f = \frac{uv}{u+v} = 62.66 \text{ mm} \tag{3.28}$$

To get the optimal light collection, we aim to fill nearly the entire width of the gap in rays with a

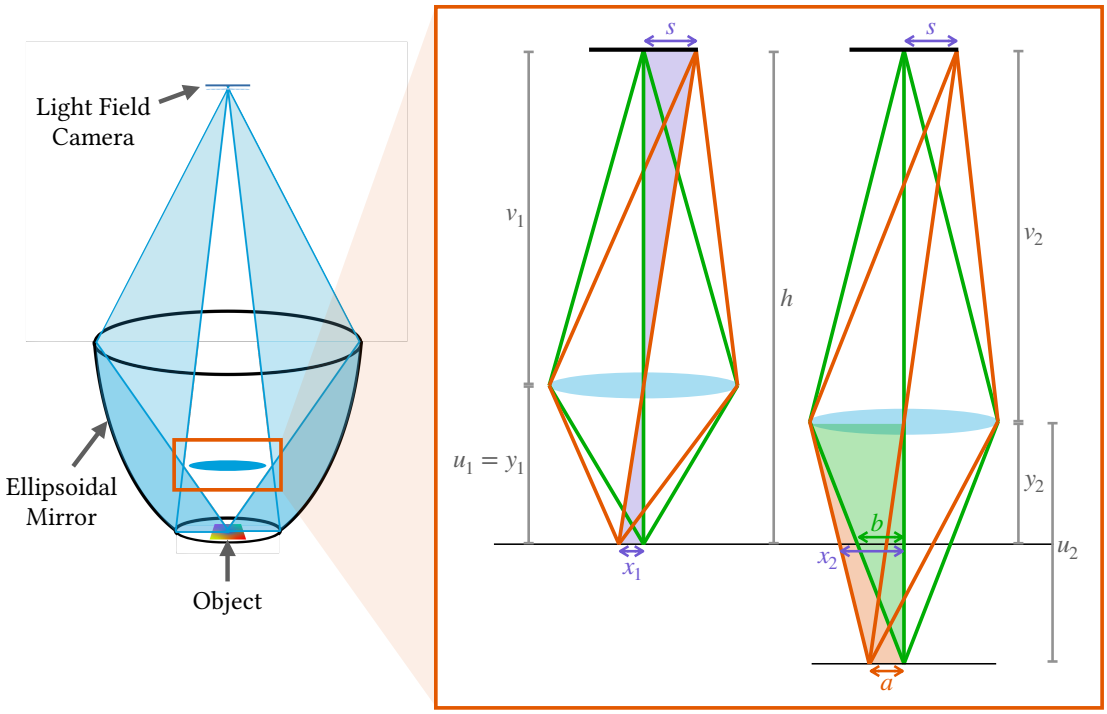

Figure 3.17: **Optimal lens placement for extended measurements.** We use an ellipsoidal mirror to map the light field from the object at one focal point to a light field camera at the other. The light field camera consists of a microlens array mounted in front of an imaging sensor, with an aperture mask aligned to each lenslet to minimize measurement blur. Due to the ellipsoidal geometry, the central cone of light is not measured; however, we mount a Fresnel lens within the ellipsoidal mirror imaging system to expand the angular range that we can measure. We optimize the lens placement to limit occlusion and to achieve the desired spatial coverage. The design parameters used in the text are defined on the diagram.

mounted lens, which would require a focal length of $f = 62.66\,\text{mm}$ and a diameter of over 5 inches, which is unachievable with spherical lenses. So, we use a Fresnel lens, as the aperture array also helps mitigate aberrations found in such a lens. Considering available Fresnel lenses that would be nearly in focus at this location, we use a lens with a 4 inch ($d = 101.6\,\text{mm}$) diameter and a focal length of 71.12 mm.

For our first lens mounting option, we consider this lens placed in focus between the sensor and the object, as in the left ray diagram in Figure 3.17. Using the focal length equation, we set $f = 71.12\,\text{mm}$, $h = 508.0\,\text{mm}$, and $v_1 = h - u_1$ to find mounting height $y_1 = u_1$ that would bring this lens to focus on the object:

$$y_1 = u_1 = \frac{h - \sqrt{h^2 - 4fh}}{2} = 85.52 \text{ mm} \tag{3.29}$$

To consider the spatial extent of the measured light field, we can use the magnification of this lens placement, since we are in focus. We consider the size of a full-frame sensor (using the half-width $s = 17.5\,\text{mm}$) magnified down to the object plane (producing total width $w_1 = 2x_1$):

$$x_1 = \frac{u_1}{v_1} s, \tag{3.30}$$

$$w_1 = 2x_1 = 7.08 \text{ mm} \tag{3.31}$$

This spatial extent is smaller than the extent achieved through the ellipsoidal mirror, where we aim for a target object size of $20{\times}20\,\text{mm}^2$ shown in Figure 3.9. To more closely align these extents at the cost of angular resolution, we move the lens down to the gap in the rays mentioned previously, with $y_2 = 73.21\,\text{mm}$, as shown on the right in Figure 3.17. We use the defocus to broaden our spatial extent due to the narrow aperture used in front of each microlenslet, which causes each light field camera pixel to approximately measure just a single light field ray. To compute the new spatial extent $w_2 = 2x_2$, we utilize additional similar triangles shown in green and orange on the diagram:

$$b = \frac{u_2 - y_2}{u_2} \frac{d}{2}, \quad a = \frac{u_2}{v_2} s, \tag{3.32}$$

$$x_2 = \frac{y_2}{u_2} a + b = \frac{y_2}{v_2} s + \frac{u_2 - y_2}{u_2} \frac{d}{2}, \tag{3.33}$$

$$w_2 = 2x_2 = 20.02 \text{ mm}. \tag{3.34}$$

This provides a similar spatial extent as the ellipsoidal mapping, which allows us to utilize this new information in a similar way.

We consider the addition in measurements that this lens addition provides. Similar to the analysis visualized in Figure 3.7, we visualize the capture of a color-mapped $20\,\text{mm} \times 20\,\text{mm}$ plane by our light field camera. Figure 3.18 shows this analysis for the imaging system with the addition of the Fresnel lens. We observe that the angular range of the measurements is increased; in particular, we extend the range of elevation angles to $0 < \phi < \pi/2$. The azimuth measurements and spatial measurements have similar ranges to those observed previously, but they feature additional measurements within that range captured by the lens.

Additionally, we consider our measurements on a point-by-point basis, similar to the analysis of Figure 3.8. Figure 3.19 visualizes the continuous region of rays that reach our sensor from a scene

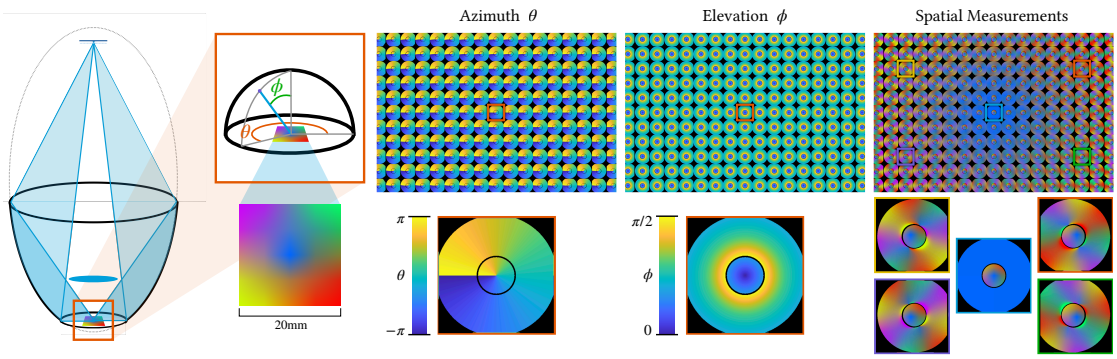

Figure 3.18: **Visualization of captured light field with added Fresnel lens.** We place a color-coded plane at the object focal point and capture a wide-baseline light field image, similar to Figure 3.7. The far right column shows this captured image, which highlights the spatial extent of our measurements. The middle two columns visualize the angular information encoded in the captured light fields; we color-code the captured object-space azimuth angle θ and elevation angle ϕ according to the convention defined in the diagram. The added Fresnel lens provides additional measurements within each lenslet view and extends the angular range of ϕ, as observed in the center of each lenslet view. Note that we closely consider only one lenslet view each for θ and ϕ; the angular measurements are nearly uniform across different lenslets with only minor variations that are indistinguishable in this visualization.

point at the focal point along with three points shifted by 5mm away from the focal point along each coordinate axis. The measurements through the Fresnel lens appear at the apex of the hemisphere under consideration. Our mounting placement for this lens does not fill the entire missing cone, but it does allow measurements through the lens even as we shift away from the focal point.

3.4.4 Prototype with Dense Illumination and Additional Lens

We implement a prototype using the dense illumination control and extended measurement cone for capturing SVBRDFs in Chapter 7. This prototype is shown in Figure 3.20. The strip of LEDs is mounted in an arc over the object imaging area inside the mirror, with the base of the mounting connected to a rotation stage to provide full illumination control. An additional shield of dark material is mounted around the LED strip to reduce caustics directly reflecting off the mirror and reaching the sensor, particularly when the fresnel lens is added. This fresnel lens is mounted above the object as described previously, using minimal mounting hardware to minimize the occlusions to the mirror measurements. The mirror, lens, and light sources are calibrated with the aid of an added helper camera aimed at the lower focal point of the system and a small LCD placed around the lower focal point, detailed in Chapter 4.

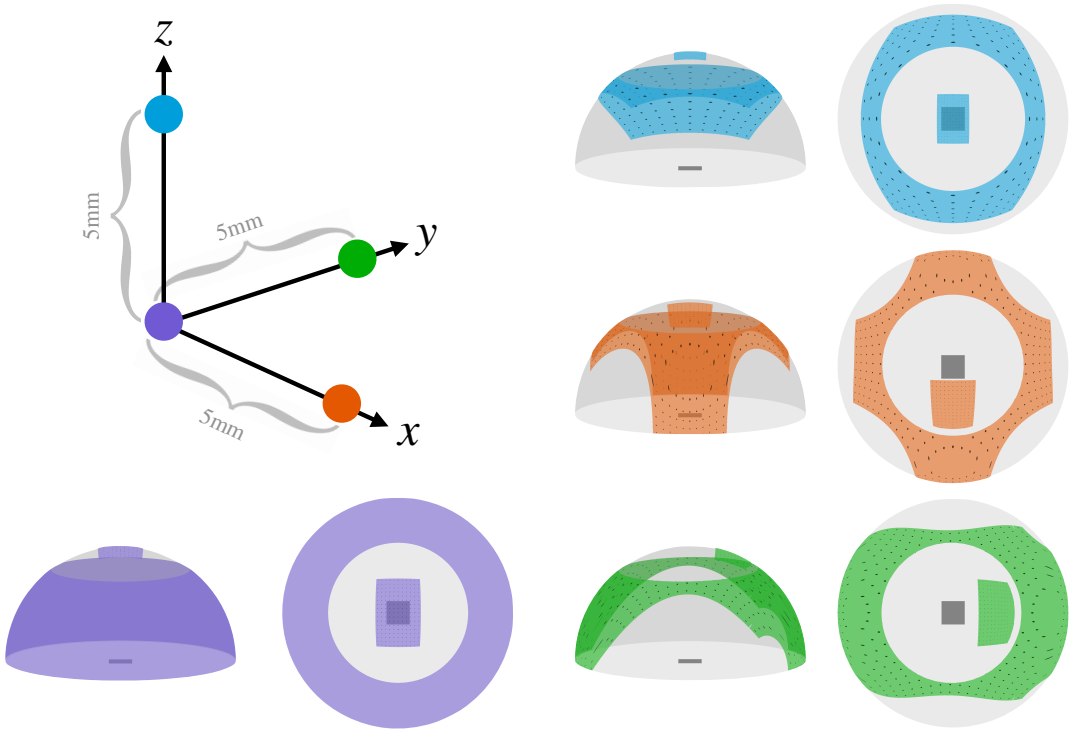

Figure 3.19: **Point-wise coverage with added Fresnel lens.** Given a full hemisphere of possible observation directions from a single point, we visualize the subset of this space that can actually be measured using the imaging system that we propose with the addition of the Fresnel lens, similar to the analysis of Figure 3.8. We again consider this region for one point at the object focal point and three points shifted by 5mm away from the focal point along each coordinate axis, illustrated on the left. Accounting for the space-angle tradeoff of the light field camera, the dark points within each region specify the actual measured rays. The effect of the added Fresnel lens is observed in each plot, as we fill in a portion of the central cone for each point; the amount of the cone that we can fill in is relatively small to allow us to achieve greater spatial coverage with the lens.

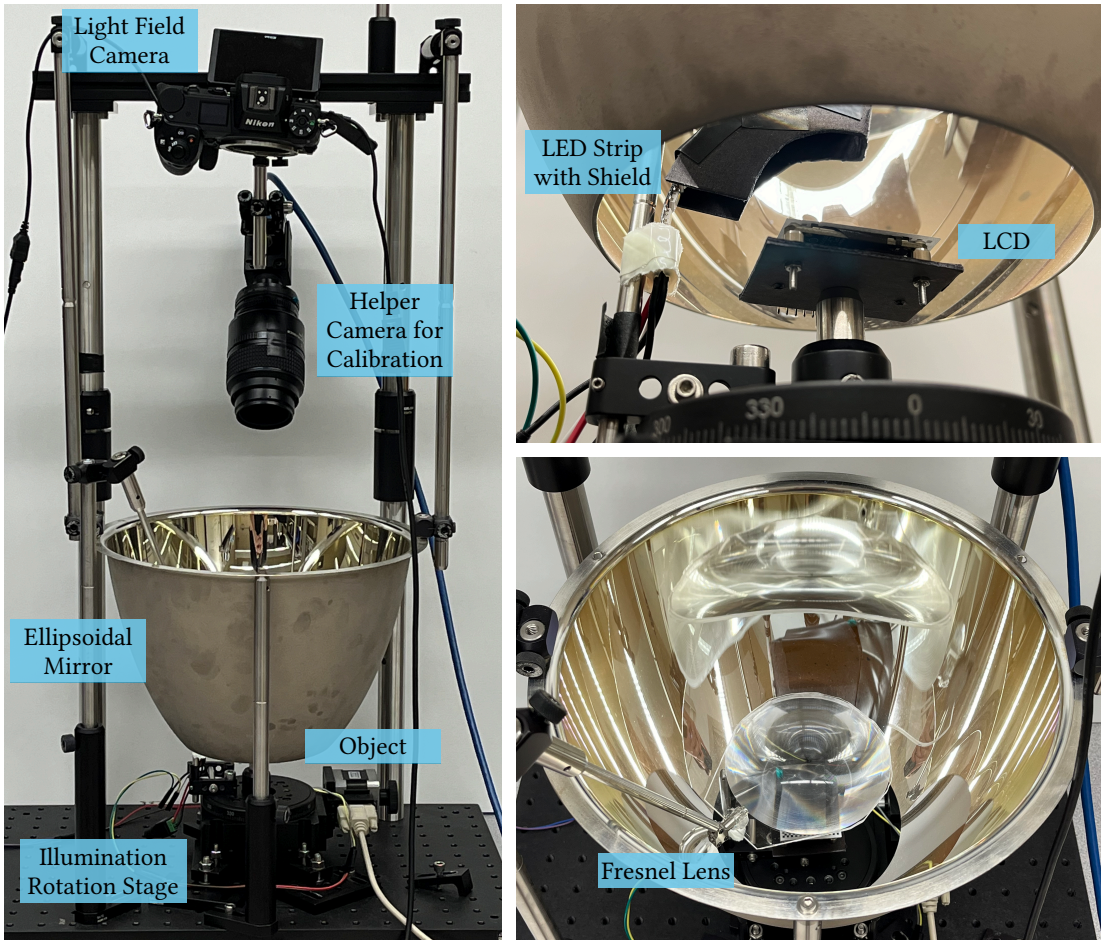

Figure 3.20: **Prototype 3.0.** Optical prototype of the wide-baseline light field camera for measuring SVBRDFs. The ellipsoidal mirror is positioned with the object at the lower focal point and the light field camera at the upper focal point. A calibrated helper camera and LCD aid in calibration and plane detection. On the right are close-ups of the key additions to the design: illumination control via an LED strip (top) and extended measurement range via a Fresnel lens (bottom).

3.5 Discussion

In this chapter, we introduced a wide-baseline light field camera design that utilizes the unique properties of an ellipsoidal mirror. We additionally analyzed the resulting light fields captured with this imaging system, including properties of resolution, angular range, coverage, and depth of field. To implement this optical design, we presented a prototype and detailed the necessary components. Finally, we presented several variations to the optical design that add illumination control and extend the measurement capabilities of the system at the cost of additional capture time or occluded measurements.

Limitations. The proposed setup inherits a number of limitations stemming from its use of mirrors and light field cameras. The ellipsoidal mirror presents a number of different constraints. Its size determines the spatial extend of the light field that we can measure; as Figure 3.9 shows, we experience a fall-off in measurement density as we move away from the focal point of the ellipsoid. This limits us to scenes with dimensions around 1 or 2 cm in each direction. The quality of the measured light field also relies on the optical quality of the ellipsoidal mirror and the light field camera, both of which have limited commercial options. This produces often significant abberations in the mapping of the light field; we aim to account for this through extensive calibration described in Chapter 4. Further, the use of ellipsoidal mirrors to relay the light field from one focus to another naturally results in a missing cone, which we mitigate through the addition of the Fresnel lens. However, this only covers an incomplete portion of the missing cone. Finally, a different set of limitations stem from the use of light-field cameras, which have implicit tradeoffs in achievable spatial and angular resolutions; we inherit these tradeoffs as well, and they affect the resolution and measurement density compared to theoretical possibilities.

4 Calibration

In the previous chapter, we discussed the optical design and implementation for the wide-baseline light field camera. The optical mapping through the ellipsoidal mirror provides key benefits in the measurements we can achieve, but introduces key challenges in the calibration of the imaging system. For synthetic data, this is performed simply by knowing the exact geometry and alignment of the ellipsoidal mirror along with other imaging optics. However, practical limitations in ellipsoidal mirror manufacturing along with strict alignment requirements for ray tracing the system result in difficult calibration.

For example, Figure, 4.1 shows a light field image captured of the material shown on the left using our lab prototype. With a proper ellipsoidal geometry for the mirror, we would expect these straight lines on the material to map to smooth curves in the light field image, similar to the straight lines of the checker pattern depicted from simulation data in the figure. However, these lines are not producing smooth curves based on small deformations in the geometry of the mirror. Calibrating this geometry in such a way as to be able to directly ray trace from the light field camera to the object is an intractable task. Instead, we introduce a novel calibration procedure that directly fits object-space light rays to each pixel of the light field camera, thereby bypassing any need to know the exact mirror geometry or alignment.

Contributions. In total, this chapter introduces a novel calibration algorithm that allows the wide-baseline light field design to be used for practical applications beyond simulation. In particular, we make the following contributions.

- *Direct pixel-to-object ray calibration.* Attempting to trace light field rays from our light field camera to the scene space relies on a detailed understanding of the geometry and positioning of all optical components, which is unreliable. Instead, we develop a novel pixel-to-ray calibration technique that directly maps light field camera pixels to the corresponding measured ray in the scene space.

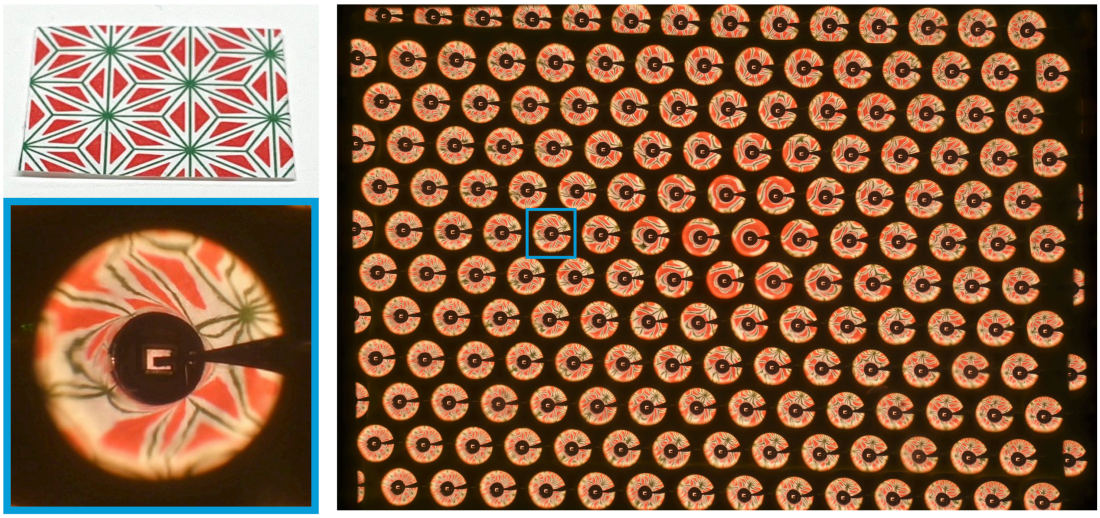

Figure 4.1: **Impact of the mirror quality.** We show a light field image captured of the patterned paper shown on the left along with a close-up of a lenslet view. We observe that the straight lines of the scene do not get mapped to smooth curves as expected; rather, the deviations in mirror geometry produce deviations in the light field mapping that need to be accounted for in calibration.

- *Iterative calibration to optimize grazing angle results.* We find that limitations of the imaging system aiding in calibration leads to unreliable ray estimations, particularly when calibrating grazing angles. We introduce an iterative calibration technique that jointly optimizes for the scene space ray parameters as well as the unreliable calibration parameters produced by our helper camera. This produces more reliable calibrated rays from our imaging system.

- *Calibrated illumination* Several applications of our imaging system apply active illumination control to enhance the measurements of our system. We introduce the techniques required to calibrate these sources of illumination for both sparse and dense illumination control.

4.1 Pixel-to-Object Ray Calibration

We aim to directly connect pixels on the light field camera to rays in the object space. We achieve this by first determining a set of 3D points through which the rays from each pixel pass through. Once we have this set of points, we fit a ray passing through these points for each pixel. This gives us the desired pixel-to-object ray calibration. The following sections go into more detail about the procedure and hardware used.

4.1.1 3D Point Mapping

To determine a set of 3D points observed by each pixel, we produce calibration patterns on a set of known 3D planes in the object space. To achieve this in practice, we utilize a small LCD and the calibrated helper camera aimed at the object-space focal point. First, we display a checkerboard on the LCD captured by the helper camera to determine the location of the screen in camera coordinates using standard techniques. This gives us our known 3D plane location and creates a mapping from LCD pixel to 3D camera coordinate, since this checkerboard is displayed on a known set of LCD pixels.

We then display a series of 8-bit Gray code patterns on the display, capturing 16 images on the light field camera for LCD rows and 16 images for columns. We capture two images per bit, displaying the Gray code pattern for one and the inverse of the Gray code pattern for the other. By comparing the two, we produce a more robust decoding of the Gray code bit sequence. The Gray code patterns and corresponding captured images are shown in Figure 4.2. This creates a unique 16-bit binary sequence for each pixel of the display. Each pixel on our light field camera, which is observing a single point on the display, should image the binary sequence corresponding to the pixel at that point. By decoding these sequences for each light field pixel, we generate a correspondence between light field pixels and LCD pixels. Since we know the correspondence from LCD pixel to 3D point from the checkerboard, we now have a correspondence from light field pixel to 3D point.

We repeat this process for a large number of LCD plane positions. The different ray intersections with each plane produces a full set of 3D points for each light field pixel.

4.1.2 Points to Rays

To turn these 3D points into calibrated rays, we fit the best ray through the observed points for each light field pixel. While this should be a simple task, the presence of outliers in the 3D point data requires us to utilize a robust line fitting algorithm. We have employed both a RANSAC algorithm based on SVD line fitting (Algorithm 1) and an iterative line fitting based on the Huber loss function (Algorithm 2). We use the Huber loss between data x and target y with parameter δ defined as

$$\ell = H(x, y, \delta) = \begin{cases} 0.5 \cdot (x - y)^2, & \text{if } |x - y| < \delta \\ \delta \cdot (|x - y| - 0.5 \cdot \delta), & \text{otherwise} \end{cases} \tag{4.1}$$

We use $\delta = 0.1$ for the Huber loss in Algorithm 2. The Huber loss combines the benefits of the mean-squared error loss and the $L1$-loss. The δ-scaled $L1$ loss when $|x - y| \geq \delta$ reduces sensitivity

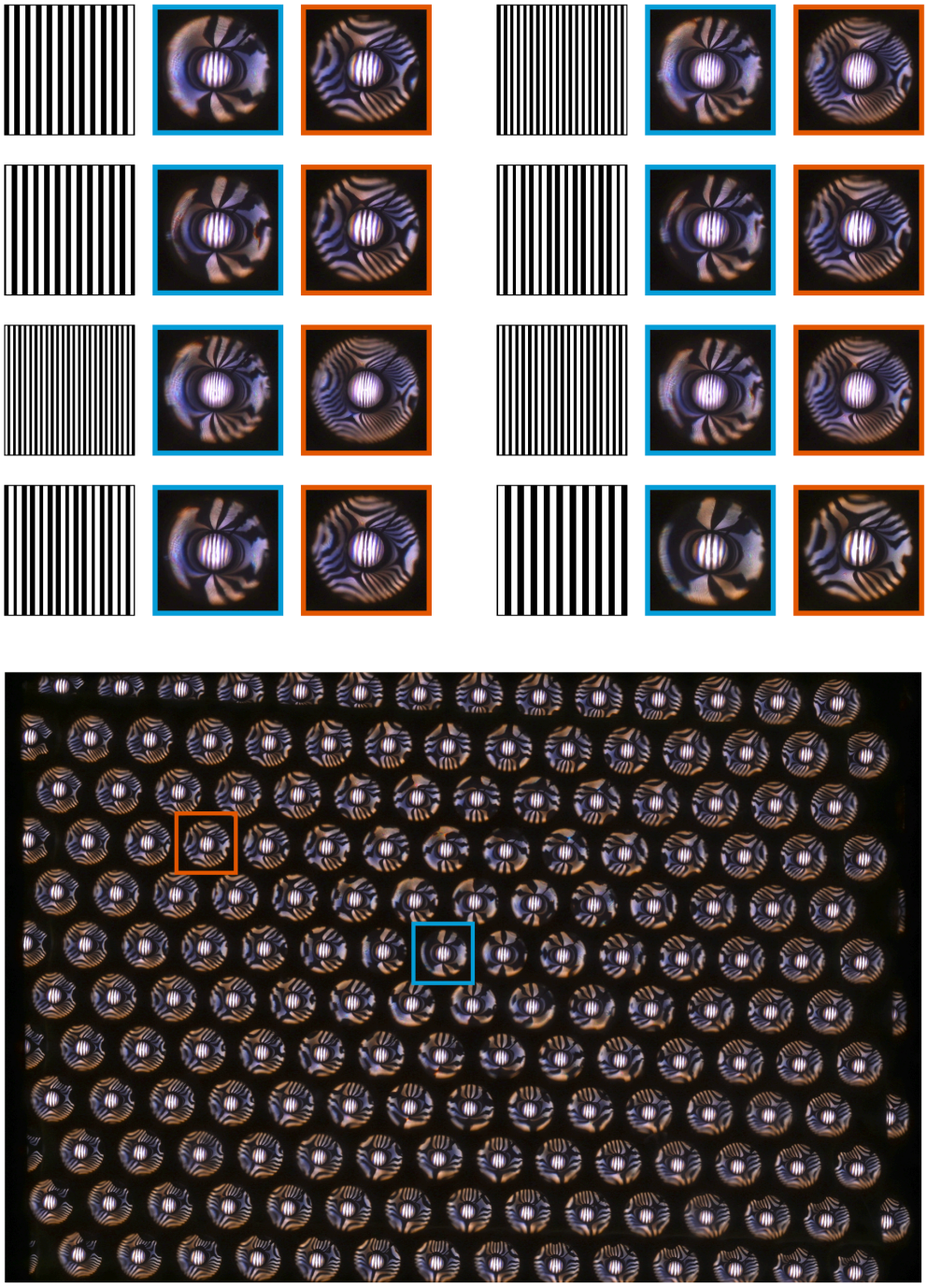

Figure 4.2: **Captured sequence of 8-bit Gray code.** We show two lenslet views of the captured light field images for each bit of the 8-bit Gray code sequence. These images encode the pixel location on the LCD. The entire light field image for one bit is shown below for reference. A similar sequence of eight images is captured oriented row-wise to complete the encoding.

to outliers, which is crucial for our setup, and the squared loss when $|x - y| < \delta$ ensures smoothness around $|x - y| = 0$.

In practice, we find that the iterative line fitting algorithm outperforms the SVD line fitting algorithm, as the resulting rays are not constrained to pass the a given point. SVD line fitting produces a ray that must pass through the zero-mean point of a set of points. By using RANSAC, we promote outlier-free sets of points, but this zero-mean point may still not be ideal. Our iterative algorithm is more flexible to the data, though it comes at the cost of computation time.

4.1.3 Sources of Error in 3D Point Estimation

Many sources of error introduce challenges into the calibration process. The first source of error is the homography estimate from the helper camera to the captured checkerboard. At the scale we are operating in (within a few centimeters), any small errors in the detected 3D location of the checkerboard plane can propagate significantly through the remaining calibration. In particular, we find that the z-coordinate of the 3D location (along the helper camera's optical axis) experiences the most noticeable errors. Due to practical constraints, we have to mount our helper camera at a distance from the scene much greater than the size of the scene. This leads to imaging a narrow field of view over a small range of locations, so our helper camera is operating in a nearly orthographic setting; thus, we do not observe much difference in the projected size of the checkerboard squares at different depths from the camera. We observe this error most prominently when using calibration planes of different orientations.

The second source of error is Gray code decoding errors along borders between changing pixels. This error is limited to begin with by using Gray codes instead of a standard binary sequence; this forces only one bit of the pattern to change for sequential locations. We further mitigate this error by using long-run Gray codes, which maximize the distance between changes for a particular bit. This corresponds to wider patterns being displayed on the LCD, which reduces the number of edges at which errors can occur. In practice, this works well to reduce this source of error, and our use of outlier-sensitive algorithms ensures that it does not impact the final calibrated rays.

Finally, we observe errors due to LCD pixel pitch limitations. Our light field camera is capable of resolving very fine spatial differences, even down to the pixel pitch of our LCD in certain areas. This produces artifacts in the measured data, where additional lines corresponding to gaps between pixels and color-shifting observations of the pixel colors appear, as shown in Figure 4.3. We mitigate this error by capturing robust Gray code data. We capture both regular and inverted Gray codes and compute our binary bit based on the relative intensity of the captured measurements for these two patterns. The

ALGORITHM 1: SVD Line Fitting with RANSAC

input : P — A set of 3D point correspondences of size $p \times n \times 3$, where p is the number of light field pixels in image **I** and n is the number of observed planes.

output: O — A set of ray origins of size $p \times 3$

D — A set of ray directions of size $p \times 3$.

```
for p ∈ I do
    points ← P_p;                          /* Get the set of point correspondences at pixel p */
    numValid ← valid(points);   /* Number of valid points (rays do not hit all n planes) */
    if numValid < 6 then
        continue;          /* With too few points, we continue and mask out this pixel p */
    end
    for α ∈ [0.8, 0.4],  β ∈ [0.8, 0.6] do
        /* Perform two passes of RANSAC algorithm, first with a larger subset ratio α
           and a more strict inlier requirement β.  Attempt a second pass if the first
           pass fails with reduced α and β.                                              */
        for i ∈ 1..20 do
            m ← α · numValid;
            subPoints ← subset(points,m);                 /* Choose random subset of m points */
            subPoints ← subPoints − mean(subPoints);                      /* Set to zero mean */
            [U, S, V] ← svd(subPoints);                     /* Perform SVD on these subPoints */
            rayDir ← V[:, 0];                    /* Get ray direction from first column of V */
            numInliers ← inliers(points, rayDir);              /* Count number of inlier points */
            if numInliers > β · numValid then
                O_p ← mean(subPoints);
                D_p ← rayDir;
                success ← True;
                break;                      /* For success, store ray information and break */
            end
        end
        if success then
            break;                 /* For a successful fitting, continue to next pixel p */
        end
    end
end
```

ALGORITHM 2: Iterative Line Fitting

input : P — A set of 3D point correspondences of size $p \times n \times 3$, where p is the number of light field pixels in image **I** and n is the number of observed planes.

output: O — A set of ray origins of size $p \times 3$

D — A set of ray directions of size $p \times 3$.

for $p \in$ **I do**

```
/* We perform this iteration for each pixel p here.  In practice, this is
   vectorized to operate on all pixels at once.                          */
points ← P_p;                          /* Get the set of point correspondences at pixel p */
o ← mean(points);                                          /* Initialize ray origin */
for α ∈ 1..10 do
    x ← subset(points, 2);                                 /* Choose two random points */
    d_α ← normalize(diff(x));   /* Find normalized direction vector between two points */
end
d ← median(d_α);       /* Initialize ray direction as median of 10 point differences */
for i ∈ 1..1000 do
    pDiff ← points − o;            /* Compute difference between points and ray origin */
    res ← norm(pDiff − (pDiff • d) · d);       /* Compute residuals between points and ray */
    ℓ ← huberLoss(res, 0);          /* Compute Huber loss between the residual and zero */
    [Δo, Δd] ← gradients(ℓ);                                    /* Compute gradients */
    o ← o − Δo · lr;                          /* Update origin with learning rate lr */
    d ← d − Δd · lr;                        /* Update direction with learning rate lr */
end
O_p ← o;                                /* Save final origin and direction for pixel p */
D_p ← d
```

end

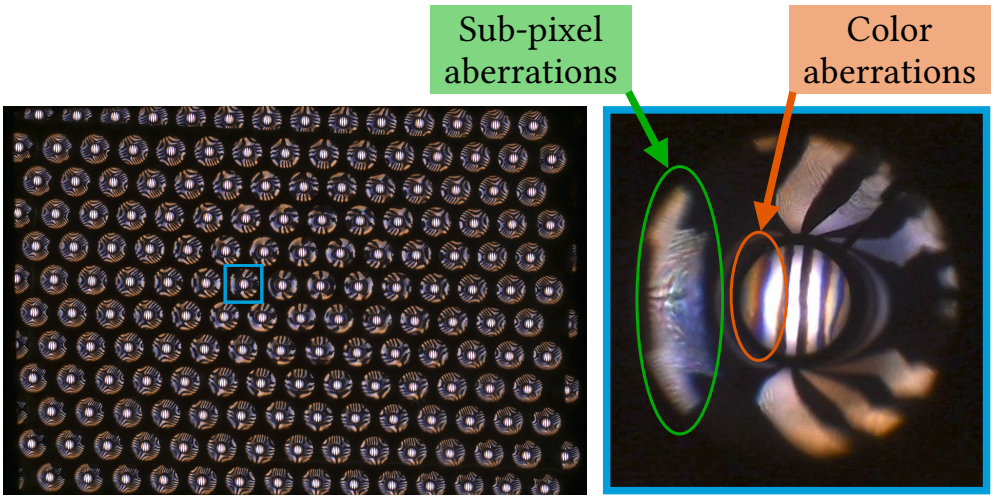

Figure 4.3: **Aberrations in captured calibration images.** We show the captured light field of one Gray code sequence along with a closer look of one lenslet view. We observe aberrations due to the broad angular range and our ability to resolve finer spatial features than the LCD pixel pitch.

artifacts impact the overall intensity of the measurements, but the relative intensity between the regular and inverted codes still successfully correlates to the displayed pattern.

4.2 Iterative Calibration Optimization

Due to these various errors, successful calibration proves challenging. In particular, the homography estimate mapping the LCD plane to camera coordinates is unable to match the precision needed, especially along the helper camera's optical axis. So, we develop a joint optimization approach to both solve for the best-fit ray for each pixel's corresponding 3D points as well as update the estimated plane parameters of the display. We continue to employ the Huber loss to update both the best fit rays and the parameters of the planes. This optimization is described in detail in Algorithm 3 and is summarized in Figure 4.4.

We visualize the Huber loss through various stages of our optimization in Figure 4.5. As noted in Algorithm 3, we initialize the line fit for every light field pixel using Algorithm 2. This provides the initial ray origins and directions by fitting to the noisy, erroneous plane parameters. Then, we perform the alternating optimization outlined in the rest of Algorithm 3. Due to memory and time constraints, particularly when performing the global update of plane parameters, we perform batch gradient descent with 1000 light field pixels at a time. This creates the stochastic noise found in the batch loss on the right.

ALGORITHM 3: Iterative Line Fitting with Plane Parameter Updates

input : P — A set of 3D point correspondences of size $p \times n \times 3$, where p is the number of light field pixels in image **I** and n is the number of observed planes.

N — A set of surface normals for the observed planes of size $n \times 3$.

X — A set of origin points for the observed planes of size $n \times 3$.

output: O — A set of ray origins of size $p \times 3$

D — A set of ray directions of size $p \times 3$.

$[O, D] \leftarrow$ lineFit(P); `/* Initialize ray origin and direction using Algorithm 2 */`

while *not converged* **do**

- points $\leftarrow$ subset(P, 1000); `/* Get a random subset of 1000 pixels */`
- $\mathbf{o} \leftarrow$ subset(O, 1000);
- $\mathbf{d} \leftarrow$ subset(D, 1000);
- points $\leftarrow$ update(points, N, X); `/* Update points based on updated plane parameters */`
- **for** $i \in 1..10$ **do**
  - pDiff $\leftarrow$ points $-$ $\mathbf{o}$; `/* Compute difference between points and ray origin */`
  - res $\leftarrow$ norm (pDiff $-$ (pDiff $\bullet$ $\mathbf{d}$) $\cdot$ $\mathbf{d}$); `/* Compute residuals between points and ray */`
  - $\ell \leftarrow$ huberLoss(res, 0); `/* Compute Huber loss between the residual and zero */`
  - $[\Delta\mathbf{o}, \Delta\mathbf{d}] \leftarrow$ gradients(ℓ); `/* Compute gradients w.r.t. rays */`
  - $\mathbf{o} \leftarrow \mathbf{o} - \Delta\mathbf{o} \cdot$ lr; `/* Update origin with learning rate lr */`
  - $\mathbf{d} \leftarrow \mathbf{d} - \Delta\mathbf{d} \cdot$ lr; `/* Update direction with learning rate lr */`
- **end**
- pDiff $\leftarrow$ points $-$ $\mathbf{o}$; `/* Compute difference between points and ray origin */`
- res $\leftarrow$ norm (pDiff $-$ (pDiff $\bullet$ $\mathbf{d}$) $\cdot$ $\mathbf{d}$); `/* Compute residuals between points and ray */`
- $\ell \leftarrow$ huberLoss(res, 0); `/* Compute Huber loss between the residual and zero */`
- $[\Delta N, \Delta X] \leftarrow$ gradients(ℓ); `/* Compute gradients w.r.t. plane parameters */`
- $N \leftarrow N - \Delta N \cdot$ lr; `/* Update plane normals with learning rate lr */`
- $X \leftarrow X - \Delta X \cdot$ lr; `/* Update plane origins with learning rate lr */`
- $O_{subset} \leftarrow \mathbf{o}$; `/* Update output origin and direction for the given subset */`
- $D_{subset} \leftarrow \mathbf{d}$;

end

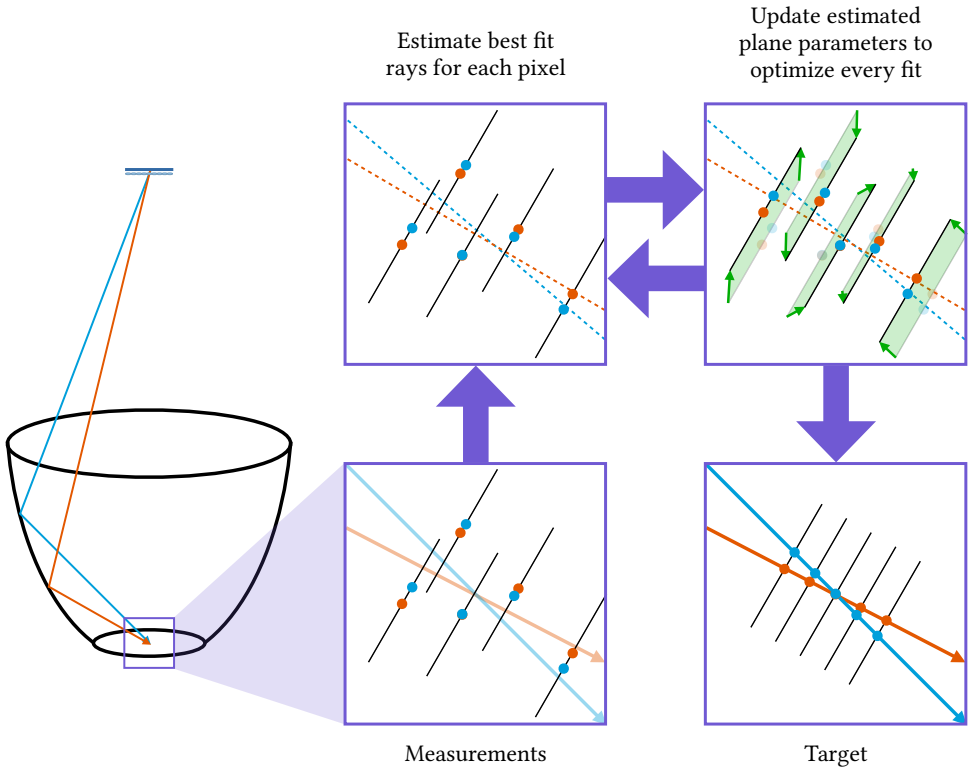

Figure 4.4: **Iterative calibration algorithm.** Illustration of our calibration procedure. Consider the two rays originating from different pixels on the sensor, shown here in blue and orange. To determine the ray parameters in the object-space (at the bottom of the ellipsoid), we need to determine a set of points that fall along the ray. We place a series of planes along each ray, as shown in the target frame in the bottom right. Using a helper camera, we determine the parameters of these planes, and we display graycodes on each plane to allow us to determine the intersection point with each plane based on the sequence of intensities measured by the camera. In an ideal world, this would produce the geometry shown in the target frame. However, due to the small scale of our measurement area, small errors in the computed homography between the helper camera and each plane correspond to significant variance in our calibration measurements, as shown in the bottom left frame. To resolve this, we implement a two stage optimization algorithm. At each iteration, we first find the best fit ray for each pixel's measurements, based on the initial homography and graycode estimates. Then, we update the parameters of each plane to attempt to align the measured points along each ray. After a step is taken to update each plane, we recompute the best fit rays and cycle the optimization again. After sufficient iterations, our measurements more closely align to our expected target.

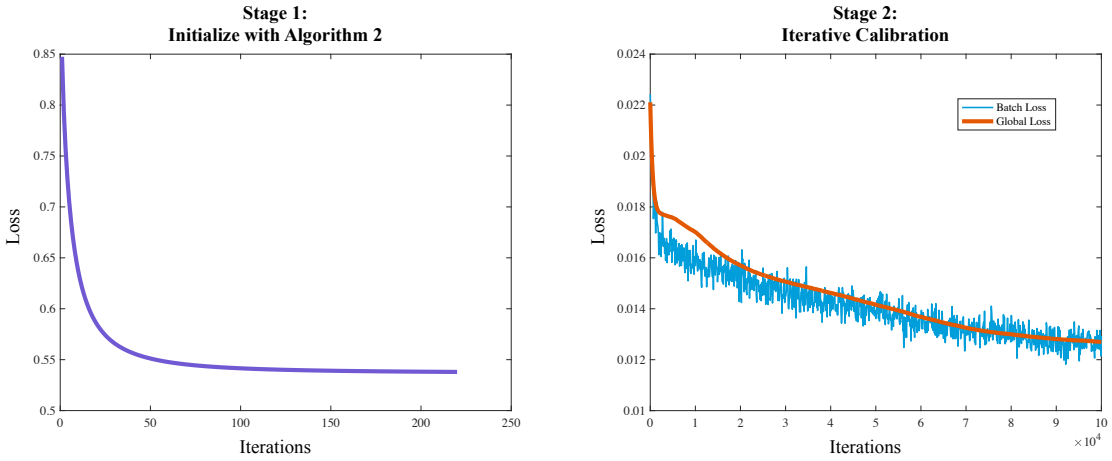

Figure 4.5: **Calibration loss.** We plot the Huber loss for the two stages of iterative calibration. In Stage 1 (left), we initialize the fitted ray origins and directions according to Algorithm 2. This provides strong initial conditions but does not account for the errors in plane parameters. In Stage 2 (right), we alternatively update the best fit rays (per light field pixel) and the plane parameters (across all pixels intersecting a given plane). The batch loss (blue) is computed over the batch of 1000 rays used in each iteration to enable practical use of memory and time. Every 100 iterations, we additionally compute the loss over all light rays, which we plot in orange. We observe that the overall loss effectively reduces along with the batch loss even with the reduced number of rays trained over in each iteration.

We additionally compute the global loss over all light rays every 100 iterations. This produces a smooth loss plot to gauge our optimization progress; we utilize this loss for our convergence testing.

After the iterative calibration procedure has converged, we visualize the impact on the 3D points use to fit each ray. Figure, 4.6 shows the 3D points for one light field ray before and after this calibration. Note that before calibration, we observe significant deviations of these points from the expected ray. These deviations exist primarily along one direction (orthogonal to our view in this case), which aligns with the optical axis of the helper camera. After calibration, the updated plane parameters produce 3D points that fall much more closely to an expected ray. This occurs for all measured light field rays, producing measurements closer to the expected points based on our calibration plane placement. Our iterative calibration procedure adjusts these 3D points accordingly and produces more accurate fitted light rays as a result.

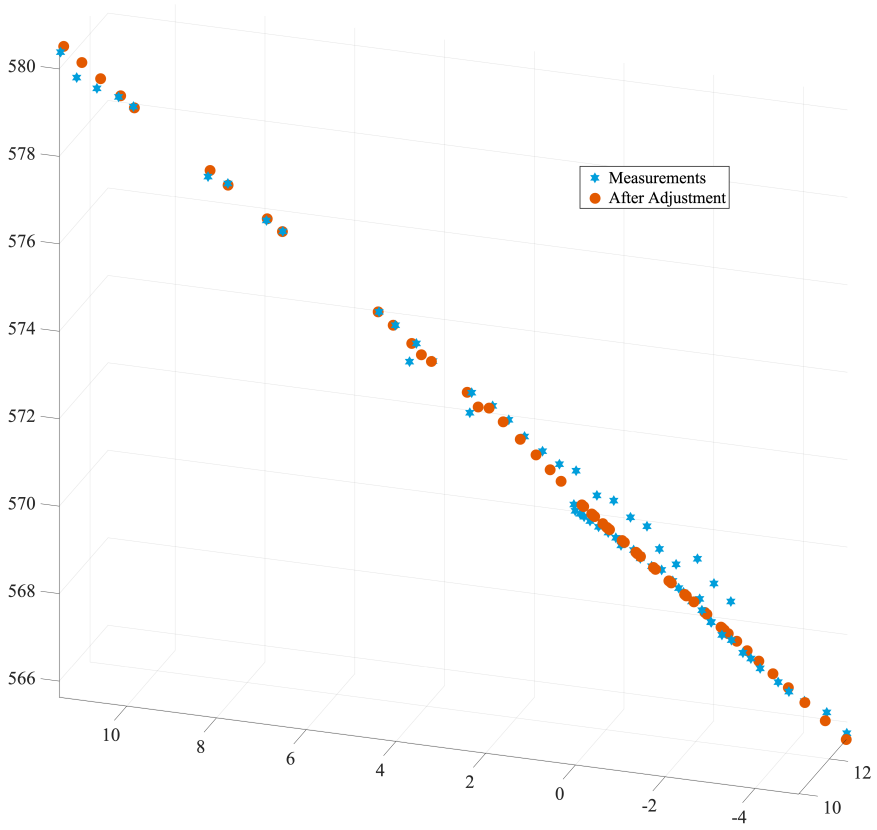

Figure 4.6: **Adjusted 3D points following calibration.** We plot the 3D points corresponding to a single light field ray both before (blue) and after (orange) the calibration procedure. We note that prior to calibration, the 3D points exhibit significant variation along one axis (corresponding to the optical axis of our helper camera). After calibration, the updates to the plane parameters lead to more linear 3D points; in turn, this produces more precise ray fitting to these points.

4.3 Analysis of Prototype Reprojection

In Section 3.2, we considered a theoretical basis for the spatial resolution of our imaging system. Now that we are able to properly calibrate our prototype , we consider the spatial resolution of our camera in terms of the reprojection error of measured light rays. We measure this error by using an LCD placed at the object focal point of our camera, similar to the calibration procedures outlined previously. By displaying a checkerboard on this display and capturing an image from our helper camera, we can determine the 3D location of every pixel on the display along with the parameters of the plane made up by these pixels.

We next illuminate a single pixel on the LCD and capture the light field image created on our sensor.

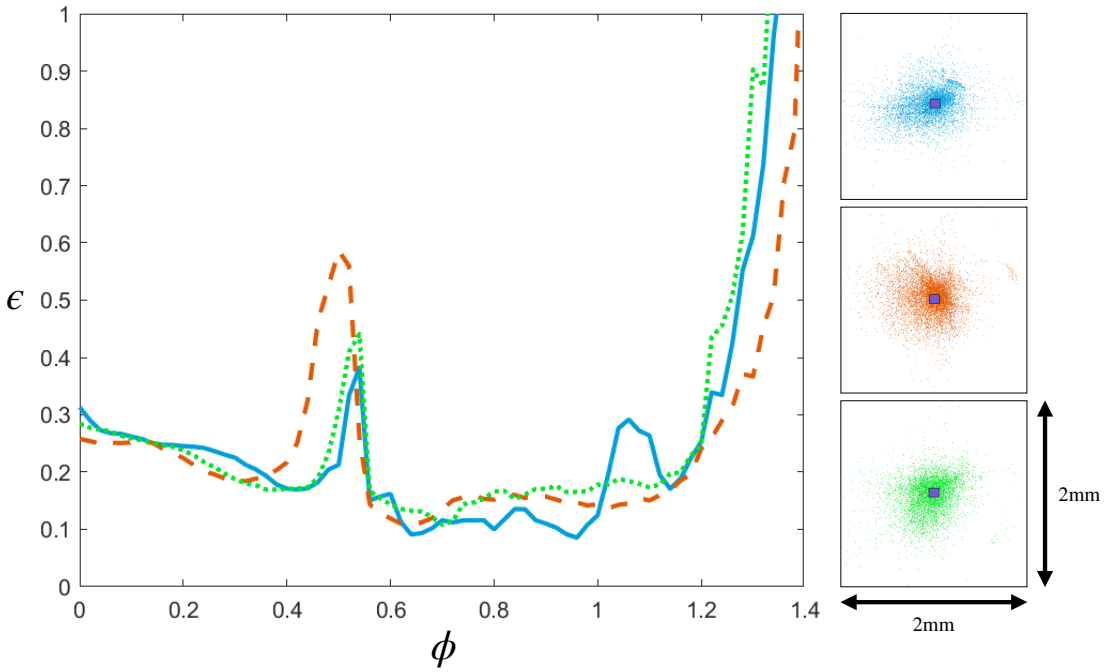

Figure 4.7: **Reprojection errors.** Plot of the reprojection errors for three illuminated points measured with our lab prototype. On the right, we show the spread of reprojected points of a measured pixel in blue, red, and green; the size of the measured pixel is shown in purple. On the left, we show the reprojection error as a function of the ray elevation angle. We observe a dramatic increase in error near grazing angles, which is expected based on prior analysis. Additionally, we note a peak around $\phi = 0.5$ corresponding to the gap in angular coverage between the Fresnel lens and the ellipsoidal mirror.

Due to the low light levels of a single pixel, we use a long exposure and capture images with this pixel both turned on and off. The difference between these two images reveals the pixels on our light field camera that observe the spot formed by the LCD pixel. The light rays corresponding to these light field pixels should correspond to measurements within the small spatial extent of the LCD pixel ($\sim 97.7\,\mu m$). We project these rays back to the calibrated plane formed by the LCD and consider the error between this projected location and the calibrated location of the illuminated pixel.

Figure 4.7 shows this reprojection for three different illuminated pixels near the lower focal point. On the right, we observe the spread of reprojected points over a $2\,mm \times 2\,mm$ area surrounding the pixel, illustrated in purple. We observe a cluster of points directly surrounding the pixel and a number of points that are scattered further from this point. On the left, we plot the distance ϵ between the reprojection and the edge of the pixel with respect to the elevation angle of the measured light ray. We

bin these ϵ measurements into elevation angles of a given resolution to visualize the trends more clearly.

Similar to the analysis in Figure 3.4, we note a significant increase in the error as we approach grazing angles ($\phi \rightarrow \frac{\pi}{2}$). This is due to small angular errors or changes producing significantly changed projection locations when the ray direction is nearly parallel to the projection plane. Additionally, there is a peak around $\phi = 0.5$. This occurs due to the gap in elevation angles measured between the Fresnel lens and the ellipsoidal mirror, measurements in this region are sparse or non-existent. Beyond these grazing regions, where resolution would be better considered for a plane of different orientation, we see errors on the order of 100−300 µm; we find that more than 50% of reprojected rays fall within 150 µm of the measured pixel for all three points. While this is worse than the theoretical analysis, which put our theoretical spatial resolution along the horizontal axis at less than 100 µm, but the added variability in ray measurements, even after calibration, adds to this limitation. We consider the average reprojection error to be comparable to the spatial resolution in this analysis; note that there are additional reprojections that fall beyond this average, but we start to consider these rays in terms of noise rather than resolution. The noise present even in these calibrated measurements encourages robust techniques in our applications explored in Chapter 5-7.

4.4 Illumination Calibration

Both dense and sparse illumination control, introduced in Chapter 3, require the calibration of the input light sources. We aim to produce calibrated illumination directions ℓ for each captured image under a different lighting condition. We introduce the simple procedures required in this section.

4.4.1 Sparse Illumination

The sparse illumination design (Figure 3.14) introduces several directional light sources by collimating the light from several LEDs. We calibrate these light directions by placing a reflective sphere of known diameter at the object plane. We image this sphere through the helper camera, with the first image captured with bright illumination and a longer exposure, as shown in Figure 4.8. This enables the helper camera to observe the outline of the sphere, which in turn gives us the needed information about its 3D geometry relative to the helper camera. We then image the sphere with a short exposure for each light source turned on. This produces a single bright spot in the captured image where the light source reflects off the mirror and reaches the helper camera. Based on the location of this bright spot, we can determine the point on the sphere where the light is reflecting. Through the use of spherical geometry and the known camera ray reaching that point from the helper camera, we determine the ray that reaches

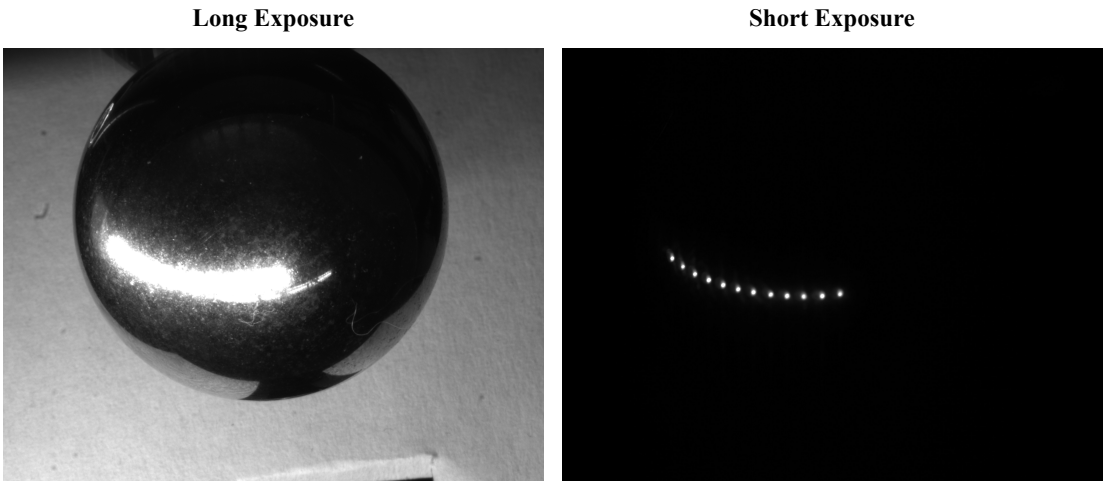

Figure 4.8: **Images for illumination calibration.** For both sparse and dense illumination calibration, we capture both long (left) and short (right) exposure images of a reflective sphere. The long exposure reveals the shape of the entire sphere, which we use along with the calibrated camera parameters to determine the 3D location and size of the sphere. The short exposure reveals the observed reflection of the light sources; in this case, we observe the dense illumination control of the LED strip. We apply the known geometry and observed reflection locations to determine the light ray from the sphere to the light source. With directional light sources (used in our sparse setup), this calibrates the illumination direction. With point light sources (used in our dense setup), we repeat this procedure with different reflective sphere locations to produce a set of rays observing each light source position. We triangulate these rays to produce the calibrated point light source.

the light source from the scene. Since we have directional light sources, this corresponds to a calibrated illumination direction for images captured under that light condition.

4.4.2 Dense Illumination

Since the dense illumination design (Figure 3.15) uses nearby point light sources, we need to calibrate the precise 3D location of each LED rather than just a single ray corresponding to the direction. We perform this calibration by again imaging a reflective sphere of known diameter placed at the object plane with the helper camera, as in Figure 4.8. We follow the same procedure for determining the 3D geometry of the sphere relative to the helper camera. We set the illumination arm to each desired illumination position and image the sphere under one illuminated LED. The observed reflection of each light source through this sphere provides a ray that observes the desired 3D point of the LED. By moving

the illumination arm through all desired positions, we produce one ray pointing to each light source. However, since we aim to calibrate a single point, we need to produce multiple rays that observe each point. So, we repeat the procedure for several different sphere positions, thereby producing a set of rays observing each light source. We find the 3D location of each LED by triangulating these rays to find the intersection point closest to each ray. To convert this illumination point into the illumination direction ℓ for each measured light field pixel, we require some understanding of the scene geometry. This is performed when inputting the calibrated light directions to the network described in Chapter 7; we discuss the details further in that section.

4.5 Discussion

We have shown progressively more robust calibration algorithms, sacrificing runtime for better ray fitting, particularly at grazing angles. These algorithms were developed along with the applications shown in Chapters 5 - 7, meaning that different iterations of these algorithms were applied to the results shown in these chapters. Applying newer calibration schemes to earlier work would improve our results.

5 Geometric Light Field Processing

Light field cameras have been shown to enable a bevy of useful post-capture processing techniques [Ng *et al.*, 2005]. This includes refocusing to different depths in the scene, estimating the geometry of the scene, or generating synthetic images from novel viewpoints. These are achieved through resampling the captured measurements based on the known geometric alignment of the imaging components and the known light rays that contribute to the desired result. Due to this, the achievable results are limited by the limitations of the captured light field itself. First, the narrow cone of light captured by such a camera limits the minimum depth of field for refocusing. Second, the narrow baseline of measurements provides limited information for 3D shape reconstruction, leading to reconstruction only along one axis (depth reconstruction). The narrow baseline also limits the range of novel viewpoints that are achievable from the data. In all, while this data allows for such reconstructions, the limitations are significant.

The measurement of wide-baseline light fields can significantly enhance these conventional uses of such light field signals while enabling a bevy of new capabilities. Like its narrow-baseline counterpart, wide-baseline light fields can be used to refocus the scene at different depths, reconstruct the shape of objects, and generate synthetic images from novel viewpoints. However, the larger angular cone can dramatically enhance these capabilities by providing extremely shallow depth of fields, the ability to image the backside of an opaque object, and an extended range of measurements from which to produce novel viewpoints. Additionally, wide-baseline light fields can provide truly novel capabilities for geometric light field processing. Namely, by using visibility as a physical cue, we can estimate the surface normals of scene points; such a capability, especially in the absence of any active illumination or strong assumptions on the scene illumination, is unique to our setup. We use this capability to jointly estimate the 3D shape and surface normals of objects using our wide-baseline light field camera.

Contributions. This chapter introduces a new methodology for processing wide-baseline light fields. In particular, we make the following key contributions.

- *Shape and normal estimation.* We describe a procedure for shape and normal estimation from wide-baseline light fields for reconstructing high-resolution 3D models while independently estimating the surface normal at each point.
- *Traditional applications.* We explore applications traditional to light fields, namely, changing of the focus and perspective. The wider cones of light that our setup measures allows for some unique capabilities; for example, we can create focus stacks corresponding to planes with surface normal perpendicular to the optical axis of the setup.

Limitations. The proposed system shares limitations from the imaging setup. In particular, the size of the mirror limits our reconstruction area to $\sim 10\,\text{mm}^3$. Additionally, the missing central cone results in non-traditional bokeh as well as reduced performance in normal estimation in some cases. Finally, the proposed shape and normal estimation algorithms use visibility analysis that relies on scene points having a supporting hyperplane; this analysis requires that the shapes under consideration are convex.

5.1 Prior Work

Our work builds upon prior work devoted to light field processing. Many of the light field imaging designs discussed in the previous chapter are used for light field rendering, which resamples the captured light field to produce novel images. In particular, these light fields are commonly applied to refocusing and viewpoint synthesis [Levoy, 2006, Ng, 2005, Ng *et al.*, 2005], though the range of novel viewpoints is limited by the baseline of the light field image capture. Larger baseline light field techniques provide a larger synthetic aperture that excels in focusing through complex occluders [Vaish *et al.*, 2006] at the cost the higher data and equipment requirements of a camera array. Similar light field rendering techniques have been applied to microscopy as well [Levoy *et al.*, 2006, Lin *et al.*, 2015].

Light field imaging is also commonly applied to shape reconstruction. Several geometric algorithms have been introduced to solve this problem, including reconstruction from Epipolar Plane Images (EPIs) [Ziegler *et al.*, 2007], stereo matching from subaperture viewpoints [Heber *et al.*, 2013], and depth from focal stacks [Tao *et al.*, 2013]. We refer to [Wang *et al.*, 2024] for a survey of light field depth reconstruction techniques. Our proposed technique is most similar to the class of focal stack techniques, though we operate within the light field ray space as opposed to considering focus on a rendered image.

Multi-view imaging techniques are well established for reconstructing 3D geometry via multi-view stereo [Seitz *et al.*, 2006]. Light stages similarly provide multiple viewpoints and illumination control for shape and reflectance acquisition [Debevec *et al.*, 2000]. These setups capture high-quality geometry

but require more extensive hardware or multiple image exposures. A comparable single-shot approach to our work is found in kaleidoscopes, used for acquiring light fields [Manakov *et al.*, 2013] as well as 3D shape [Ahn *et al.*, 2021, Lanman *et al.*, 2009, Xu *et al.*, 2018]. Kaleidoscopes can provide multi-view information in a single exposure, but they require the solving of a complex ray labeling problem due to multiple bounces through the mirror configuration. Additionally, all these techniques produce sparse measurements of any given scene point, which can limit viewpoint synthesis reconstruction.

5.2 Overview

We utilize the base wide-baseline light field imaging setup described in Chapter 3. We do not implement the extended measurements via the Fresnel lens or the activation illumination control. The capability of 3D reconstruction under passive illumination is a key result in this work. We show our 3D reconstruction technique in Section 5.3. We additionally provide results for refocusing and viewpoint synthesis, whose much simpler algorithms are described briefly below.

5.2.1 Refocusing

Similar to other light field cameras, we are able to produce refocused images at varying depths. However, our setup can achieve very small depths of field due to the wide angular range of captured data. We follow a simple procedure to generate these images. First, we define a plane at some point and orientation that intersects the object where we wish to produce a focused image. Then, by accumulating the measured rays that intersect pixels along that plane, we can produce an image focused at that plane. While other setups are only able to produce focal sweeps along the direction parallel to the optical axis of the light field camera, our imaging setup allows for producing focused images at arbitrary locations and orientations.

5.2.2 Viewpoint Synthesis

We can also simulate other cameras and imaging setups by resampling the wide-baseline light field data in different ways. We can accumulate light rays that reach a simulated sensor pixel through any optics that we choose to define, including both pinhole and lens-based cameras. Compared to other light field viewpoint synthesis methods, our imaging setup provides a wider range of possible viewpoints but generally reduces the measurements available for any given viewpoint.

5.3 3D Point and Normal Estimation

We now design algorithms for recovering shape, in the form of 3D points and associated surface normals. We approach shape reconstruction from the mindset of scanning the volume, one voxel at a time, and establishing whether or not a voxel lies at (or near) the surface of an object. Our approach for determining this relies on an observation on determining the tangent plane to the scene point from light field measurements.

Observation. Consider the example in Figure 5.1 where we visualize the radiance from a point that is *on* the surface of a textured sphere; this point under consideration has an amber-color diffuse albedo and is mildly specular. The figure shows the radiance over a full sphere of outgoing directions, i.e., in every possible outgoing direction (θ, ϕ), where θ and ϕ are azimuth and elevation angles about the vertical line. This radiance image has two clear segments: a near constant region that corresponds to light that is emitted by the surface, and a textured region that corresponds to light from *other* scene points that occlude the point under consideration when observed along the corresponding direction. *The boundary between the two corresponds to the directions on the tangent plane to the surface at the point under consideration, and maps to a circle on the unit sphere of directions.* When we consider a point that is even slightly off the surface—either above or below—as seen in Figure 5.1 (bottom), this behavior changes completely; *in particular, the radiance map is not crisply split into two regions with differing behavior, nor is the separating boundary between them guaranteed to be a circle.* The observations do require the sufficiently rich textures and fairly mild assumptions on the specularity of the surface. More importantly, this observation provides the basis for identifying between voxels that contain valid scene points and those that do not; intriguingly, they also allow us to estimate the surface normals.

In essence, our approach for 3D point and normal estimation relies on whether or not we can find a valid surface normal at each candidate voxel; failure to find such a normal indicates that there is no valid 3D point in the voxel.

5.3.1 Normal Estimation

To solve for the surface normal $\vec{n}$ at each point, we utilize the visibility clues that we gain from the wide-baseline light field. In general, an outgoing direction $\vec{v}$ from a point will belong to the hemisphere around that point's surface normal if $\vec{v}^\top \vec{n} > 0$. For convex objects, this hemisphere is the set of direction from which the point is visible, which means that the radiance of measured rays from within this hemisphere

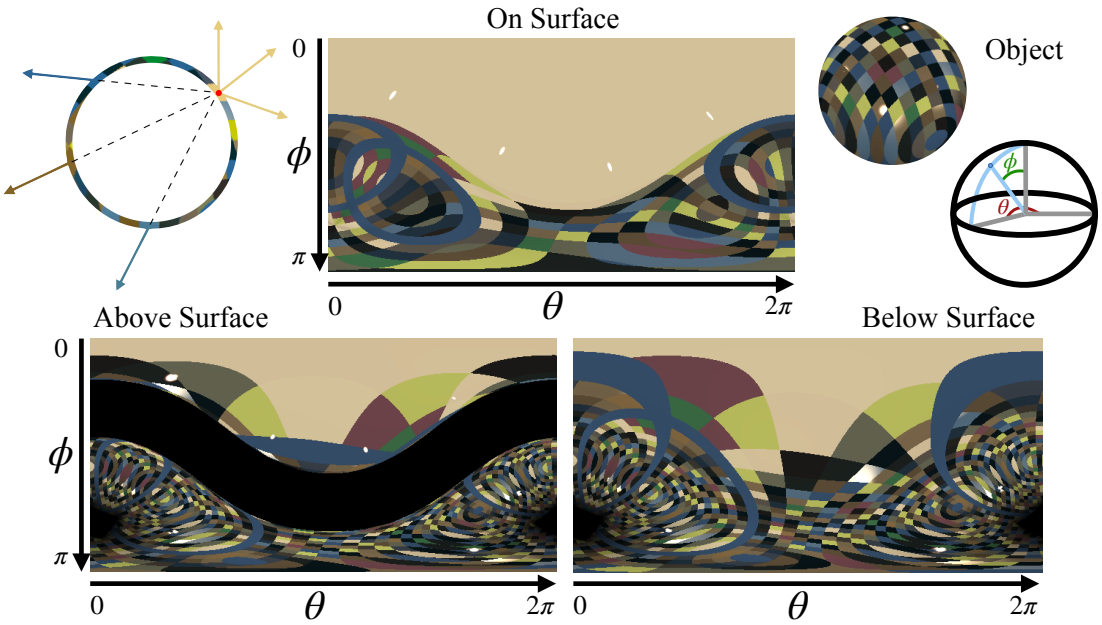

Figure 5.1: **Light field radiance plots.** Plot of the entire light field of radiance measured along each outgoing direction (θ, ϕ) (defined in the upper right) from a single point either on (top), above (bottom left), or below (bottom right) the textured sphere shown in the top right. The ray diagram in the top left shows how different directions from this single point map to different radiance measurements. When the point is on the surface, the set of all measured light rays over the hemisphere above that point will correspond to the reflectance of that point. This results in the consistent reflectance observed in the upper half of the topmost plot. This scene was illuminated using four directional light sources, which appear as the four specular highlights in the plot. Measured light rays from the hemisphere below the surface point will actually be measurements from various points on the other side of the sphere, which results in the varying, checkered reflectance pattern observed in the lower half of the plot. The transition between these two regions corresponds to the visibility horizon. When not on the surface, this transition is not defined and so we cannot estimate a surface normal.

will be measurements of the reflectance at that point. This hemisphere is visualized by the constant region in the upper half of the plot in Figure 5.1.

Conversely, directions $\vec{v}$ such that $\vec{v}^\top \vec{n} < 0$ in the lower region correspond to measured directions at which the surface point is not visible. These directions instead measure the radiance of various points on the other side of the object.

If the object is textured, as it is in this case, we can describe these two regions as "low variance" (for

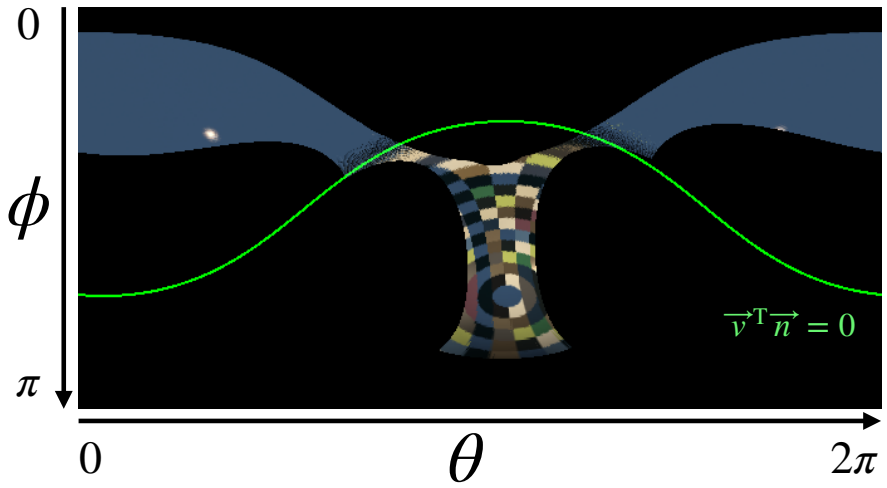

Figure 5.2: **Measured light field and visibility horizon plane.** Plot of the portion of the light field measured by our imaging setup as radiance measured along each outgoing direction (θ, ϕ) from a single point, similar to Figure 5.1. The visibility horizon plane, plotted in green, still adheres to the boundary between high and low variance regions.

the region above the visibility horizon plane) and "high variance" (for the region below the visibility horizon plane). The measurements in the low variance region all correspond to the reflectance at a single point, which has a large amount of consistency across outgoing angles. The measurements in the high variance region correspond to the reflectance at a variety of surface points, which all could have unique reflectance properties. Therefore, we can consider the plane $\vec{v}^{\top}\vec{n} = 0$ as corresponding to the visibility horizon. In Figure 5.1, the visibility horizon plane corresponds to the transition from the low variance region in the upper part of the plot to the high variance region in the lower part of the plot.

With our imaging setup, we are not actually able to measure every outgoing ray from every point in the scene. Instead, for a given point in the scene, the outgoing rays that we can measure would produce a plot like the one in Figure 5.2. The visibility horizon plane, plotted in green, is still observable as the boundary between high and low variance regions. To detect this horizon plane for every point analyzed in the scene, we follow a four-step algorithm, with each step visualized with an example in Figure 5.3.

- First, we color-normalize our measurements, i.e., to ensure that effects of shading are not considered as high variance effects, we normalize every measurement by the sum of its color channels.
- Second, we compute a local variance metric for each measured direction. This metric takes measurements from the neighborhood around a given direction and computes the sum of the variances of the color-normalized red, green, and blue channels.

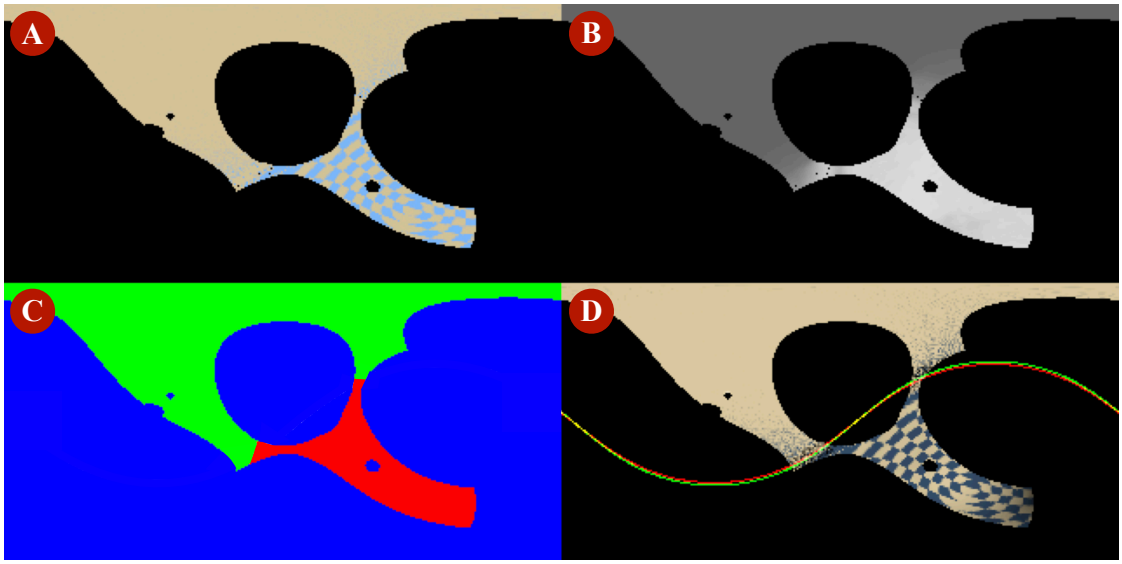

Figure 5.3: **Normal estimation.** Visualization of the four stages of the normal estimation pipeline. Each plot is consistent with the (θ, ϕ) axes defined previously. A: color normalized reflectance plot. B: local variance metric computed for each direction (brighter corresponds to higher variance). C: variance labels (green is low variance, red is high variance). D: visualization of the fitted visibility horizon plane (red) along with ground truth visibility horizon plane (green) over the original reflectance plot.

- Third, we assign a label y for each direction based on a threshold chosen empirically. We set $y = 1$ for low variance points and $y = 0$ for high variance points.

- Fourth, we train a simple classifier to separate the low and high variance regions. The classifier network structure is shown in Figure 5.4. The output of the network is a class probability predictor p, defined as

$$p = \frac{1}{1 + e^{-\vec{v}^\top \vec{n}}}. \tag{5.1}$$

The connectivity of the network ensures that the input to the logistic function is $\vec{v}^\top \vec{n}$, so the function will output a class probability based on whether that dot product is greater than or less than zero, which corresponds to the visibility horizon plane classification discussed previously. The class probability $p \in [0, 1]$ will be closer to 0 when $\vec{v}^\top \vec{n} < 0$ and closer to 1 when $\vec{v}^\top \vec{n} > 0$. We train our network using a standard binary cross-entropy loss function defined as

$$L = -\left(y \log(p) + (1 - y) \log(1 - p)\right), \tag{5.2}$$

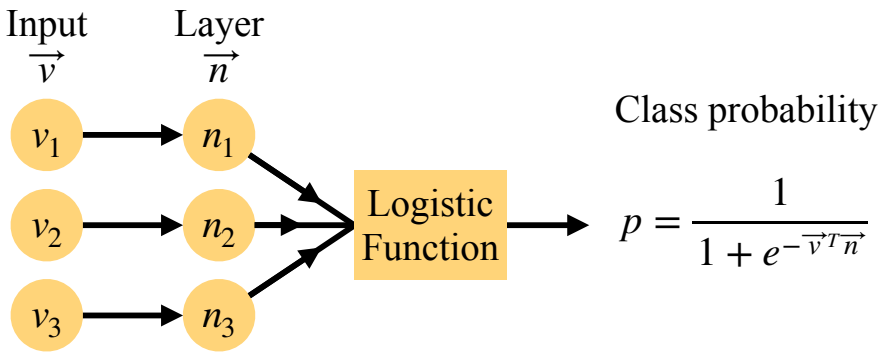

Figure 5.4: **Variance classifier.** Network trained to classify high and low variance regions.

where y is the label produced in the third stage. The weights of the network are the elements of the surface normal estimate $\vec{n}$ once normalized to a unit vector.

This algorithm will estimate a surface normal for any point in the scene. For points on the surface of an object, the visibility horizon plane corresponds to the delineation of high and low variance regions in the reflectance. For points that do not fall on the surface, such a plane does not exist. Therefore, the ability to detect and estimate an effective visibility horizon plane provides depth cues that can be used to generate a point cloud of the object.

5.3.2 3D Shape Reconstruction

We use the surface normal estimation procedure, described earlier, to determine whether or not a point belongs to the surface. Given a 3D point under consideration, we first identify visible light rays from it, using the variance metric. Using just these visible rays, we can compute the total variance of rays above the visibility horizon. If the total variance is below a certain threshold, then we denote that point as belonging on the surface. If the overall variance is above that threshold, we reject the point as not belonging to the surface. Similarly, if the normal estimation algorithm does not find a large enough set of rays that are considered visible, we reject the point as likely falling within the object. In either case, these tests determine whether or not the normal estimation algorithm produced a visibility horizon plane that successful delineates a significant region of low variance reflectance measurements. The variance of the visible region is ultimately a function of a point's distance from the surface. Figure 5.5 shows a histogram of the variance of visible rays for a textured sphere as a function of the distance from the surface. We see that the relationship between distance from the surface and variance holds well, for moving a small distance away from the surface produces a large jump in variance relative to points on or very close to the surface.

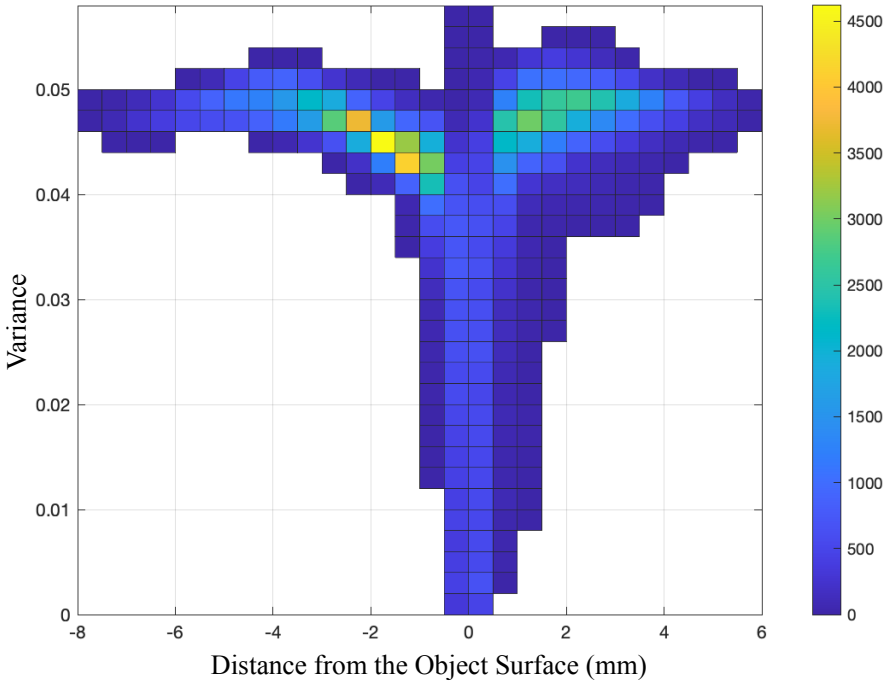

Figure 5.5: **Total variance histogram.** Histogram of the total variance of visible rays as a function of distance from the surface. Moving away from the surface in either direction results in a significant increase in variance, so we set a variance threshold to detect point that are on or very close to the surface.

5.3.3 Volumetric Sweep Speedups

Since our depth and normal estimations are computed point by point, we need to perform a volumetric sweep of our object space. In simulation, our target point cloud resolution is 50μm, while the total volume considered is 10.8mm$\times$10.8mm$\times$10.8mm. If we were to do a naive volumetric sweep, this would require over 10 million points. To speed up our processing, we implement a coarse-to-fine algorithm.

We start with a volumetric sweep at resolution 450μm. This requires the analysis of 13824 points, which is a very reasonable number for our algorithm. We perform point-by-point depth and normal estimation as before, but with a slightly higher total variance threshold. Points that are kept in the point cloud are then used to generate the next level of points at resolution 150μm. For this new level, we only analyze points that are within 150μm of the points kept at the previous resolution. We again perform depth and normal estimation with a reduced variance threshold to generate the next level of points at the final resolution of 50μm.

As an example of how this fills in a densely sampled point cloud, Figure 5.6 shows a this algorithm applied in simulation. Using a textured bunny model as our object, we first scan at 450μm resolution,

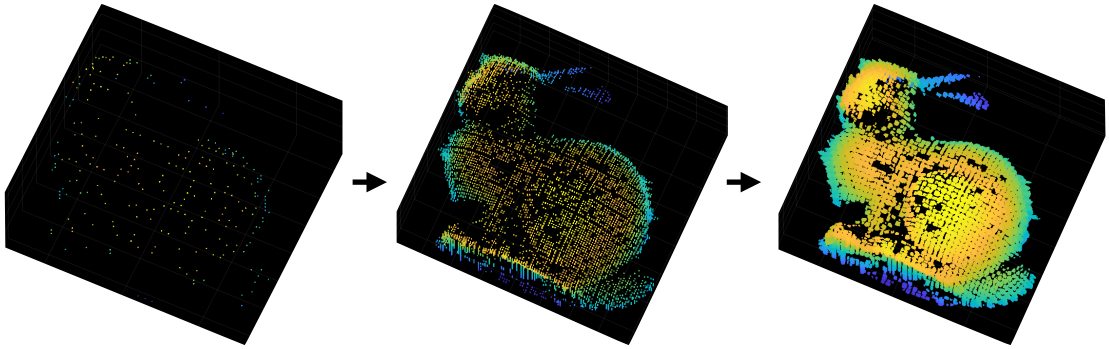

Figure 5.6: **Coarse-to-fine example.** Examples of reconstructed point clouds using the coarse-to-fine algorithm. The point clouds progress in resolution from left to right at 450μm, 150μm, and 50μm respectively.

then 150μm, then finally 50μm. While some gaps remain that were not detected in the first stage, as a whole the algorithm produces a densely sampled point cloud at the desired resolution.

5.3.4 Confidence Filtering

We can improve our point clouds by considering the calculated variance at each point kept as belonging on the surface. The lower the computed variance, the higher the confidence that we can have in that particular point. So, we can clean up our point clouds by removing points that have nearby points with much higher confidence. In particular, our algorithm searches a set of k neighbors around a given point p_i and sample resolution r. For p_i and each point p_j in the neighbor set, we have an associated confidence metric c_i and c_j as well as a distance d_{ij} from p_i to p_j. To filter out point p_i, we search for any point p_j that satisfies:

$$\frac{d_{ij}}{r} \leq 2^{c_j - c_i - 1} \tag{5.3}$$

This removes points that have very close neighbors with a moderate confidence improvement and point with slightly farther neighbors with significant confidence improvement.

5.4 Simulated Results

We evaluate the proposed setup and associated algorithms with simulated measurements.

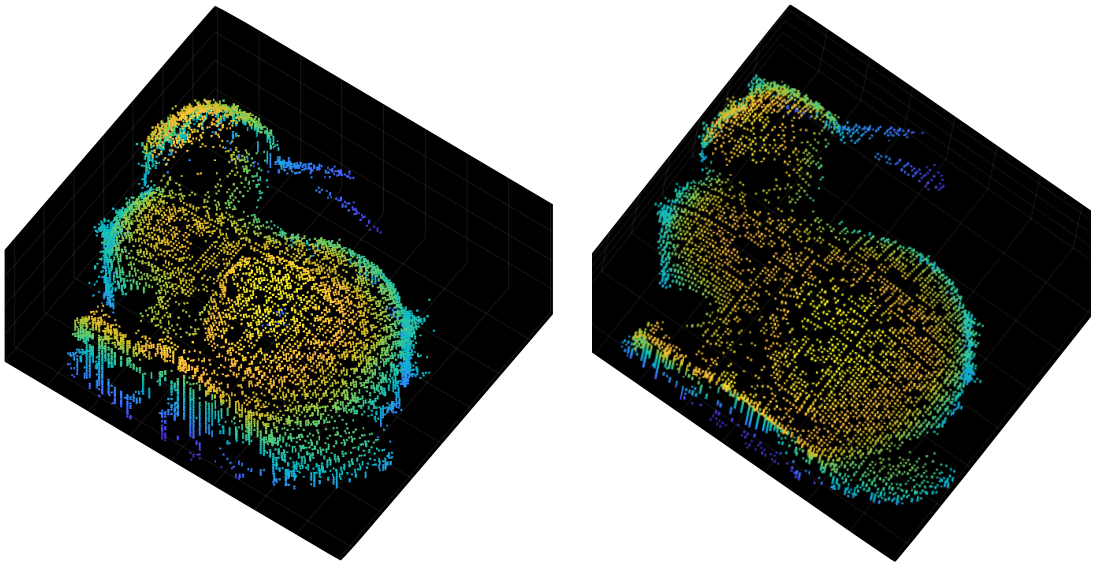

Figure 5.7: **Point clouds.** Comparison of point clouds produced at 150μm resolution for data from a single exposure and a 6×6 grid of exposures. The color map is based on the depth of the points.

5.4.1 Shape Reconstruction

We perform shape estimation using the procedure outlined above. The output of the algorithm is a point cloud where each point has an associated surface normal and computed variance. To improve the density of our measurements of the light field, we capture results from a 6×6 grid of shifted camera positions. This aids the variance computations by providing more data, which helps improve the reconstructions. However, capturing with a single exposure can still produce quality results. Figure 5.7 shows a side-by-side comparison of point clouds generated at 150μm resolution for a single exposure and for the multiple exposure method. The multiple exposures reduce some artifacts and produce slightly better results, but the single exposure still captures the shape qualities of the object.

In both cases, we observe that gaps in the point clouds occur due to missed points in the first phase of the coarse-to-fine volumetric sweep that we perform. These gaps most often occur in areas where the convexity of the object least holds, for that reduces the size of the cone of light that would be considered low variance for points on the surface. Occasionally, these gaps occur in other areas as well, which occurs due to the coarse sampled point that is nearest the surface in that area being very near to a textural transition, which can introduce more variance.

We run our algorithm on four objects: a sphere, bunny, armadillo, and dragon. Due to the wide

angle of measured light, we are able to observe points on the far side of these objects from the camera. However, depending on the shape of the object, the amount of measurements on the backside may be limited. In particular, the bunny model has a large area on the backside with a surface orientation directly away from the camera, so we are unable to acquire sufficient measurements there. For the sake of visualization, we merge point clouds generate from setups configured above and below the object to fill in the full reconstruction, with each view producing a high quality reconstruction of the side of the object facing the camera. The sphere, armadillo, and dragon models are sufficiently measured on the backside to fill out a 3D reconstruction, so these objects are reconstructed from the single view.

Figure 5.19 shows reconstruction results for these four objects captured in simulation. The objects were rendered with high frequency diffuse textures and analyzed down to a point cloud resolution of 50μm. The top three rows of the figure show the ground truth normal map visualized on the objects followed by two views of the estimated normals visualized on the reconstructed surfaces. We observe that while some of the high-frequency information from the ground truth surface normals is not preserved, the overall surface orientations are effectively estimated.

The bottom three rows of the figure show ground truth objects from one view and reconstructed objects from two different views. The reconstructed objects were generated from the point cloud locations and normals using MeshLab. We observe that for relatively simple, mostly convex shapes like the sphere and the bunny, our reconstructions are quite accurate. The method begins to break down as we try to resolve the fine shape details and non-convex areas of the armadillo and dragon, but the overall shape characteristics are reconstructed still.

We consider the accuracy of our directly estimated surface normals compared with those derived from the reconstructed meshes from the point cloud. Figure 5.8 shows the visualization of the surface normals on the reconstructed sphere along with a plot of the surface normal estimation errors in degrees for both methods. We show that our technique is better able to directly reason the surface normals for convex objects like this sphere. This is supported by the histogram shown in Figure 5.9, which plots a histogram of the normal reconstruction error for the sphere. The majority of our directly estimated surface normals have an error less than 1°, while deriving the surface normal from the shape has a much higher distribution of errors. This shows that even when our point cloud reconstruction has errors that introduce shape reconstruction artifacts, our directly estimated surface normals are able to still match the ground truth effectively for convex objects.

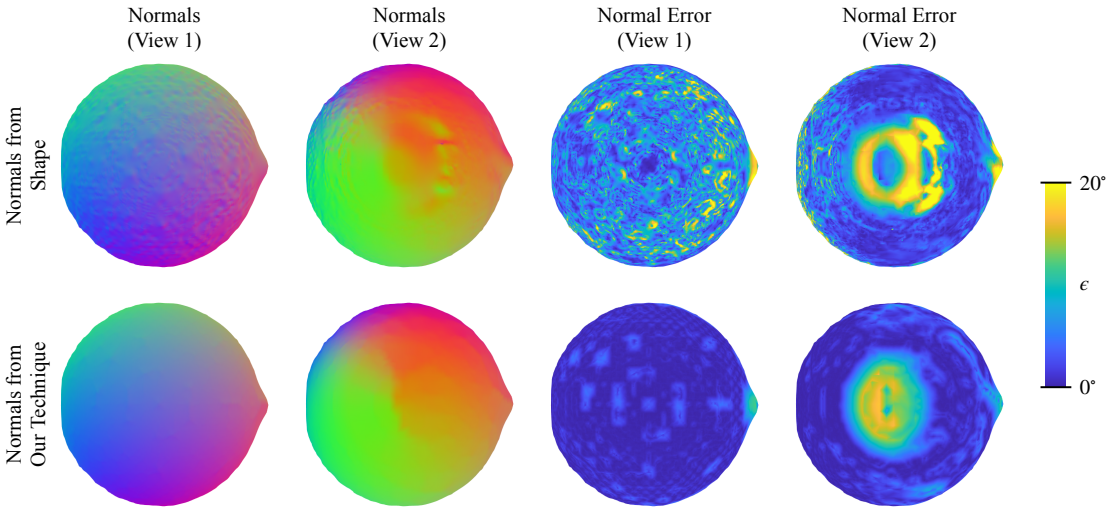

Figure 5.8: **Surface normal error visualization.** We determine the error between our estimated surface normals and the ground truth surface normals for the sphere (bottom). We compare these errors to the surface normals derived from the 3D meshes produced from our point cloud (top). The columns denote different views of the sphere for a visualization of the surface normals (left) and a visualization of the normal estimation error (right).

5.4.2 Refocusing

The wide baseline of measured light enables us to generate focal stacks along different directions. Figure 5.10 shows three sweep directions for a two-color diffuse textured sphere. Here, our setup allows for views around the backside of the object from the camera. In the sweep from the side, we have more measurements in the region toward the camera (on the top of the images), but we still focus on the texture at the bottom of the sphere.

5.4.3 Viewpoint Synthesis

In addition to refocusing, we can resample our light field to generate synthetic views of our objects. However, the sparse spatial measurements of our light field in a single exposure hampers our ability to produce high quality pinhole views. Similar to shape reconstruction though, we can capture a set of data with shifted camera positions to fill in these sparse measurements. Figure 5.11 shows the results of four different views generated for different objects with different reflectance properties. Note in particular the captured variations of the specular highlights of the different objects. The mixed textures are a random assortment of diffuse and specular materials based on those captured from the MERL dataset

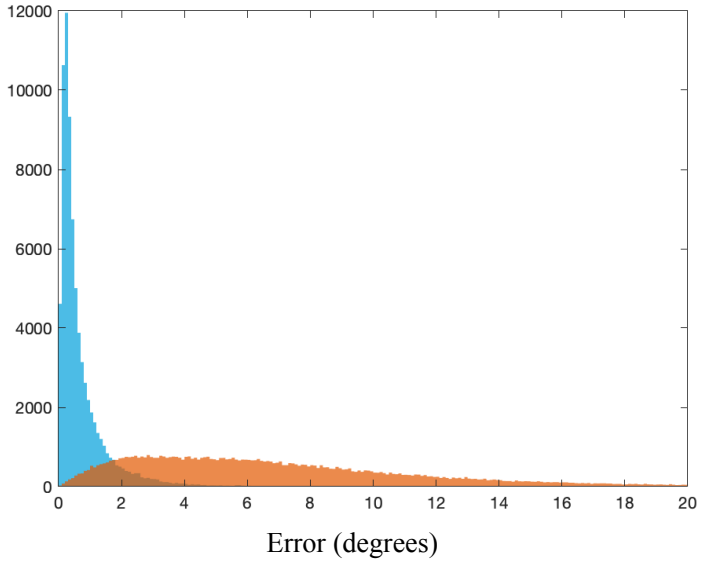

Figure 5.9: **Surface normal error histogram.** We determine the error between our estimated surface normals and the ground truth surface normals for the sphere. We plot a histogram of these errors over the entire mesh for both estimation techniques: direct estimation (blue) and derived from 3D mesh (orange). We show that our technique is significantly better when reconstructing convex geometry, though our surface normal estimation in particular relies on this convexity assumption.

[Matusik *et al.*, 2003]. In addition, the bottom row shows a sphere modeled with an iridescent thin film coating. The different synthesized viewpoints observe a wide selection of specular angles that cause the specular highlight to shift in color according to the model.

5.4.4 Real Results

With the lab prototype, we can capture single-exposure images of wide-baseline light fields. We use this data to perform the same set of processing that we showed in simulation.

Shape Reconstruction. The blur that persists in our real imaging system prevents the processing of more complex objects. However, Figure 5.12 shows a proof of concept of this processing technique on a planar textured surface. Using the joint depth and normal estimation, we produce the point cloud shown in the figure.

Refocusing. We generate focal stacks with a small depth of field in the same way that we did for the simulated results. Figure 5.13 shows results for two objects, all generated with a vertical focal sweep

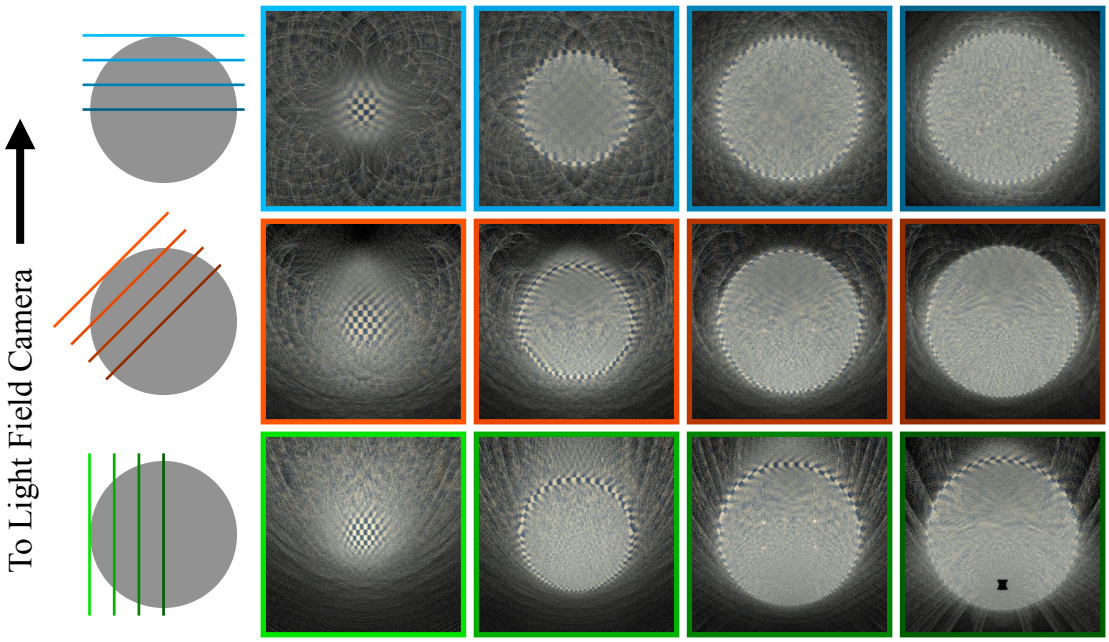

Figure 5.10: **Refocusing.** Simulated refocusing of the sphere model with a diffuse texture at different depths. Each row corresponds to a different focal sweep direction as visualized by the graphics on the left. All directions are with respect to the light field camera direction, which is labeled. The hourglass artifact in the last row is a consequence of the central subaperture view being aligned with the ellipsoid focal point; in this case, all rays from this subaperture map exactly to the other focal point, while neighboring sub-apertures map to points some distance away.

from under the object surface up toward the camera. The first object is a ThorLabs resolution chart printed on glass. The first two rows of the figure show data captured in two different areas of the chart. For both locations, we observe two apparent focal planes: first in column (b), then in column (d). While column (b) seems less in focus than column (d), we note that it is more in focus than the focal plane between the two in column (c). This dual focusing is caused by the shadow cast from the printed pattern through the glass to the plane on which the glass is resting. Column (b) is the depth focused on this shadow, which is inherently less sharp than the original pattern and is shifted slightly in one direction according to the direction of the light source. Column (d) is focused on the pattern printed on the glass, which is sharpest. On either side of these focused planes, the focal sweep quickly defocuses. The final object shown is a blue die. We see that we come clearly into focus in column (d), but fall into defocus blur in either direction from that point.

As we showed in simulation, we can also perform focal sweeps in different directions as well. Figure

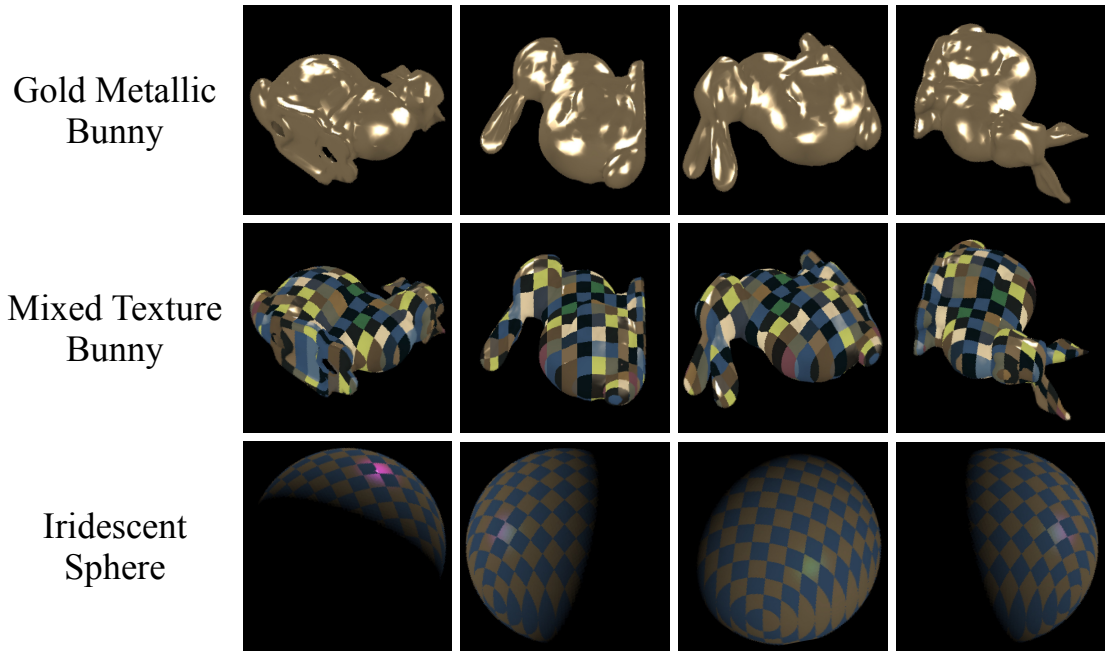

Figure 5.11: **Viewpoint synthesis.** Simulated viewpoint synthesis of the sphere and bunny models with different textures and reflectances. The gold and mixed texture bunnies were illuminated with four directional light sources. The sphere model was illuminated with a single light source. The iridescent sphere was modeled using a thin film reflectance model, which produces angle-dependent changes in reflected wavelengths in the specular highlights.

5.14 shows two focal sweep directions for a stack of two small dice, particularly focused on the top die where we have more measurements. Sweeping from inside the die up through the top toward the camera, we come into focus in the fourth frame. The second frame is focused on two of the dots on the side of the die, so the in focus area resolves to a small strip from this view. The sweep from the side of the die brings a different face of the cube into focus and effectively illustrates the narrow depth of field. Since we are not sweeping perfectly square to the side of the cube, one part of the face comes into focus before the other, as shown in the third and fourth frames. This is true for the first two frames as well, where different parts of the row of three dots on top come into focus in each frame.

Viewpoint Synthesis. In our simulations, we improved the quality of viewpoint synthesis results by capturing a set of images with a shifted camera. This produced high-quality pinhole images over a wide range of viewpoints. With our physical setup, we capture just a single image with a stationary camera. This does not provide the dense measurements needed for these sharp synthetic views. Instead, we

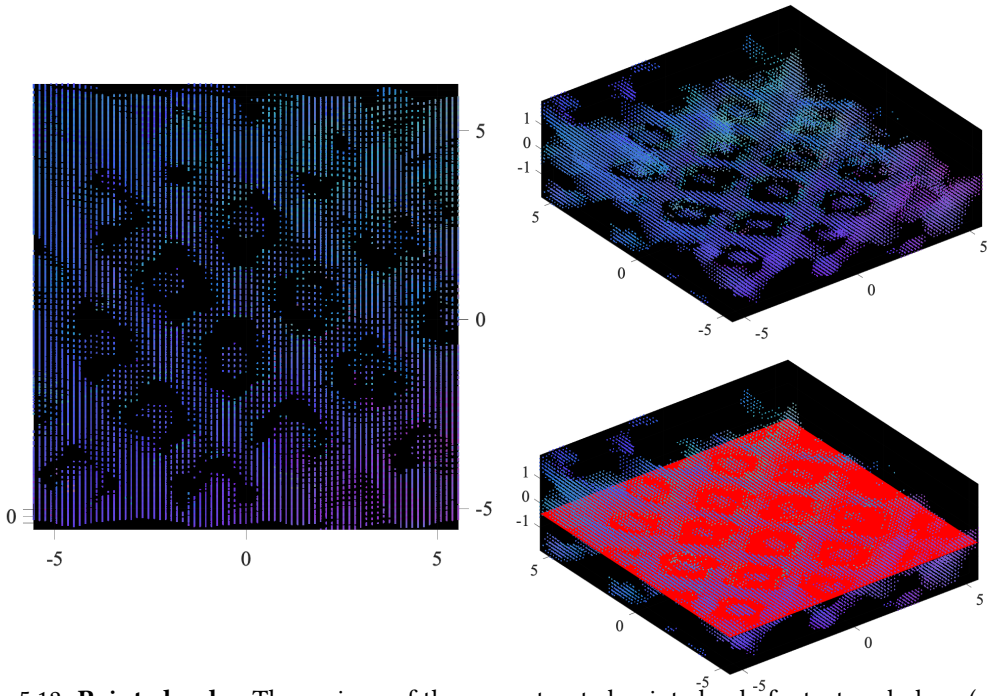

Figure 5.12: **Point cloud.** Three views of the reconstructed point cloud of a textured plane (a bayer pattern printed on paper) with computed surface normals. These images show how the blur in the system allows proper measurements mostly just in the center of each texture element, as the measurement gaps follow a regular grid. The estimated surface normals are all fairly consistent with the proper orientation of the plane, but they do tend to be skewed toward whichever end of the volume the point is on. In the bottom right, we visualize the point cloud along with a fitted plane.

simulate a lens-based synthetic camera and apply interpolation to the rays we have measured to fill in the missing gaps. This produces blurrier images than what we showed in simulation, but we are able to produce these results from a single exposure. Figure 5.15 shows results for four different objects. In particular, note how the different views capture the different angular responses to the single light source. The specular highlight in the third column for blue die and the different illuminated or shadowed side faces on the green illustrate this behavior effectively.

5.5 Discussion

We introduced a novel imaging device that senses wide-baseline light fields through the mapping properties of an ellipsoidal mirror. We showed how this data enhances the capabilities of standard light field processing techniques by achieving extremely small depth of fields and increasing the area of the object

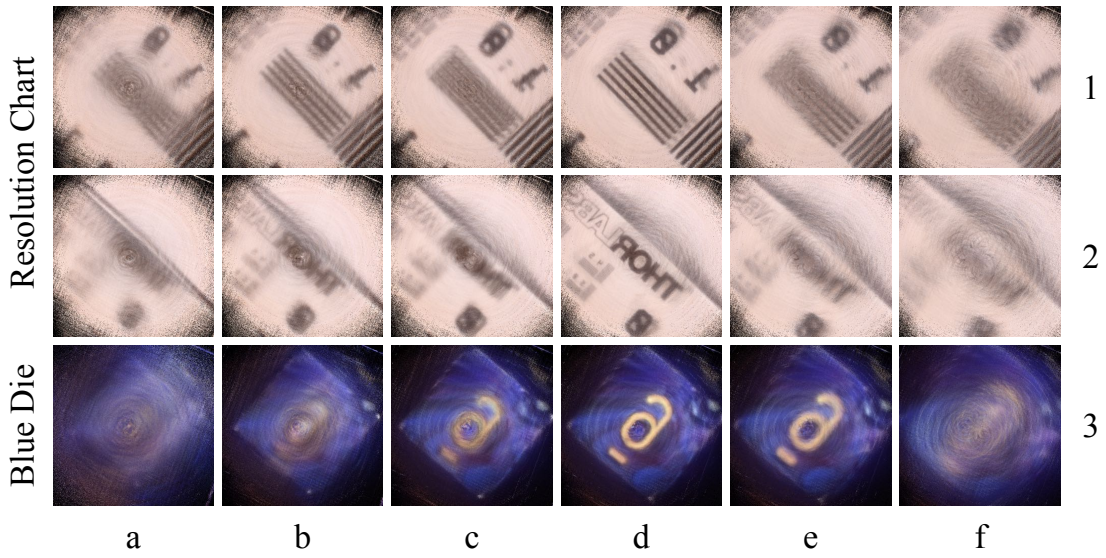

Figure 5.13: **Refocusing.** Focal sweeps of real data captured of a ThorLabs resolution chart and a blue die. Depths are swept from lower to higher in the direction of the light field camera.

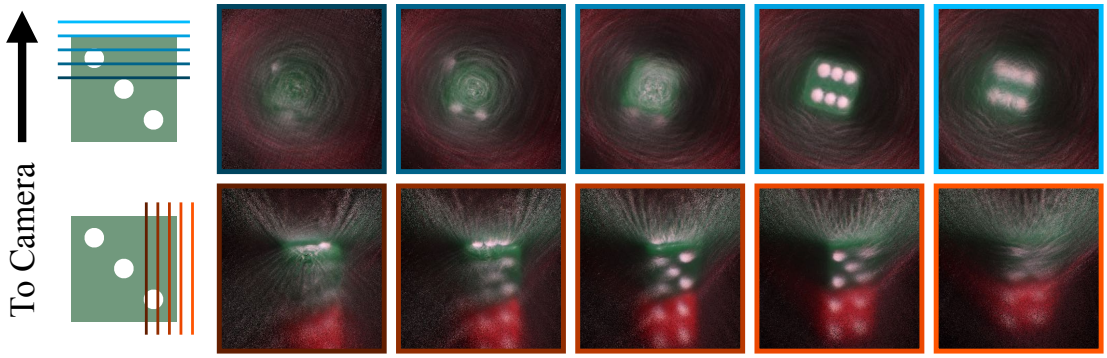

Figure 5.14: **Refocusing from alternate directions.** Focal sweeps in vertical (top) and horizontal (bottom) directions, as illustrated on the left. We are able to resolve different faces on the die through the different directions.

that we can observe. Wide-baseline light fields also open the door to a new set of capabilities. By using visibility cues, we detailed a procedure for estimating surface normals of scene points and integrated this ability into a shape reconstruction algorithm. We also demonstrated the ability to capture high-frequency reflectance phenomena like iridescence. In total, our proposed camera design pushes into new territory for light field acquisition and analysis.

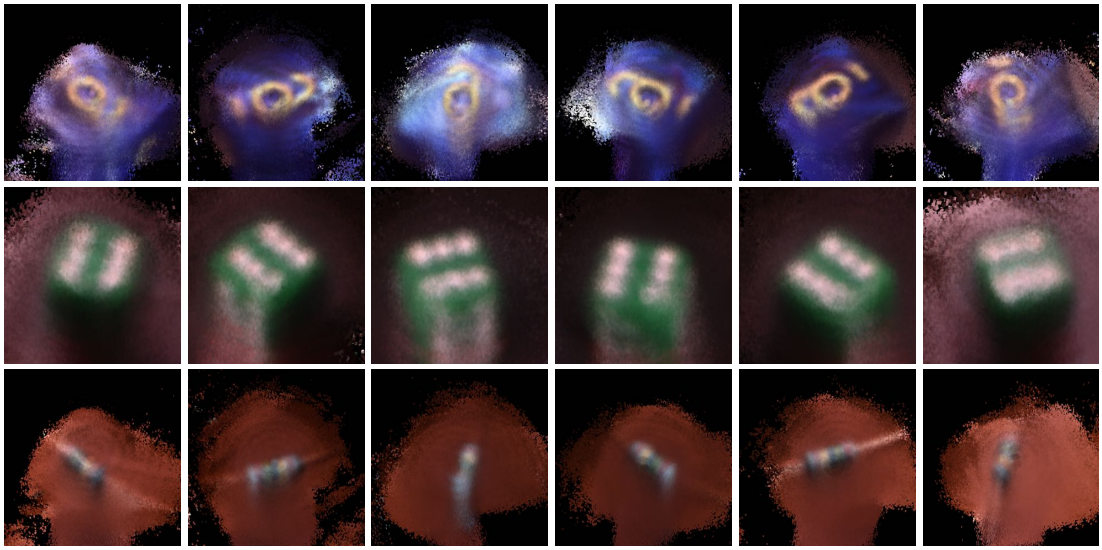

Figure 5.15: **Viewpoint synthesis.** Viewpoint synthesis for a blue die (top row), a small green die (middle row), and a blue resistor (bottom row). Each view reflects the angular changes in reflected light based on a single illuminant.

Assumptions on object shape . A key highlight of wide-baseline light fields is their ability to resolve surface normals from a visibility analysis. This relies on the existence of a supporting hyperplane at a scene point, such that there are no self-occlusions in the cone of light emitted on one side of the hyperplane. This property is satisfied by a convex object. In practice, while this technique does work on non-convex shapes as seen in Figure 5.19, the reconstruction is worse in regions that are locally non-convex. We explore this in simulation by using a 3D bowl geometry, as outline in Figure 5.16. We define a set of different bowl geometries by intersecting spheres of different radii with the planar side of a cube, enforcing a fixed intersection diameter of 6 mm. We place these different bowls within our simulated imaging setup and aim to reconstruct the bowl using the techniques defined in this chapter.

Figure 5.17 shows the results of this experiment. We show the reconstruction profiles for seven different bowl depths, from a flat plane to a fully hemispherical bowl. We observe that we are able to relax our convexity assumption to an extent and still reconstruct the bowls with accuracy, as noted on the left side of the figure. We note average errors less than or equal to 64 µm when the depth of the bowl is less than or equal to 1.2 mm. Deeper than this, we note average errors in the hundreds of µm and a complete failure to reconstruct the deepest of bowls. This aligns with our empirical observations, where we can account for some non-convexity, but significantly non-convex objects can cause difficulties for our reconstruction.

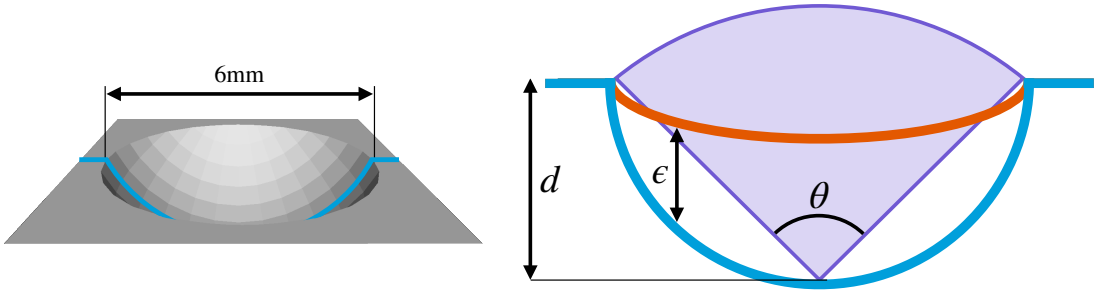

Figure 5.16: **Convexity experiment setup.** We consider the impact of convexity on our reconstruction results by capturing light field images of a spherical 3D bowl. We create the bowl model by intersecting spheres of increasing diameters with a plane, enforcing a fixed intersection diameter of 6 mm. This forms a set of bowls with different depths d. We further consider the bowls in terms of the upper hemisphere coverage by considering the percentage of elevation angles that can be observed for a point at the bottom of the bowl. We compute this metric as $p = \frac{\theta}{\pi} \times 100\%$. When reconstructing the bowl, we consider the depth error ϵ for each point in the bowl.

Effect of the missing cone of light. As highlighted earlier, light field capture using ellipsoidal mirrors alone is incapable of measuring the cone of light centered around the major axis of the ellipsoid. This has important consequences both in shape estimation as well as image-based rendering.

Our shape estimation techniques rely on the large angular span of the measurements, as opposed to completeness of the measurements in that range. This large angular span allows us to obtain measurements on either side of the visibility horizon plane—which is critical for accurate determination of the surface normal as well as depth. This is visualized in Figure 5.2, which shows successful normal recovery from observing a small portion of the light directly on either side of the visibility horizon plane.

The missing cone also results in some artifacts (seen in Figures 5.14 and 6.3) as the rendered images have unusual bokeh, which we simulate in Figure 5.18 for focus stacks with different orientations. The upper portion of the plot shows the rendered bokeh for just the basic wide-baseline light field camera used for the results in this chapter. We also show results in the lower portion corresponding to the variation on the WBLF camera design where an additional lens placed within the ellipsoidal mirror expanded the angular range of the captured light field, as discussed in Chapter 3. We note that when we fill in part of the central cone, the bokeh is more filled in, producing more standard disc bokeh when refocusing along the optical axis. We still observe an unusual crescent-shaped bokeh when focusing along a tilted plane, for there is a portion of the hemisphere over these points that falls below the hemisphere measured by our camera.

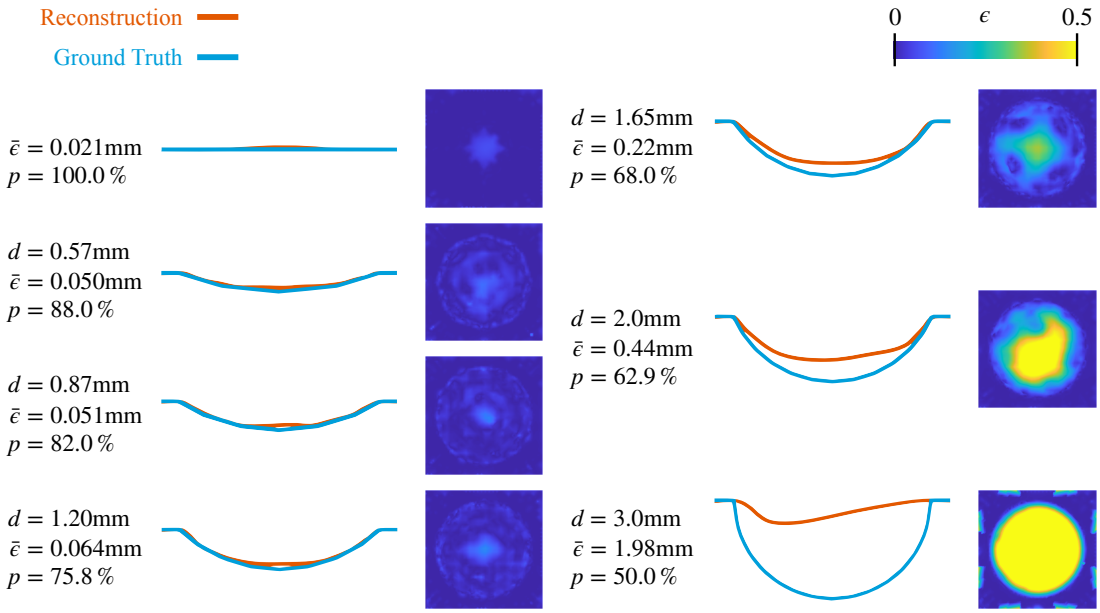

Figure 5.17: **Convexity experiment results.** For seven bowls of different depths and hemisphere coverage percentage, we reconstruct the bowl geometry using the techniques outlined in this chapter. We visualize the central profile of the ground truth and reconstructed bowls for each depth along with a visualization of the depth reconstruction error ϵ for every point of the bowl, visualized looking down at the bowl from above. We further display the average error $\bar{\epsilon}$ over the central profile for each depth. We observe that our technique is able to relax the convexity assumption somewhat (up to bowls of depth $d = 1.2$mm), but the more we relax this assumption beyond this point the worse the reconstruction quality (right side). Note that the bowl depth d and coverage percentage p are defined in Figure 5.16.

The missing cone also restricts the view points where we can render perspective images. However, this is still less restrictive than what's possible with a standard light field camera, where the viewpoints are limited to the very narrow measured cone.

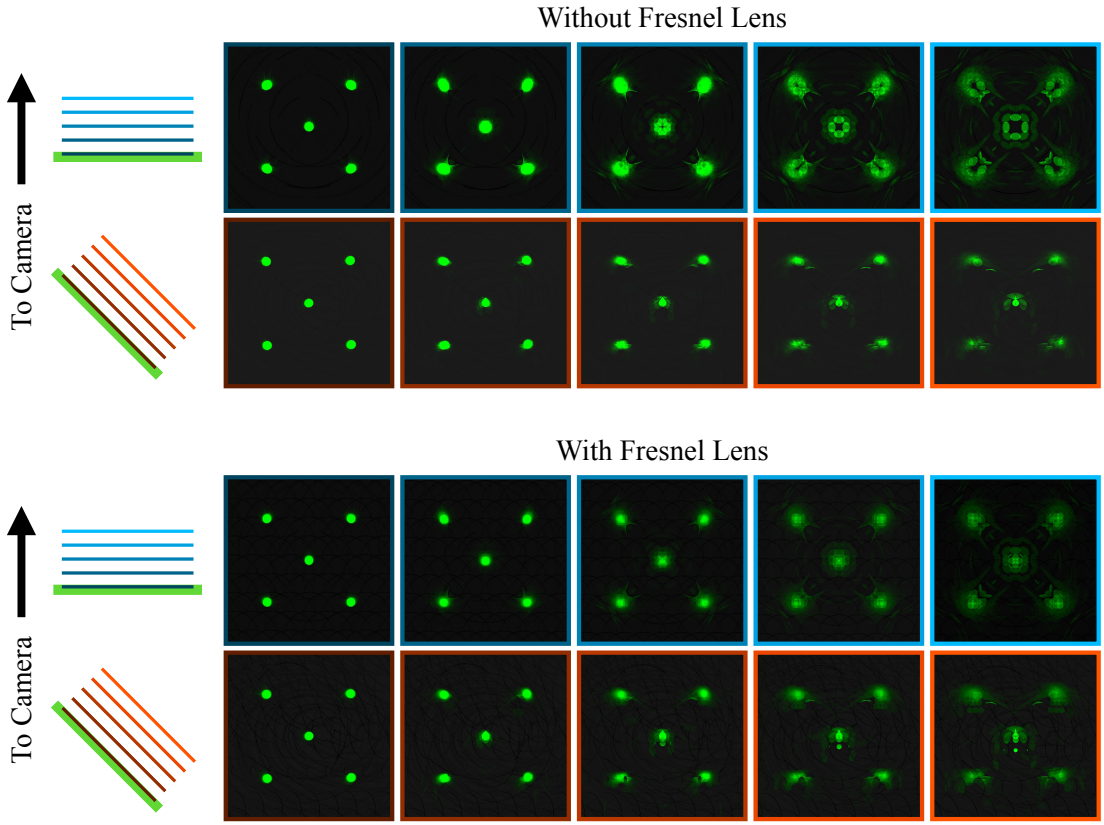

Figure 5.18: **Bokeh visualization.** Focus stacks for a plane near the focal point with five green dots (gamma-corrected). The defocus of these dots reveals the unconventional bokeh resulting from our setup and measurements. This varies depending on the orientation relative to the mirror, as each row designates. When straight on, we observe a ring-shaped bokeh when not using the additional Fresnel lens, but the ring is filled in to a disc when adding the lens. When considering a tilted plane, we observe a crescent-shaped bokeh with and without the lens, though the lens does fill in some missing components of the bokeh.

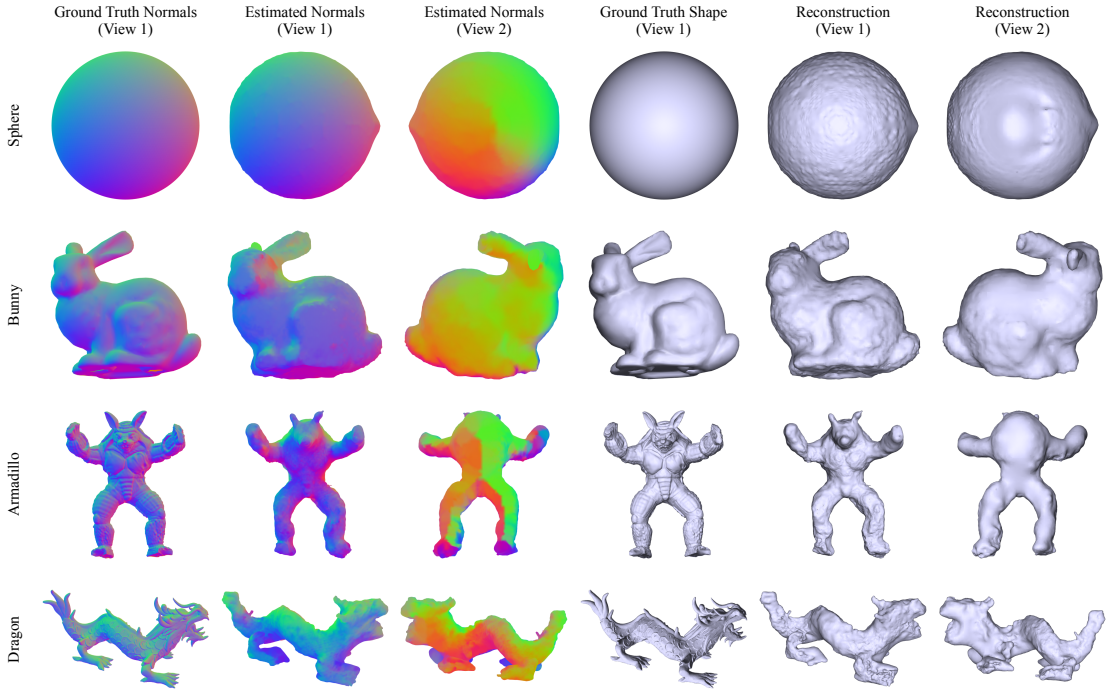

Figure 5.19: **Simulated 3D shape reconstructions.** Shape reconstructions for the sphere, bunny, armadillo, and dragon models. Using the point clouds with per-point surface normals, we use MeshLab to generate reconstructed surfaces. To visualize the original surface normal estimates, we reassign the surface normals for each vertex in the reconstructed mesh to equal a weighted sum of the normal estimates for the ten nearest points in the original point cloud. The left column visualizes the ground truth surface normals. The second and third columns show two views of each reconstructed object with the estimated surface normals visualized on the surface. The fourth column shows ground truth renderings of the different models. The final two columns show two views of rendered images for the reconstructed meshes. We also compute the average reconstruction error of the original estimated point cloud for each of the objects: sphere - 0.116mm; bunny - 0.584mm; armadillo - 0.125mm; dragon - 0.093mm. Each measured object fits snugly within a 1 cm^3 volume, which means our error range is approximately 1-5%.

6 Neural 3D Reconstruction from Wide-Baseline Light Field Data

In the previous chapter, we discussed the geometric light field processing equivalents for wide-baseline light fields. This provides a useful baseline for comparison to conventional light field cameras, which can produce limited shape reconstruction, refocusing, and viewpoint synthesis. However, the geometric 3D reconstruction technique is limited by key assumptions. Namely, it relies on generally convex objects and the presence of high-frequency textural elements. This limits the applicability to general 3D object scanning, where we do not wish to limit ourselves to these objects. Additionally, our reconstruction algorithm has to be applied independently at each point, which is time consuming as well as incognizant of the surface-specific properties of the object.

In addition, refocusing and viewpoint synthesis rely on interpolating between measured rays. For conventional light field cameras that have a grid structure to the measurements, this is well established, as discussed in Chapter 2. For our setup, this is feasible only in simulation, where the exact mirror geometry and configuration is known. In this case, it is a simple matter to trace the rays back to the sensor and perform pixel-based interpolation. However, moving to our optical prototype, we no longer have any knowledge or confidence in the geometry of the mirror or its configuration. This means that the final calibrated output from our camera is an unstructured list of light rays with the corresponding measurements for each. Interpolating in this space is difficult, and it is unclear which measure of distance is proper. For example, do we consider two rays close in angle or two rays close in space to be more similar? This challenge leads to the artifacts in the viewpoint synthesis results for our previous geometric processing technique shown in Figure 5.15.

Both of these limitations lead to the conclusion that we need a representation of our data that allows for shape reconstruction and easier interpolation. This aligns well with a great volume of recent research on neural scene representations [Mildenhall *et al.*, 2020, Tewari *et al.*, 2022, Yariv *et al.*, 2020], which have become a popular method for representing the appearance and geometry of arbitrary scenes. However, most of the research in this space focuses on the model and its outputs, relying on the same input

methodology: a collection of 2D image captures with viewpoints that are sparsely distributed around the object. While this method is effective at capturing the spatial extent of the scene, each image provides just a single additional sample from a scene point. In order to fully understand how light interacts with the scene, we must take measurements of the reflectance at each point more densely; this requires a dense collection of images whose acquisition is invariably time and data intensive.

In this chapter, we propose a shape and reflectance recovery procedure from images acquired from our wide-baseline light field camera. Inspired by recent work on implicit neural representations, we regularize the inverse problem of shape and reflectance recovery using an implicit differentiable renderer (IDR) [Yariv *et al.*, 2020]. We utilize the addition of a sparse collection of directional light sources to the imaging setup to increase the richness of the data capture through illumination control, as discussion in Chapter 3. The use of implicit functions for shape allows us to benefit from the robustness of surface smoothness constraints and avoid per-pixel reasoning. This relaxes the need for high contrast and frequency textures; for example, we are able to successfully scan objects with piecewise-homogenous materials.

Contributions. This chapter presents a novel application for neural scene representation to model 3D scenes using WBLF data captured using an ellipsoidal mirror and a light field camera. In particular, we make the following contributions:

- *Implicit Differentiable Renderer (IDR) from WBLF.* We build a model based on IDR that operates on wide-baseline light field data as opposed to sparse, 2D image inputs.
- *Full Reflectance Modeling.* Wide-baseline light field data provides dense measurements of the reflectance from each scene point. This allows for a more accurate renderer based on a learned BRDF model. In particular, by introducing calibrated illumination and an improved input representation to the rendering network, we are able to render not only synthetic viewpoints but synthetic illumination conditions as well.
- *High Frequency Reflectance.* Existing neural scene representations struggle to represent high-frequency visual effects, including strong specularities. Our method utilizes the rich angular data from the wide-baseline light field camera to improve the modeling of these high-frequency effects.

We characterize the performance of our approach using a set of simulated and real results. These results show significant improvements over prior work, and lays the foundation of fast reflectance acquisition of complex visual phenomena.

Limitations. The limitations of this work are primarily driven by the size of the ellipsoidal mirror used, which dictates the size of the objects that can be measured—this limit is $\sim 10\,\text{mm}^3$ in our setup. This leaves us at a different scale than existing methods, making direct comparisons difficult. Additionally, we apply sparse illumination to our imaging system; this limits the expressiveness of illumination synthesis and causes limitations for objects with strong self-occlusions. These objects may have regions in shadow for all or most illumination conditions, which negatively impacts the reconstruction quality.

6.1 Prior Work

This chapter builds upon recent advances in implicit neural representations [Sitzmann, [n.d.]] for (inverse) rendering of shape and reflectance [Tewari *et al.*, 2022]. Our work broadly uses the framework proposed in IDR [Yariv *et al.*, 2020], which uses signed distance function (SDF) to (neurally) represent the shape and a texture network to capture the viewpoint dependent appearance of the object. Our approach, due to its reliance on active lighting, requires additional modifications. Zhang *et al.* [2021] allow for explicit illumination modeling so as to decompose "appearance" more broadly into its physically-meaningful constituents, namely, lighting and BRDF; however, the underlying representation of the scene is still via volumetric radiance fields. Yariv *et al.* [2021] and Wang *et al.* [2021] provide a hybrid volume-surface representation to address this gap between shape and volumetric radiance, with the eventual goal of a principled handling of reflectance and illumination. Sang *et al.* [2023] provide a framework for shape, illumination and reflectance using SDFs. Our approach broadly borrows ideas from these, tailoring them to the specifics of our imaging system.

6.2 Overview of Implicit Differentiable Rendering

We implement the scene representation based on the optimization techniques of IDR, which we will outline in brief here. With IDR, we represent the scene using two neural networks, F and M, which respectively define the scene's geometry and appearance. The F network, an MLP with weights θ, defines the scene geometry through an SDF, where the surface of the scene corresponds to the zero level-set of the SDF. This network takes as input a 3D location $\mathbf{x}$ and returns the value of the SDF along with a feature vector $\mathbf{z}$ used for rendering:

$$F(\mathbf{x};\theta) = (f(\mathbf{x};\theta), \mathbf{z}(\mathbf{x};\theta)) \tag{6.1}$$

The M network, an MLP with weights γ, approximates the surface light field radiance L for points in the scene based on their location $\widehat{\mathbf{x}}$, surface normal $\widehat{\mathbf{n}}$, direction of observation $\mathbf{v}$, and feature vector $\mathbf{z}$:

$$L(\widehat{\mathbf{x}}, \mathbf{v}) = M(\widehat{\mathbf{x}}, \widehat{\mathbf{n}}, \mathbf{v}, \mathbf{z}(\mathbf{x}; \theta); \gamma) \tag{6.2}$$

Performing a forward pass through these networks to render a given ray $r = \{\mathbf{c} + t\mathbf{v} \mid t \geq 0\}$ first intersects the surface level-set through the sphere-tracing algorithm [Hart, 1996]. This finds the intersection point $\widehat{\mathbf{x}}$, and the gradient of the SDF produces the surface normal $\widehat{\mathbf{n}}$. These vectors, along with the feature vector $\mathbf{z}$ and the input view direction $\mathbf{v}$, are used as inputs to the texture network M, which produces the RGB value for that ray.

These networks are trained using a set of input images $\{\mathbf{I}\}$ and corresponding masks $\{\mathbf{S}\}$. Rays from the cameras producing these images are traced through the SDF and intersected with the surface level-set. Whether or not an intersection is found produces an estimated set of masks $\{\widehat{\mathbf{S}}(\theta)\}$, and when an intersection is found ($\widehat{\mathbf{S}}(\theta) = 1$) the texture network M is used to produce an estimated RGB value $L_p(\theta, \gamma)$ for a given pixel p. To ensure differentiability, an approximation of $\widehat{\mathbf{S}}(\theta)$ is used:

$$\widehat{\mathbf{S}}_\alpha(\theta) = \text{sigmoid}\left(-\alpha \min_{t \geq 0} f(\mathbf{c} + t\mathbf{v}; \theta)\right) \tag{6.3}$$

The network is trained end to end using three loss components, with the overall loss computed as the weighted sum of these terms.

$$\text{loss}(\theta, \gamma) = \text{loss}_{rgb}(\theta, \gamma) + \lambda_{mask}\text{loss}_{mask}(\theta) + \lambda_{eik}\text{loss}_{eik}(\theta) \tag{6.4}$$

The first loss term is the RGB loss between the input image pixels and the rendered RGB values from the texture network. This operates over the set of pixels $p \in P_{in}$, where P_{in} is the set of pixels where both the ground truth mask $\mathbf{S}$ and the estimated mask $\widehat{\mathbf{S}}$ indicate an intersection.

$$\text{loss}_{rgb}(\theta, \gamma) = \frac{1}{|P|} \sum_{p \in P_{in}} \left|\mathbf{I}_p - L_p(\theta, \gamma)\right| \tag{6.5}$$

The second loss term is the mask loss, which computes the cross entropy loss between the ground truth and estimated masks over the remainder of the pixels $P_{out} = P - P_{in}$.

$$\text{loss}_{mask}(\theta) = \frac{1}{\alpha|P|} \sum_{p \in P_{out}} \text{CE}\left(\mathbf{S}_p, \widehat{\mathbf{S}}_{\alpha,p}(\theta)\right) \tag{6.6}$$

The final loss term is the eikonal loss, which guides the formation of the SDF by enforcing the gradient of the SDF to equal 1.

$$\text{loss}_{eik}(\theta) = \mathbb{E}_x\left(\|\nabla_x f(\widehat{\mathbf{x}}; \theta)\| - 1\right)^2 \tag{6.7}$$

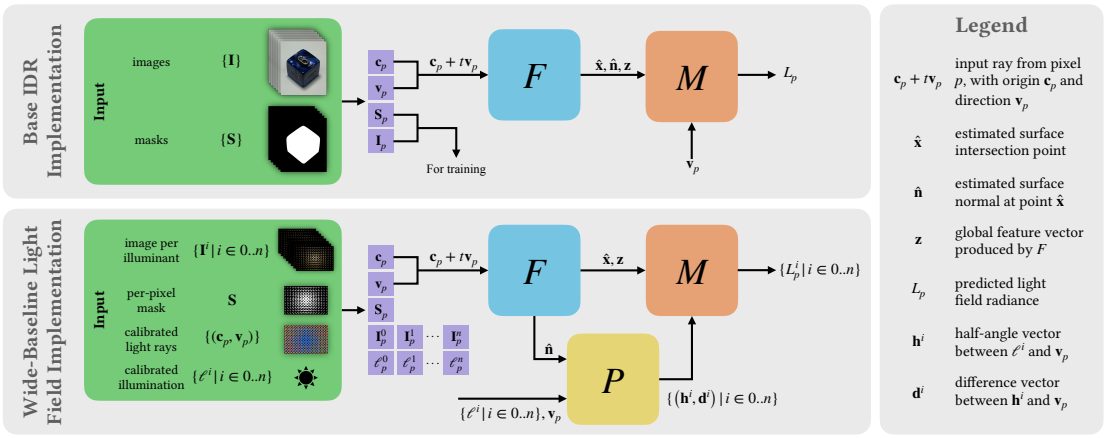

Figure 6.1: **Network architecture.** Network design for the base IDR implementation (top) and our proposed implementation (bottom). The F and M blocks shown are the geometry and texture networks from IDR, described previously. Here, we include the sphere-tracing in the F block so that it outputs $\widehat{\mathbf{n}}$ and $\widehat{\mathbf{x}}$ directly. As opposed to ray tracing from given images and masks, we input from a set of per-pixel calibrated rays and mask values. Additionally, for each input ray, we obtain a set of n images corresponding to n illumination directions. The corresponding RGB values and calibrated illumination directions are included as a set with each individual ray (shown in purple). As such, a single forward pass produces a rendered RGB vector for each of these inputs per ray. Additionally, we add a parameterization block P that converts the set of illumination directions $\{\ell^i | i \in 0..n\}$ and viewing direction $\mathbf{v}$ into a set of half-angle parameterizations for each illumination, set in the coordinate frame of the normal $\widehat{\mathbf{n}}$.

We can utilize the resulting networks to compute a mesh based on the zero level-set of the SDF as well as to render synthetic viewpoints by intersecting the desired rays with the SDF and passing the resulting parameters to the texture network.

6.3 Proposed Method

In this section, we introduce key modifications to the IDR algorithm to enable shape and reflectance estimation from WBLFs. We introduce a sampling framework for our input as well as calibrated illumination, which requires changes to the IDR's network design. Figure 6.1 provides an overview of the proposed technique.

6.3.1 Network Inputs and Sampling

The existing IDR implementation samples a batch of light rays at each iteration by sampling pixels from each 2D input image and tracing the ray through that pixel based on the camera pose and intrinsics. In our method, each pixel on the light field camera has a unique ray path based on how it reflects off of the ellipsoidal mirror below. Rather than trying to represent that geometry directly, we provide each pixel with its corresponding object-space ray through calibration. The network then samples directly from a list of measured rays, each of which also carries the corresponding observed RGB value of the pixel along with a foreground/background indicator, which serves the purpose of the masks provided previously.

Additionally, when using calibrated illumination, each ray is associated with the illumination direction under which the pixel measurement was taken. We note that since our imaging optics are fixed, further measurements with different illumination conditions correspond to the same rays as previous measurements. This has positive and negative effects. The primary downside is that we do not gain any data diversity or expansion despite the additional exposures. To the positive, training can be made efficient, with each ray being trained with multiple rendered radiances at once. Thus, when performing a forward pass through the geometry network, we train using just the set of unique rays and their mask indicators. When passing through the texture network, we train over all of the observations for the sampled rays at once. These inputs to the network are depicted in the first two sections of Figure 6.1.

6.3.2 Modeling the Spatially-Varying BRDF

The texture network used in IDR approximates a function defining the light field radiance $L(\widehat{\mathbf{x}}, \mathbf{v})$. This is defined after integration is performed in the standard rendering equation, with $\mathbf{w}_o = -\mathbf{v}$:

$$L(\widehat{\mathbf{x}}, \mathbf{w}_o) = L^e(\widehat{\mathbf{x}}, \mathbf{w}_o) + \int_\Omega B(\widehat{\mathbf{x}}, \widehat{\mathbf{n}}, \mathbf{w}_o, \mathbf{w}_i) L^i(\widehat{\mathbf{x}}, \mathbf{w}_i)(\widehat{\mathbf{n}} \cdot \mathbf{w}_i) d\mathbf{w}_i, \tag{6.8}$$

where $B(\widehat{\mathbf{x}}, \widehat{\mathbf{n}}, \mathbf{w}_o, \mathbf{w}_i)$ is the spatially-varying BRDF, and L^i and L^e are the incoming and emitted radiance, respectively. Training the MLP to approximate this equation does well to reproduce the appearance of the scene and even transfer that appearance to other geometries. However, this bakes the incoming illumination into the function, rather than treating it as an input. This creates limitations when rendering, particularly when applied to object scanning, where the user may desire to render this object in novel illumination conditions. To do this requires the network to directly approximate the spatially-varying BRDF $B(\widehat{\mathbf{x}}, \widehat{\mathbf{n}}, \mathbf{w}_o, \mathbf{w}_i)$, which in turn requires knowledge of the illumination $\mathbf{w}_i$. Our optical configuration lends itself well to adding calibrated light sources.

To further improve our ability to model the spatially-varying BRDF, we use the half-angle reparameterization of Rusinkiewicz [1998], which is known to improve the reconstruction of high-frequency reflectance functions. Given the direction to the light source ℓ and the viewing direction $\mathbf{v}$, we compute the half-angle vector $\mathbf{h}$ and difference vector $\mathbf{d}$ as follows. First, we define a coordinate system on a given point using its surface normal $\widehat{\mathbf{n}}$, a surface tangent $\widehat{\mathbf{t}}$, and the surface bi-normal $\widehat{\mathbf{b}} = (\widehat{\mathbf{n}} \times \widehat{\mathbf{t}})$. Next, we transform ℓ and $\mathbf{v}$ into this coordinate system, producing the incoming and outgoing rays $\mathbf{w}_o$ and $\mathbf{w}_i$, respectively. Finally, we compute the half-angle vector $\mathbf{h}$ and the difference vector $\mathbf{d}$:

$$\mathbf{h} = \frac{\mathbf{w}_o + \mathbf{w}_i}{||\mathbf{w}_o + \mathbf{w}_i||}, \quad \mathbf{d} = \text{rot}_{\widehat{\mathbf{b}}, -\theta_h} \text{rot}_{\widehat{\mathbf{n}}, -\phi_h} \mathbf{w}_i, \tag{6.9}$$

where (θ_h, ϕ_h) are the spherical coordinates of the half-angle vector in the $\widehat{\mathbf{n}} - \widehat{\mathbf{t}} - \widehat{\mathbf{b}}$ coordinate system. The spherical coordinates of the difference vector (θ_d, ϕ_d) define the difference in angle between $\mathbf{h}$ and $\mathbf{w}_i$ in the same coordinates. The rotation functions shown represent the rotation by the given angle ($-\theta_h$ or $-\phi_h$) around the given axis ($\widehat{\mathbf{b}}$ or $\widehat{\mathbf{n}}$).

In place of passing $\widehat{\mathbf{n}}$ and $\mathbf{v}$ to the rendering network M, we instead pass $\mathbf{h}$ and $\mathbf{d}$. This both enables improved high-frequency reflectance performance and training over calibrated illumination.

6.4 Results

We now evaluate the performance of our technique, in both simulations and real captures. Specifically, we look at reconstructions of shape as well as appearance, both in novel viewpoints as well as illumination.

6.4.1 Simulation

A primary benefit of this technique over that of Chapter 5 is the efficiency of data usage. There, we capture additional data by shifting the light field sensor over the range of the lenslet size (diameter of 2.2mm). Our simulated results for geometric 3D shape reconstruction are generated from a 6×6 grid of shifted views, requiring 36 exposures overall, while viewpoint synthesis required 529 exposures from a 23×23 grid of shifted views. These additional views provided a densely sampled light field at the cost of higher data and time requirements. This technique requires no such shifting of the light field camera, making it akin to a single-shot technique (we apply additional exposures to capture illumination information, which is not helpful to the prior technique).

Additionally, our previous shape reconstruction algorithm is a depth-from-focus technique, which requires textural information in the scene to provide the necessary depth cues. This technique does not assume anything about the texture or reflectance found in the scene.

These two key distinctions make our neural technique important for scanning general scenes, where high-frequency texture is not typically present and fewer exposures are desired. Therefore, we compare the results with this use case in mind; our scenes contain isolated objects with relatively few materials, and we perform 3D scene reconstructions based on the set of eight illuminated images captured as discussed.

Figure 6.5 shows a results comparison of 3D reconstructions from this simulated data. The scale of these objects is $\sim 10\,\text{mm}$ across. Our technique produces effective 3D geometry for each object, with some loss of fine details but the overall structure preserved. The mean error between our reconstructions and ground truth is $\sim 70\,\mu\text{m}$, while the previous technique did not well preserve all parts of the geometry and had higher average errors ~ 100 to $500\,\mu\text{m}$.

Additionally, Figure, 6.6 shows multiple renderings of synthetic viewpoints or synthetic illumination conditions. Our reconstruction is able to properly reproduce specular highlights on the bunny and armadillo upper body while still reconstructing the diffuse color as well. Some very high frequency reflectance, such as that on the legs of the armadillo model, is not able to be reproduced, leaving the material appearing diffuse. Our geometric techniques do not scale down as well to single-shot data effectively, leaving blurry and choppy viewpoint renderings; additionally, we did not have any modeling of specific illumination, so rendering with novel light sources was not possible.

6.4.2 Real Data

Our lab prototype was built with three calibrated light sources, so we capture three images for each of the depicted results. Figure 6.2 shows a results on measurements from our lab prototype. These results show that we are able to produce 3D shape reconstructions for real objects using the neural scene representation in a way that our previous technique could not. One limitation present in these reconstructions is the spatial extent of our measurements. Note that the pizza cutter object, while quite small, is still too big for our setup and leads to portions of the reconstruction being cut off. Additionally, the three calibrated light sources are all mounted above the mirror; the lack of illumination from beneath the object constrains the resistor reconstruction to the upper half of the cylindrical shape.

Additionally, we show viewpoint and illumination synthesis in Figure 6.3. Here, we see the materials of the pizza cutter being well reconstructed; being generally upward facing, each visible scene point is

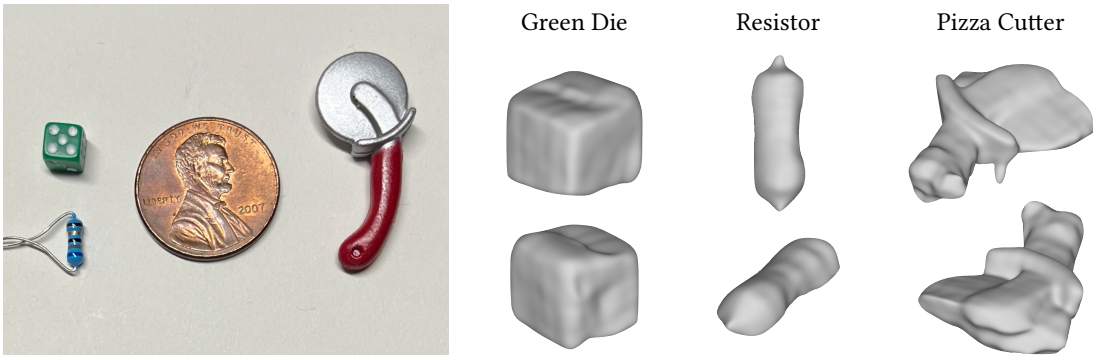

Figure 6.2: **3D reconstruction results.** 3D shape reconstruction results for three objects scanned using our wide-baseline light field camera prototype. The penny is shown for scale.

illuminated by multiple of the light sources, which tends to improve the results. The resistor is more in shadow, which is more difficult to reconstruct completely. Further, practical constraints on the frequencies that we can model limit the full reconstruction of finer details, like the colored bands on the resistor, but they are recovered here with some blur.

Finally, we show a comparison to the viewpoint synthesis presented in Chapter 5 for the green die in Figure 6.4. While neither technique fully reconstructs high quality appearance, our technique constrains the renderings by tracing to the surface representation, which leads to a sharp silhouette of the object. Prior results interpolated directly from the captured rays and were correspondingly blurred over the entire image.

6.5 Discussion

We introduced a neural scene representation capable of operating on wide-baseline light field data to reconstruct 3D shapes and render novel images under different viewing and illumination conditions. This work improves upon the capabilities presented in prior wide-baseline light field research, particularly in severing the reliance on high-frequency textures and providing illumination calibration for further expressivity. We applied calibrated illumination to the task of modeling the reflectance of scene points through a learned BRDF model, which showcased effective modeling of several specular surfaces. As we built upon the baseline established for wide-baseline light field processing, we explored how the rich angular data being captured provides modeling capabilities with fewer exposures yet equally diverse measurements.

Future work in this space can build upon the limitations present in the results shown. In particular,

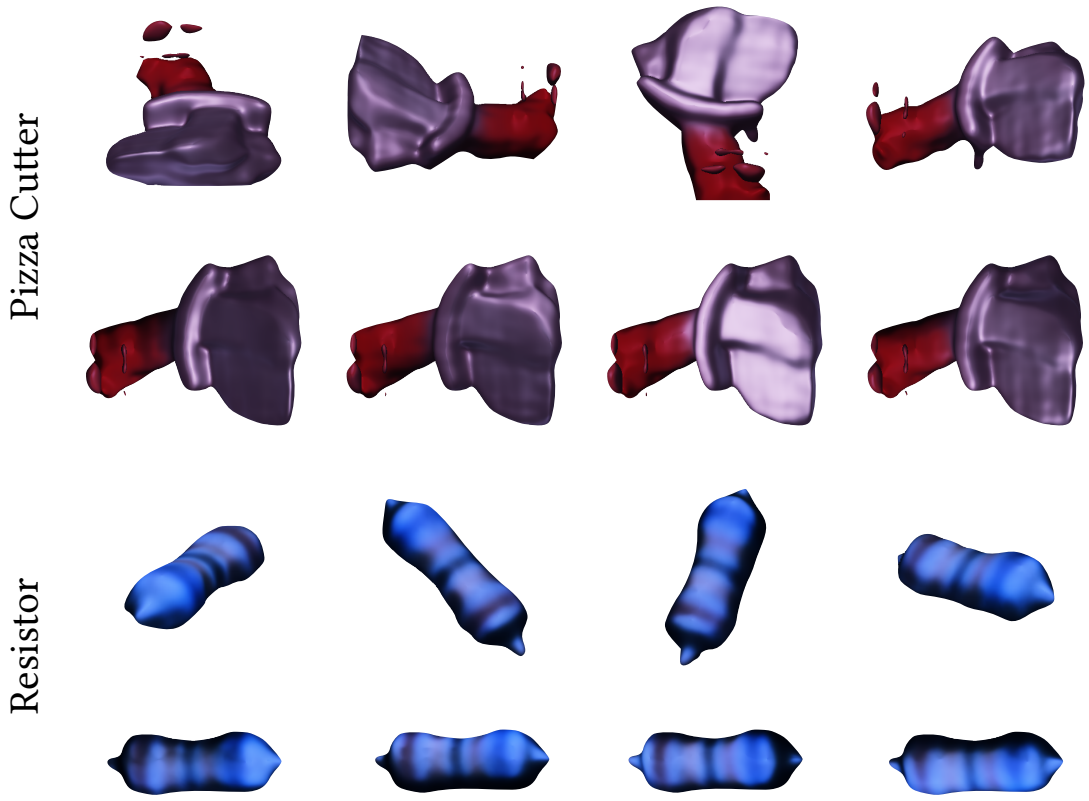

Figure 6.3: **Viewpoint synthesis.** Viewpoint synthesis (top row for each object) and illumination synthesis (bottom) for the pizza cutter and resistor objects. The pizza cutter appearance is well reconstructed, while the higher-frequency details on the resistor are lost.

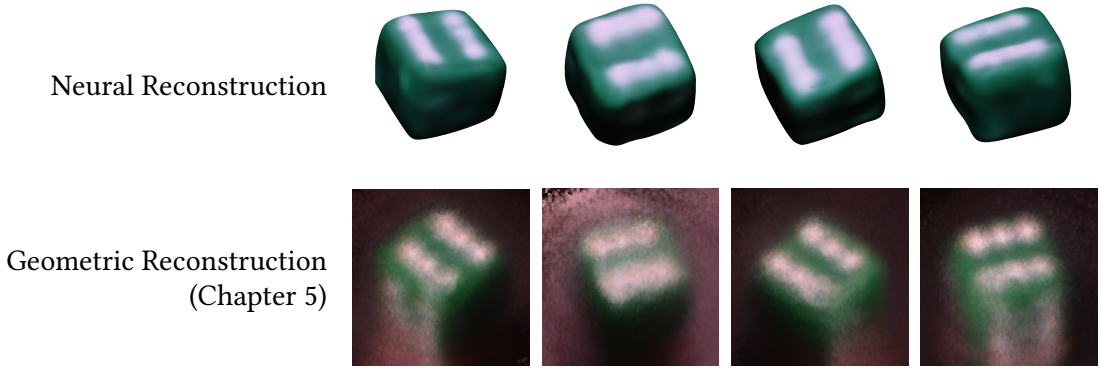

Figure 6.4: **Viewpoint synthesis comparison.** Comparison of the viewpoint synthesis of a green die from our neural approach to that of Chapter 5.

high frequency elements, be they in reflectance (sharp specular highlights) or in texture (sharp changes in material or color), proved difficult to reconstruct in many cases. Additional development is needed to properly account for these effects in the model. Furthermore, the sparse, directional illumination used in this work provided a simple framework for including calibrated light sources in the model, but shadowed regions produced as a result of this impacted the reconstruction quality, particularly from our lab prototype. Devising a more sophisticated scheme for calibrated illumination could provide benefits in the coverage of reconstruction results.

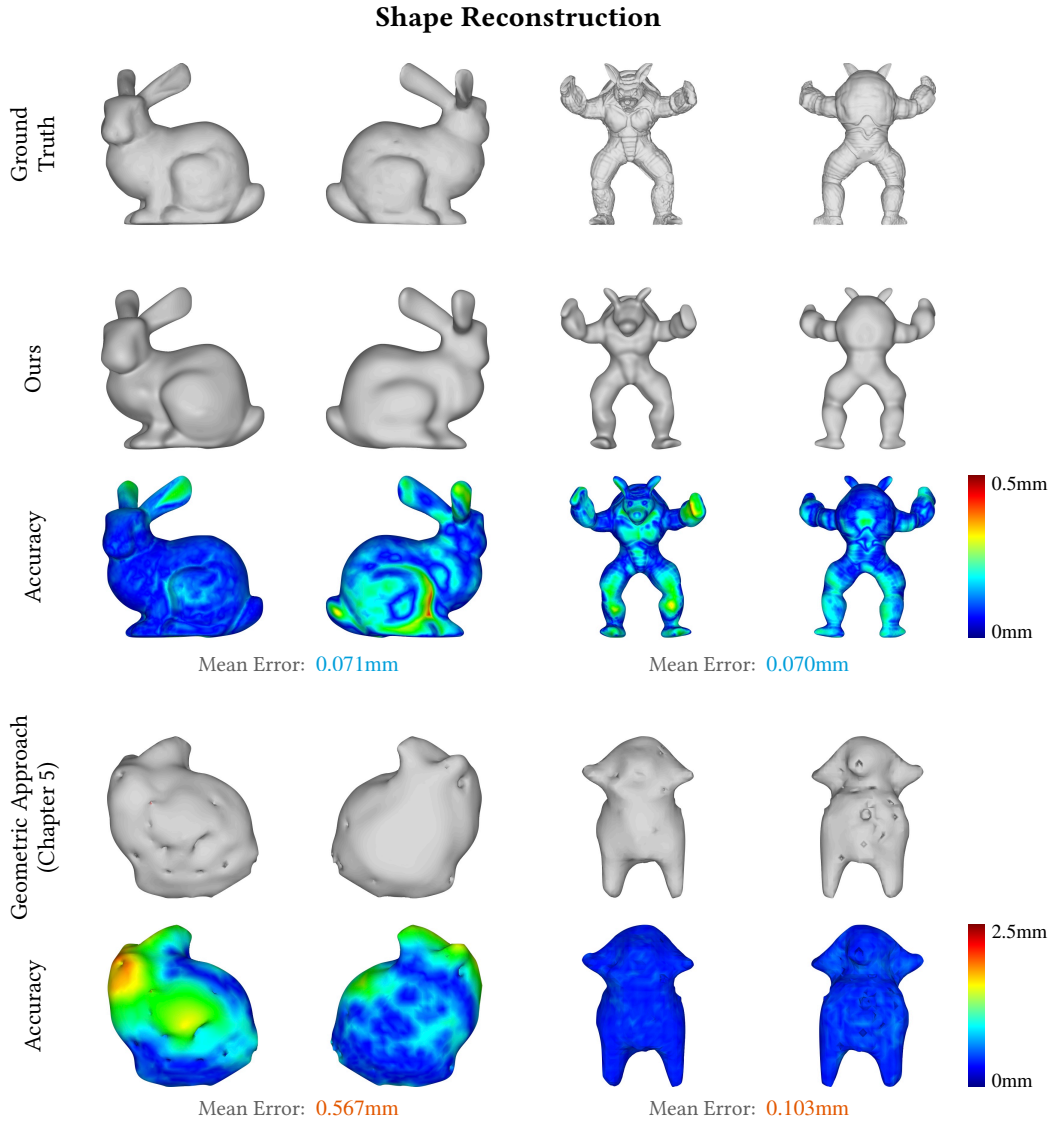

Figure 6.5: **Simulated 3D reconstructions.** Simulated 3D shape reconstruction results for our technique compared to Chapter 5. We visualize the ground truth mesh, the estimated meshes, and the Hausdorff distance used to show the reconstruction error. The prior technique requires very high frequency textural information and fails when presented with reflectance distribution that are more commonly encountered, as we have shown here. Using data that is piecewise-continuous in reflectance, we are able to reproduce all but the finest of shape details. Both the bunny and armadillo reconstructions have a mean error $\sim 70\mu$m. The strongest error is seen on the backside of the bunny, which is the hardest for our imaging setup to reconstruct as it is oriented away from the sensor. Nevertheless, we are able to reconstruct full 3D reconstructions from a single perspective.

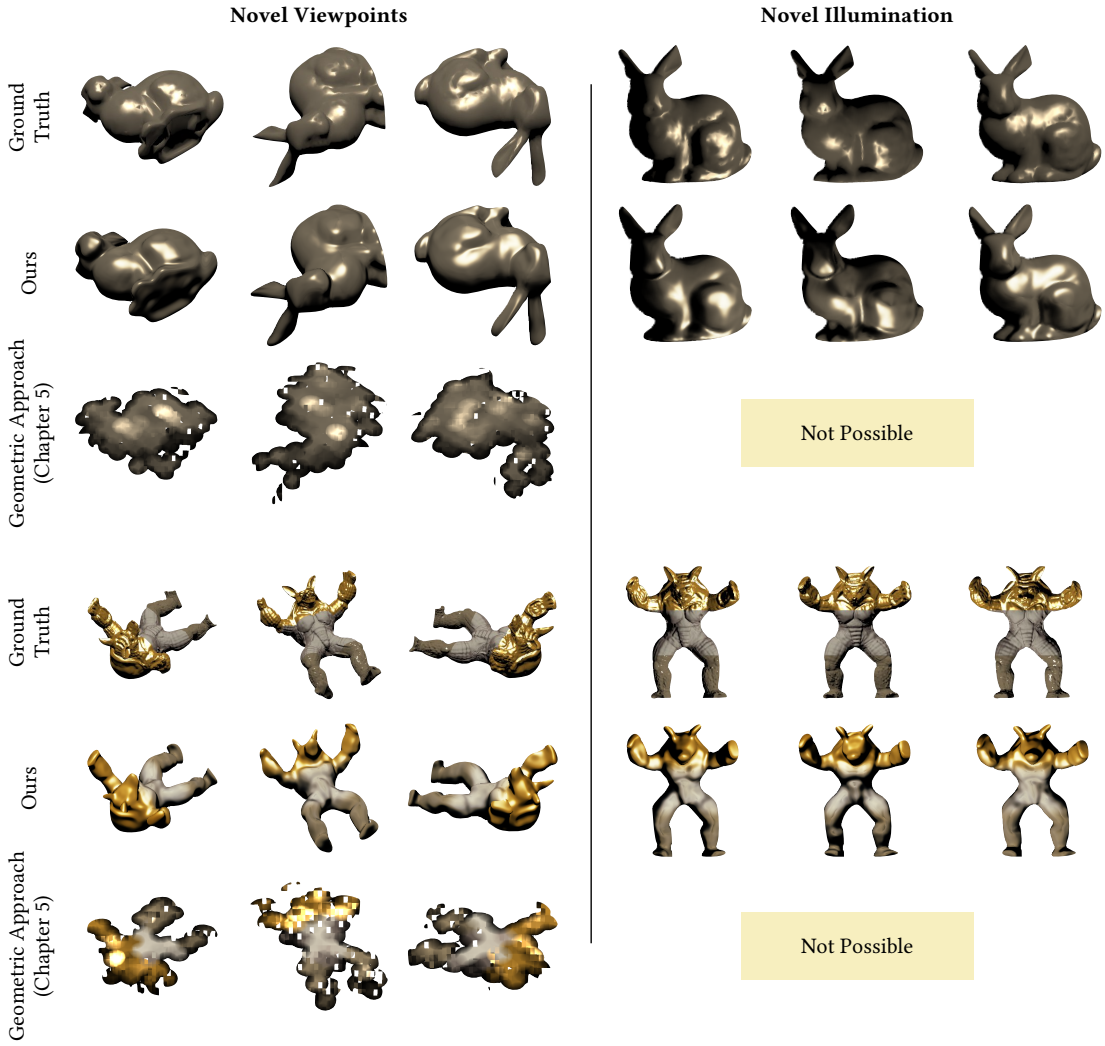

Figure 6.6: **Simulated viewpoint and illumination synthesis.** Simulated viewpoint and illumination synthesis for our technique compared to Chapter 5. We show ground truth renderings of each of the novel viewpoints and lighting directions that we are reconstructing, along with the results from our technique and that of our prior work. We did not have novel illumination rendering previously, and our novel viewpoint renderings are very coarse without the additional exposures that we applied. Our results show proper rendering of moving specular highlights, even those that were not observed by our sensor. In particular, specular highlights with a wider lobe, such as the bunny material or the golden upper body of the armadillo, are able to be reconstructed quite effectively. Very high-frequency specularities, such as can be observed on the legs of the armadillo, are more difficult to reproduce. In all, the interpolation benefits of the neural scene representation allow us to render a broader range of scenes and scene conditions with higher quality.

7 Reconstructing Spatially-Varying Iridescent Reflectance

Characterizing the complex ways in which light interacts with the world around us is a challenging problem. Many materials exhibit particularly intriguing reflectance phenomena that prove especially difficult to both capture and digitally represent. This is particularly true of iridescent objects, where the observed color of the object varies significantly with viewpoint and illumination changes, presenting a challenge for data capture, for we would need to capture every possible combination of observation and illumination directions to fully represent the reflectance characteristics. Additionally, this poses a difficulty for representation, for the color variation occurs at high frequency in angle, can be produced through different underlying physical mechanisms (such as thin-film interference, diffraction and structural coloration), and exhibits characteristics not found in standard reflectance models that focus on diffuse and specular modes.

As a result, most of the recent research in reflectance acquisition has revolved around faster data capture techniques and improved representations that allow for both reduced representation size and less required input data. However, many of these techniques achieve their excellent results through sacrificing one or more elements that make up the full characterization of the scene, whether that be spatial information, high-frequency reflectance representation, or illumination variance. This makes it difficult to fully represent all possible scenes, particularly those with highly complex visible phenomenon such as iridescence.

In this chapter, we apply the wide-baseline light field camera to efficiently capture dense measurements in space, observation angle, and illumination angle using the modified imaging setup of Figure 3.20. We pair this with a neural spatially-varying bidirectional reflectance distribution function (SVBRDF) that can be utilized to capture complex materials with high-frequency features in both space and angle. A summary of the proposed pipeline is introduced in Figure 7.1.

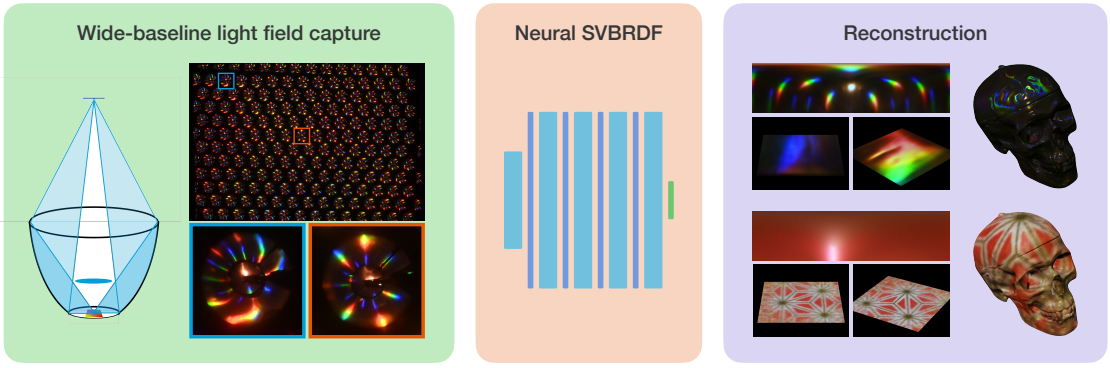

Figure 7.1: **Summary of contributions.** We first measure the SVBRDF of various complex materials through wide-baseline light field measurements, optimizing the imaging system for efficient data capture. We then represent this large dataset as a neural SVBRDF, which can be applied to reconstruct the appearance of the captured materials. We show that this system is capable of capturing and reconstructing highly complex and spatially-varying iridescent reflectance.

Contributions. This chapter presents an optical design and digital reflectance representation to extend the usage of the wide-baseline light field camera to reflectance capture. In particular, we make the following contributions:

- *Efficient SVBRDF Capture using Wide-Baseline Light Field (WBLF) Camera.* We show how the extensions of the WBLF camera for dense illumination control and extended angular range efficiently measure the broad gamut of the SVBRDF.
- *Neural SVBRDF Modeling.* We build a neural architecture that is capable of characterizing high-frequency information in space and angle. This model is utilized in other renderers as a texture map of BRDF features.

Our system is capable of scanning objects that exhibit rich iridescence, including the eye of a peacock's feather that shows structural coloration and craft materials that have diffractive light dispersion similar to what is observed from a grating. Such results present an advance in our ability to scan materials that exhibit extremely high-frequency variations in spatial and angular dimensions of its reflectance.

Limitations. Several key limitations of this work are driven by the size of the ellipsoidal mirror used, which dictates the spatial extent of the materials that can be measured—this limit is 20 mm × 20 mm in our setup, with reconstruction quality fading some toward to outer extreme of that range. Large ellipsoidal mirrors are also extremely expensive to manufacture, especially at the surface quality typically

seen in imaging instruments. Consequently, our system, which is built out of a significantly lower-quality mirror, requires painstaking calibration; despite this, the quality of results we obtain is still adversely affected by deviations from the ideal shape and reflectance of the mirrored surface. Another limitation induced by the ellipsoidal mirror is the shallowness of the depth of field on the target, which limits us to planar targets. We additionally share limitations from the use of the light field camera, which dictates a space-angle trade-off in measuring capabilities. While the mapping through the ellipsoidal mirror changes how this is realized, we still fall short of theoretical limits on angular and spatial resolution based on this trade-off. Finally, while we introduce additional optics in the form of a Fresnel lens to extend the angular range of the system, the resulting system still fails to provide a complete sampling of the reflectance cone.

7.1 Prior Work

There are two broad categories of work that broadly intersect with the proposed work: acquisition systems for reflectance and neural representations for the same. We briefly discuss both.

7.1.1 Systems for Reflectance Measurements

Complete measurements of the BRDF are classically achieved using a gonioreflectometer, where rotation control of the illumination and sample with a fixed detector is utilized to capture all possible illumination/observation angle pairs [Li *et al.*, 2006]. Assuming a uniform material allows for a reduction in capture time and data requirement by using a camera to simultaneously measure from multiple surface points at once [Marschner *et al.*, 2000, Matusik *et al.*, 2003], including for anisotropic materials [Filip *et al.*, 2014]. These systems don't account for spatial variance in the material samples, instead using the spatial extent captured by the camera to generate additional samples of the uniform material. We can account for the spatial variance with additional hardware, such as using a light stage [Debevec *et al.*, 2000, Ghosh *et al.*, 2011, Joo *et al.*, 2017, Schwartz *et al.*, 2013] comprising of multiple cameras, light sources and projectors. These systems are resource intensive, often consisting of tens to hundreds of devices that need to be mechanically secured and calibrated.

An alternate approach, more geared for low-cost applications, is to use a catadioptric camera, which adds mirrored optics to an imaging system. One catadioptric design uses a kaleidoscope to scan the scene [Ahn *et al.*, 2021, Reshetouski *et al.*, 2011, Xu *et al.*, 2018]; in principle, the kaleidoscope acts like a multi-camera system, providing a plethora of novel views of the object which can be used to reconstruct the geometry and appearance of objects. Smooth mirror geometries provide continuous measurements

of material reflectance as opposed to the sparse captures from multiview imaging or kaleidoscopes. In particular, the use of ellipsoidal [Mukaigawa *et al.*, 2007, 2009] and parabolic [Dana and Wang, 2004, Ghosh *et al.*, 2007, Zhang *et al.*, 2015] mirrors takes advantage of the focal points of these geometries to get many observations of a single point at once, providing dense BRDF measurements over a broad angular range. However, the measurements all come from the focal point, so they do not account for spatial variance. The wide-baseline light field camera that we introduce in Chapter 3 combines the ellipsoidal mirror with a light field camera to capture a large range of angular and spatial measurements. We utilize the variation shown in Figure 3.20 to further add dense illumination control and extended angular range measurements.

7.1.2 Neural Reflectance Representations

Traditional BRDF representations are usually either model-based [Soldado and Almagro, 2012] or data-driven [Hui and Sankaranarayanan, 2015, Matusik *et al.*, 2003]. Neural networks provide a way to either improve interpolation of large datasets or learn representations with less input data.

Much of the research in this space involves using neural networks to learn SVBRDF parameters like diffuse albedo, specular albedo, glossiness, and more [Dong, 2019]. These techniques are able to learn these parameters from very few images, but are limited to materials that match their model assumptions. More complex materials, like iridescent objects, are unable to be captured either due to the limited data missing the regions of interest or the limitations of the model itself.

Recent work in implicit neural 3D scene representations [Sitzmann, [n.d.], Tewari *et al.*, 2022] is able to reconstruct more general shape and appearance information. Much of this work is focused on recovering the scene geometry and appearance, and does not focus on illumination variation or producing BRDF representations that can be rendered in other settings. Yariv et al. [Yariv *et al.*, 2020] use a signed distance function (SDF) to (neurally) represent the shape and a reflectance network to capture the viewpoint-dependent appearance of the object. The separability of these networks allows the reflectance network to be transferred to alternate geometry, though the illumination conditions are baked into the representation. Zhang et al. [Zhang *et al.*, 2021] allow for explicit illumination modeling so as to decompose "appearance" more broadly into its physically-meaningful constituents, namely, lighting and BRDF; however, the underlying representation of the scene is still via volumetric radiance fields, which are not transferrable beyond the trained scene. Yariv et al. [Yariv *et al.*, 2021] and Wang et al. [Wang *et al.*, 2021] provide a hybrid volume-surface representation to address this gap between shape and volumetric radiance, with the eventual goal of a principled handling of reflectance and illumination.

Our approach aims to leverage the generalizability of implicit scene representations to create a neural representation of the SVBRDF itself. Our architecture is most similar to the reflectance network in IDR [Yariv *et al.*, 2020], with the added measures of illumination control, input reparameterization, and improved high-frequency response. This has the added benefit of allowing our learned SVBRDFs to be easily rendered on trained IDR geometry.

7.2 Neural SVBRDF Representation

In this section, we introduce our neural representation for the SVBRDF, which trains over the data measured with the previously outlined optical system.

7.2.1 Network overview

Capturing a full dataset with this optical design with 20° illumination spacing in azimuth illumination measurements produces $9172877 \times 18 \times 12 \times 3$ measurements (our calibration produces 9172877 measured rays from the light field camera, the illumination accounts for 18×12 measurements, and we capture three color channels for each measurement). This prompts us to develop a representation of the data that is equally as expressive but requires less storage. We achieve this with the neural architecture shown in Figure 7.2. The network consists of four fully-connected layers of size 512, comparable to the rendering network employed in Implicit Differentiable Rendering (IDR) [Yariv *et al.*, 2020]. We pass as inputs for each measured pixel p the (x, y) coordinates on the observed plane, the viewing direction $\mathbf{v}$, the set of m observed lighting directions $\{\ell^i | i \in 0..m\}$, and the surface normal $\mathbf{n}$ of the observed plane. The network outputs the set of predicted RGB colors $\{\mathbf{c}^i | i \in 0..m\}$ for each lighting direction m, which is trained based on the set of ground truth RGB observations $\{\mathbf{I}^i | i \in 0..m\}$. This setup allows the network to be used as a BRDF texture map, where (x, y) act as the texture coordinates. This makes it easy to transfer and render our material captures on other geometries.

7.2.2 Generating Inputs

We apply the calibration procedure of Chapter 4 to calibrate a light field ray origin and direction for each light field pixel. We additionally apply the helper camera to capture a checkerboard on the LCD where we will place the object. Once the measurement plane origin and orientation are known, we project the light field rays to this plane, providing 2D texture coordinates (x, y) based on the intersection location. These spatial coordinates paired with the observation direction of each ray form the main input to the

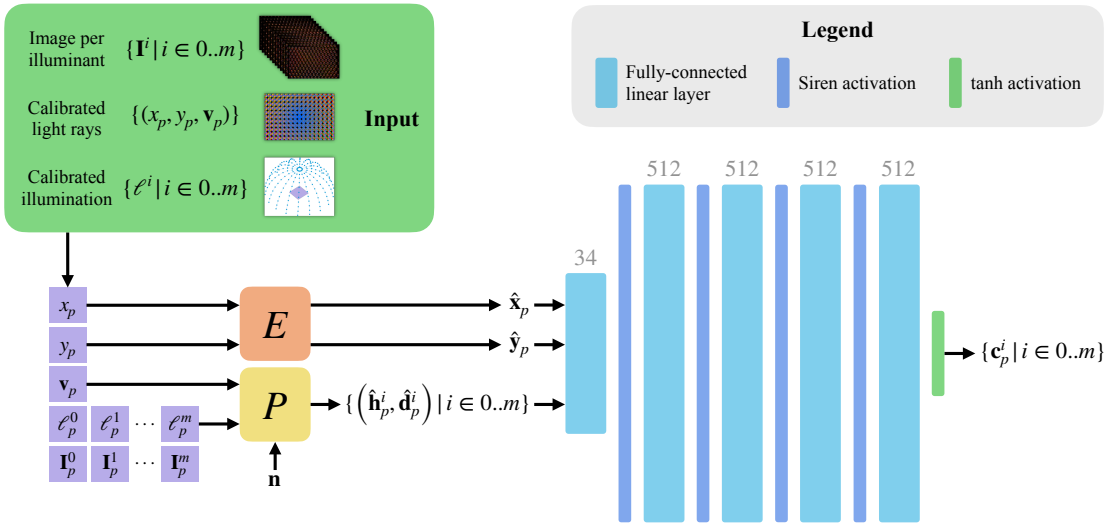

Figure 7.2: **Overview of the network design to represent the SVBRDF.** We take as input the captured image data, one image per illumination direction, along with the calibrated light rays for each light field pixel and the illumination direction for each pixel p and illuminant i. The observation and illumination angles are transformed by the block P along with the surface normal $\mathbf{n}$ to produce the reparameterized features $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{d}}$ for each pixel. The spatial coordinates (x, y) for each pixel are encoded into Fourier features $(\widehat{\mathbf{x}}_p, \widehat{\mathbf{y}}_p)$ by the block F. The resulting input vector is of size 34. The main network consists of four fully connected layers of size 512 (light blue), with a sinusoid activation function between each layer (dark blue). The final output is a standard tanh activation (green), producing the predicted color $\mathbf{c}$ at a given pixel p for each illumination direction i (color values during training are mapped from $[-1, 1]$).

SVBRDF network. The calibrated plane also means the surface normal $\mathbf{n}$ is known and passed to the network as input as well.

The dense illumination is calibrated as 3D point locations for each light source, according to Chapter 4. We produce the desired input light direction ℓ for each light source i and pixel p by finding the difference vector between the scene point measured by the light field camera and the location of the light source. The scene point is the intersection point of the light field ray from pixel p and the measurement plane (i.e. the 3D spatial coordinate of the 2D texture coordinates (x, y) produced previously). This means that we produce a unique illumination direction ℓ for each point observed on the plane, providing additional variability of the illumination measurements that we take.

7.2.3 Encoding High-Frequency Features

The fully-connected network should be able to model the complexity of the desired BRDFs, but as Sitzmann et al. [Sitzmann *et al.*, 2020] and Tancik et al. [Tancik *et al.*, 2020] have shown, the network tends to learn low-frequency representations of the data. To improve the reconstruction of high-frequency details, the latter work employs spatial encoding via Fourier features, which have become common in neural scene representations. This involves encoding the inputs (such as the spatial coordinates) as a series of sines and cosines of these inputs with increasing frequencies. We found that while this technique does improve high-frequency reconstruction for spatial locations, it struggles to reconstruct high-frequency features in reflectance when used on viewing or lighting direction inputs. Instead, we employ the Siren activation function technique [Sitzmann *et al.*, 2020], which has also been shown to similarly improve high-frequency performance. This technique uses a sinusoid activation function between layers as opposed to the more common ReLU activation. We observed that Siren improved our reconstruction of high-frequency reflectance features, but failed to do the same for spatial textures. Ultimately, we employed both techniques; we applied spatial encoding to our texture coordinates (x, y) and utilized the Siren activation function throughout our network. Section 7.3.4 goes into detail on our observation that the combination of spatial encoding and Siren activation functions performs best for both complex reflectance and complex texture.

7.2.4 Input Parameterization

To further improve our ability to model the SVBRDF, we use the half-angle reparameterization of [Rusinkiewicz, 1998], which is known to improve the reconstruction of high-frequency reflectance functions. Given the direction to the light source ℓ and the viewing direction $\mathbf{v}$, we compute the half-angle vector $\mathbf{h}$ and difference vector $\mathbf{d}$ as follows. First, we define a coordinate system on a given point using its surface normal $\widehat{\mathbf{n}}$, a surface tangent $\widehat{\mathbf{t}}$, and the surface bi-normal $\widehat{\mathbf{b}} = (\widehat{\mathbf{n}} \times \widehat{\mathbf{t}})$. Next, we transform ℓ and $\mathbf{v}$ into this coordinate system, producing the incoming and outgoing rays $\mathbf{w}_o$ and $\mathbf{w}_i$, respectively. Finally, we compute the half-angle vector $\mathbf{h}$ and the difference vector $\mathbf{d}$:

$$\mathbf{h} = \frac{\mathbf{w}_o + \mathbf{w}_i}{||\mathbf{w}_o + \mathbf{w}_i||}, \quad \mathbf{d} = \text{rot}_{\widehat{\mathbf{b}}, -\theta_h} \text{rot}_{\widehat{\mathbf{n}}, -\phi_h} \mathbf{w}_i, \tag{7.1}$$

where (θ_h, ϕ_h) are the spherical coordinates of the half-angle vector in the $\widehat{\mathbf{n}} - \widehat{\mathbf{t}} - \widehat{\mathbf{b}}$ coordinate system. The spherical coordinates of the difference vector (θ_d, ϕ_d) define the difference in angle between $\mathbf{h}$ and $\mathbf{w}_i$ in the same coordinates. The rotation functions shown represent the rotation by the given angle ($-\theta_h$ or $-\phi_h$) around the given axis ($\widehat{\mathbf{b}}$ or $\widehat{\mathbf{n}}$).

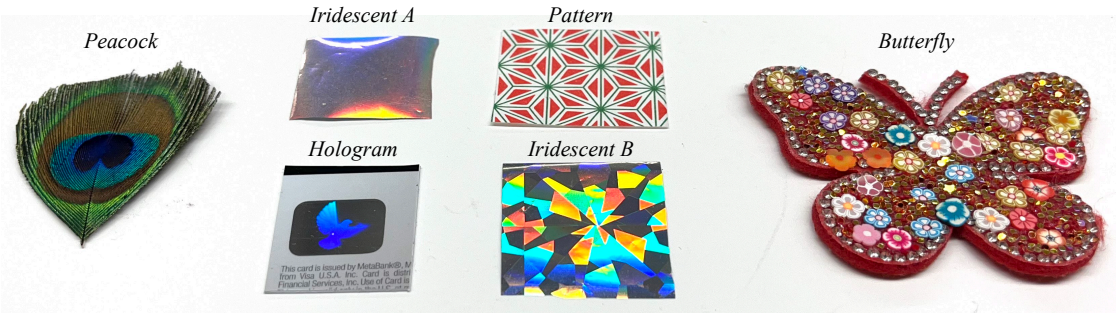

Figure 7.3: **Objects used for evaluation.** We test our proposed imaging and representation using a peacock feather (*Peacock*), uniform iridescent material (*Iridescent A*), patterned origami paper (*Pattern*), holographic dove from a credit card (*Hologram*), spatially-varying iridescent material (*Iridescent B*), and butterfly hair clip (*Butterfly*).

To account for the cyclical and fixed-magnitude nature of the vectors $\mathbf{h}$ and $\mathbf{d}$, we represent them as vectors of the sines and cosines of the azimuth and elevation angles for each vector:

$$\widehat{\mathbf{h}} = [\sin\theta_h, \cos\theta_h, \sin\phi_h, \cos\phi_h]\,, \tag{7.2}$$

$$\widehat{\mathbf{d}} = [\sin\theta_d, \cos\theta_d, \sin\phi_d, \cos\phi_d]\,. \tag{7.3}$$

In place of passing the surface normal $\widehat{\mathbf{n}}$, the viewing direction $\mathbf{v}$, and the lighting direction ℓ to the network, we instead pass only $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{d}}$. This improves our high-frequency reflectance performance.

7.3 Results

We show a range of results for reflectance with high-frequency variation in space and/or angle using the objects shown in Figure 7.3. We produce visualizations of the learned SVBRDF for different input parameters along with renderings of the objects under novel viewpoints and illuminations. Finally, we show reproductions of the captured materials rendered on alternate geometry by texture mapping our learned SVBRDF onto these objects.

7.3.1 Learned SVBRDF Visualization

We first analyze our results by visualizing the learned SVBRDF. We fix the spatial coordinates at $(x, y) = (0, 0)$ and visualize the observed color at every observation angle over the hemisphere for a set of different illumination conditions. Figure 7.4 shows these results for five different material captures under three different novel illumination directions. We observe that we capture the different ways that iridescent

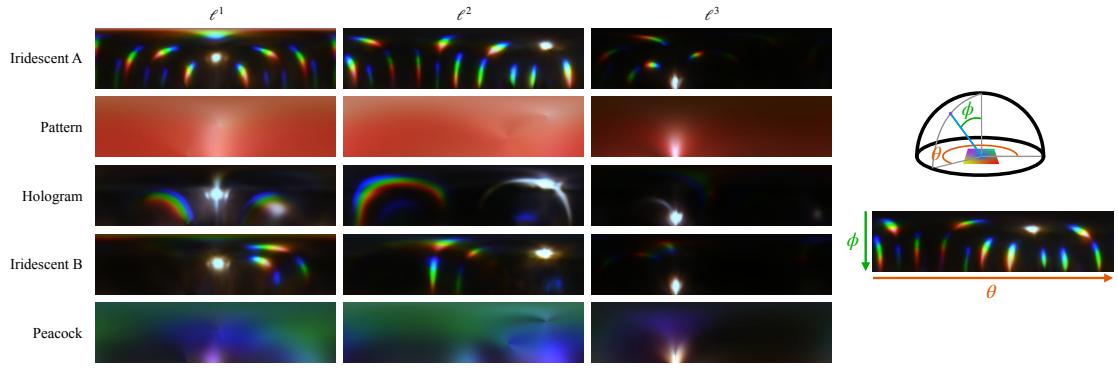

Figure 7.4: **Visualization of the SVBRDF learned for various input materials.** We fix the spatial coordinates at $(x, y) = (0, 0)$ and plot the observed color at viewing direction $\mathbf{v}$ by parameterizing $\mathbf{v}$ according to its azimuth angle θ and elevation angle ϕ, as defined in the illustration on the right. Each column is generated with a different fixed illumination direction ℓ^i.

material can present itself, with the objects *Iridescent A*, *Iridescent B*, and *Hologram* displaying different iridescent diffraction effects. Also, the mostly-diffuse *Pattern* object shows consistent diffuse reflectance when the illumination elevation angle is high (columns one and two) but behaves more specularly at grazing angles, consistent with known behavior.

Some artifacts do appear in these results, particularly noticeable for the diffuse *Pattern* object. First, there is a noticeable color shift moving toward the upper portion of these plots. This is due to the differences in imaging between the lens (upper portion of plot) and the mirror (lower portion). The quality of the Fresnel lens does tend to degrade the color reproduction in practice, leading to more washed out colors in the upper portion of the plot.

7.3.2 Novel Renderings

We next consider our SVBRDF representation by producing novel renderings of the captured material planes. Figures 7.11 and 7.12 show a set of renderings at different viewpoints with two different illumination directions. We render a 20mm $\times$ 20mm area of the captured material. These results highlight the angular and spatial effects that we are able to reproduce. We can reconstruct materials with either high angular frequency (*Iridescent A*), high spatial frequency (*Pattern*), or both (*Iridescent B*, *Hologram*) Particularly, the *Hologram*, a common security feature on many credit cards, is exemplary of the capabilities of the proposed method. The dove appears in iridescent color in certain observation/illumination conditions, but completely disappears into the background in others, which is consistent with the observed effect from our ground truth observations.

Figure 7.5: **Fixed viewpoint renderings.** We render the captured *Hologram* with a fixed viewpoint while moving the illumination in an arc over the object from right to left (see diagram). Note that the dove's left wing is more extended when illuminated from the right (orange) and more retracted when illuminated from the left (green). This creates an effect where the dove appears to rotate in space.

We further render the *Hologram* under a fixed viewpoint with varying illumination. Figure 7.5 shows this set of renderings, which reveal the holographic nature of this material. The dove is designed to appear to rotate slightly under changing illumination, which we observe in this set of renderings as well. Capturing this result requires the measurement of fine details in both space and angle, which our design delivers in an efficient way.

We observe some artifacts from errors in the object plane estimation; due to the very shallow depth of field of the wide-baseline light field camera, small errors in plane estimation lead to computed intersection points (x, y) having errors that are most significant at these grazing angles. These artifacts are most apparent near the outer range of the material, most visibly on the *Pattern* object. This is expected due to the fall-off in the number of measurements, as discussed in Chapter 3. The quality of the reconstruction is also dependent on the observation elevation angle. Figure 7.6 shows a set of renderings with fixed illumination at various observation elevation angles. As we move into the previously missing central cone, which we aimed to fill with the Fresnel lens, additional artifacts appear. Relative to the wider set of angles captured through the mirror, we capture far fewer measurements in these regions. Some degradation in results is expected therefore, but the filled cone allows for any observations in this range at all, as discussed in Section 7.3.4. Overall, the central region reproduces the captured material faithfully and the full area is captured with some artifacts.

7.3.3 Material Transfer

Finally, we demonstrate the applicability of our captured SVBRDFs by rendering them as BRDF texture maps on different geometry. Figure 7.7 shows each captured material rendered on the skull geometry

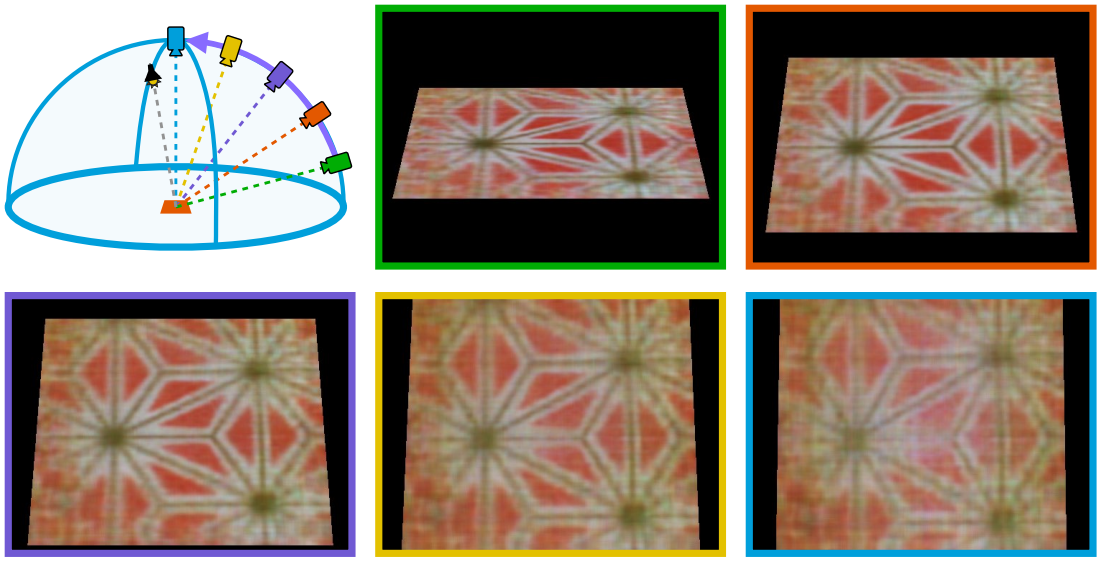

Figure 7.6: **Fixed illumination renderings.** We render the captured *Pattern* with a fixed illumination direction while moving the viewpoint in an arc in elevation (see diagram). We note the increase in artifacts as we move closer to a head-on view; this is due to the relatively limited number of observations from these angles and the optical quality of the Fresnel lens. However, we only achieve these viewpoints at all due to the addition of the Fresnel lens.

from a pre-trained IDR network [Yariv *et al.*, 2020], where we replace the rendering network with our own trained SVBRDF representation. This shows the captured materials under a range of different observation angles and surface normals that were not part of training. The illumination is set to shine top-down in these images; note that we do not model self-occlusion in these renderings.

7.3.4 Ablation Study

These results exemplify the capability of our proposed imaging system and neural SVBRDF representation. To empirically analyze the impact of particular design elements on our reconstructions, we consider removing one element at a time in Figure 7.8. In order, we consider our system with no added Fresnel lens, no Siren activation between linear layers, and no positional encoding on the input spatial locations. We finally consider removing the Siren activation and replacing it with Fourier feature encoding (similar to the positional encoding) on the observation and illumination angle inputs.

The material transfer results in particular highlight the impact of the Fresnel lens addition on our ability to capture on model these materials. We observe how the lack of this lens prevents the recon-

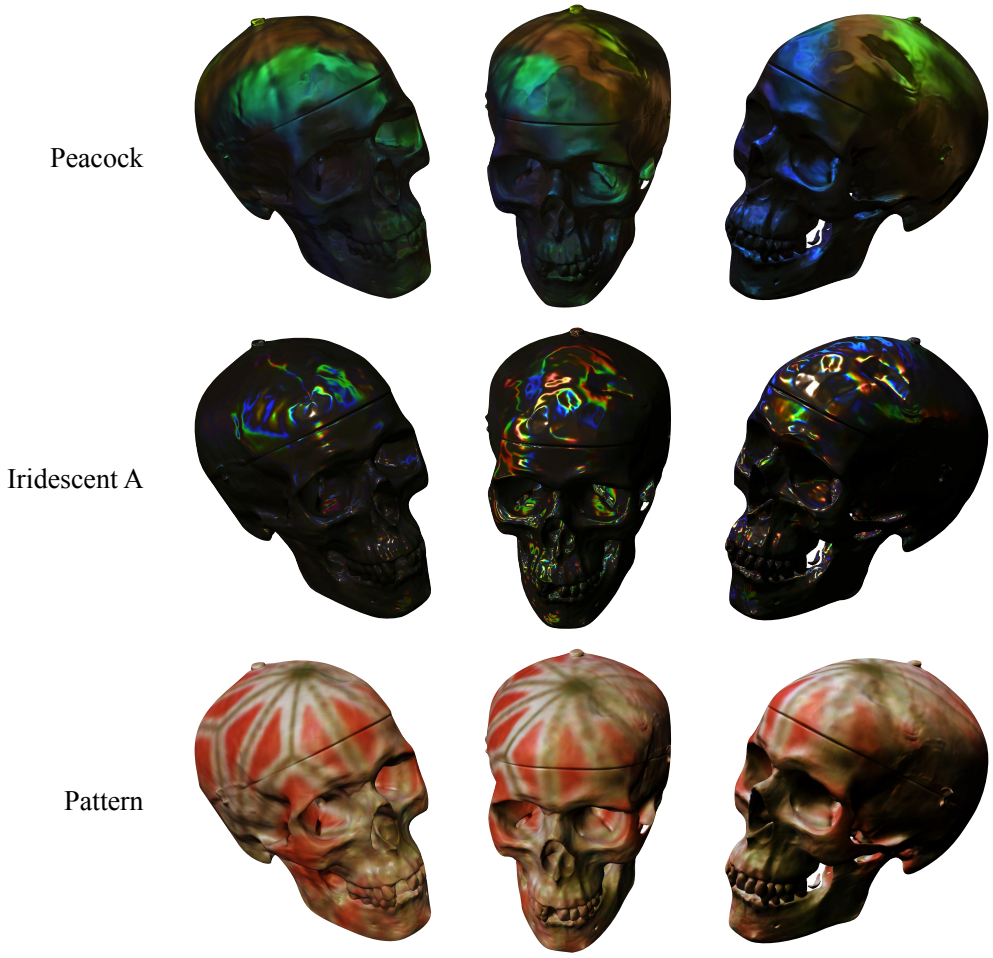

Figure 7.7: **Rendering on other geometry.** We render pre-existing geometric models using our captured SVBRDFs as texture maps. This tests our reconstruction over a range of different surface normals and observation angles, and we see both the iridescent effects and the spatial textures successfully rendered.

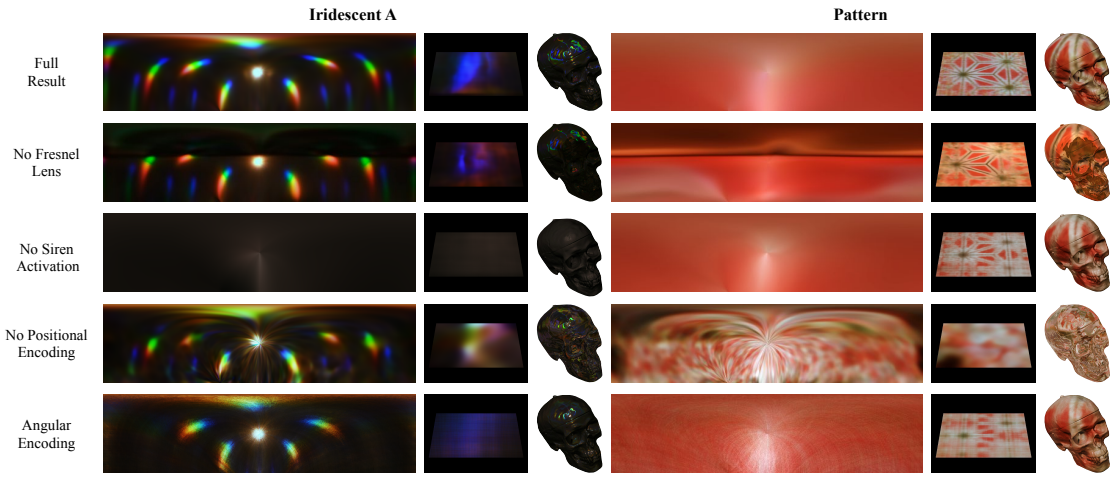

Figure 7.8: **Ablation study.** The impact of major design decisions on reconstruction results. We show a BRDF plot, rendering, and material transfer for two materials (*Iridescent A* and *Pattern*) both with and without the added Fresnel lens, positional encoding, and Siren activation. The addition of the lens fills in the missing central cone from the ellipsoidal mirror, allowing the results to actually be used in rendering over a full set of viewpoints. Siren activation ensures that we can characterize high-frequency angular effects like iridescence; without it, we lose all of these effects but still maintain the ability to reconstruct spatial texture through the positional encoding. Without the positional encoding, the reflectance of each point is accumulated from a large area of neighboring points, removing our ability to reconstruct high-frequency effects in either angle or space. The final row illustrates our reconstruction when removing the Siren activation and utilizing Fourier feature encoding on the observation and illumination angular inputs to the system. While we are able to resolve some high-frequency reflectance effects, we miss many of the highest-frequency elements and observe introduced angular artifacts.

struction of the central cone of light leaving each scene point. This cone corresponds to the ability to reconstruct the appearance of surface normals oriented toward the observer.

The use of positional encoding ensures that high-frequency spatial features can be faithfully reconstructed. Without it, even with the Siren activation function, neighboring spatial regions are combined and treated as angular variation, leading to the BRDF plots shown.

The Siren activation technique improves our angular reflectance reconstruction. We observe the loss of many high-frequency reflectance elements when we revert back to just using spatial encoding to treat the angular features.

Removing the Siren activation and using Fourier feature encoding on the angular components of our system is able to resolve some high-frequency angular effects. However, we still miss some of the more

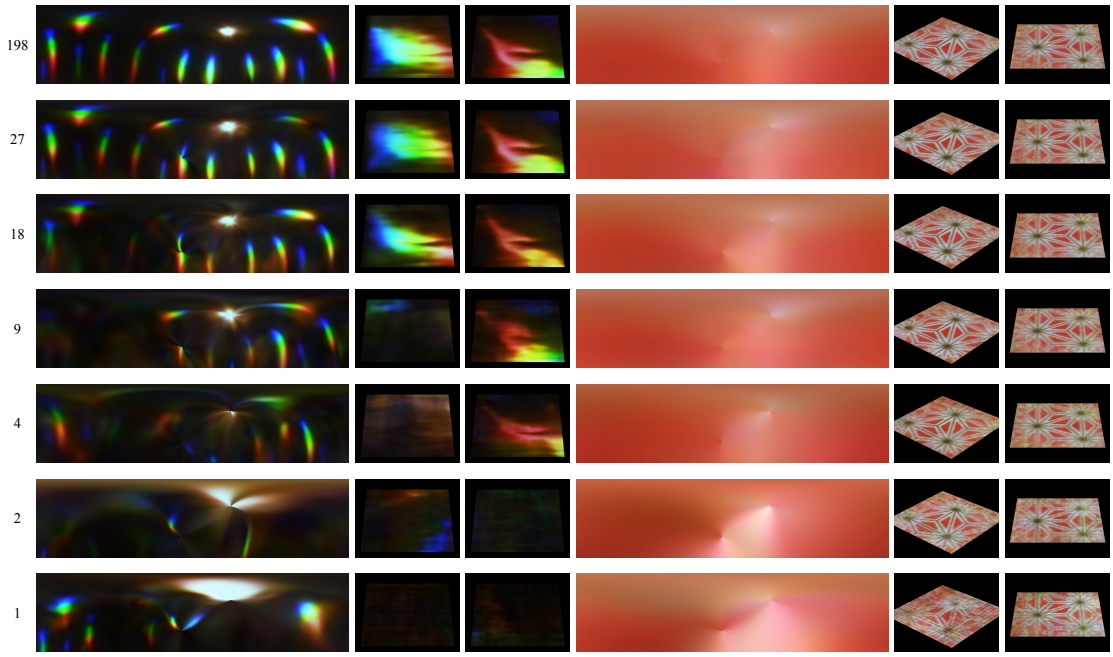

Figure 7.9: **Consideration of the number of training images.** We consider the impact of reducing the number of measurements that we train over. We show a BRDF plot and two novel view renderings for two materials (*Iridescent A* on the left and *Pattern* on the right) with decreasing number of measurements, listed on the left. For the diffuse object (right), the impact is minimal. For the iridescent object (left), the impact is significant. This indicates that the density that we can achieve with our imaging system is necessary to capture these high-frequency iridescent effects.

narrow (and thereby high-frequency) elements of the iridescent reflectance, and we start of observe some angular artifacts in our reconstruction due to the high-frequency encoding. This supports our empirical analysis that the Fourier features work well in our system for positional encoding of spatial features while the Siren activation enables the reconstruction of high-frequency angular reflectance.

We further consider the impact that the number of captured images has on our reconstructions. Recall that each additional image captures an additional illumination condition, but the viewing angles and spatial measurements remain constant. We show the BRDF visualization and viewpoint synthesis for the *Iridescent A* and *Pattern* objects in Figure 7.9. All reconstructions are shown for an illumination condition not seen in training, generally falling in between captured measurements, which are spaced out over the hemisphere.

Our main results have utilized 198 images, aligning with the top row. This takes advantage of the dense measurements that our imaging device can achieve. However, the neural SVBRDF representation

is powerful enough to interpolate between sparser measurements. As we move to fewer measurements, we observe that our reconstruction of spatial textures for the diffuse object is hardly impacted. Our network is able to interpolate easily when there is a high amount of viewpoint consistency in color, which is true for diffuse objects like this one. However, as we do the same with the iridescent object, we note that we lose more and more lobes of color and introduce additional artifacts as we reduce the number of measurements. The dense measurements of our wide-baseline light field camera are necessary to capture these effects.

7.4 Discussion

We introduced an imaging system and neural SVBRDF representation capable of capturing and reconstructing the reflectance of complex materials. Our proposed imaging system extends the capabilities of a wide-baseline light field camera to capture observations over a full hemisphere surrounding the material in a single exposure, enabling efficient reflectance measurements. The light field imaging capability also provides measurements in space along with these diverse angular measurements in the same exposure. We additionally utilize an illumination system to provide illumination control over a full hemisphere. This imaging system provides efficient high-quality measurements that we can use to train a neural SVBRDF representation for a broad range of complex materials. We show that we are capable of reconstructing complex, high-frequency details in angle (iridescence) and space (spatial textures).

Limitations We conclude by considering the limitations of our approach. We noted at the start of this chapter several limitations, including the size of the capture area, limitations in quality imaging optics, and standard light field space-angle trade-offs. These limitations are largely fixed into the optical design based on the ellipsoidal mirror and light field camera parameters, though further refinement of the calibration techniques outlined here could overcome some limitations in imaging optics.

We further note that we are limited in our capture to planar materials, as we do not reconstruct any geometry information in our training. This creates problems when reconstructing even mostly-planar materials, as our depth of field is very shallow. Figure 7.10 shows the attempted reconstruction of a butterfly hairclip. While mostly planar, the varying heights on the surface prevent quality reconstruction at these different depths. We are able to reconstruct several features, including the texture of the flowers and the iridescent star, but many details at different depths are lost. We limit ourselves to planar objects due to the high-frequency details we aim to capture in angle. This makes common shape reconstruction techniques, which rely on color consistency across observations, difficult to apply to our data.

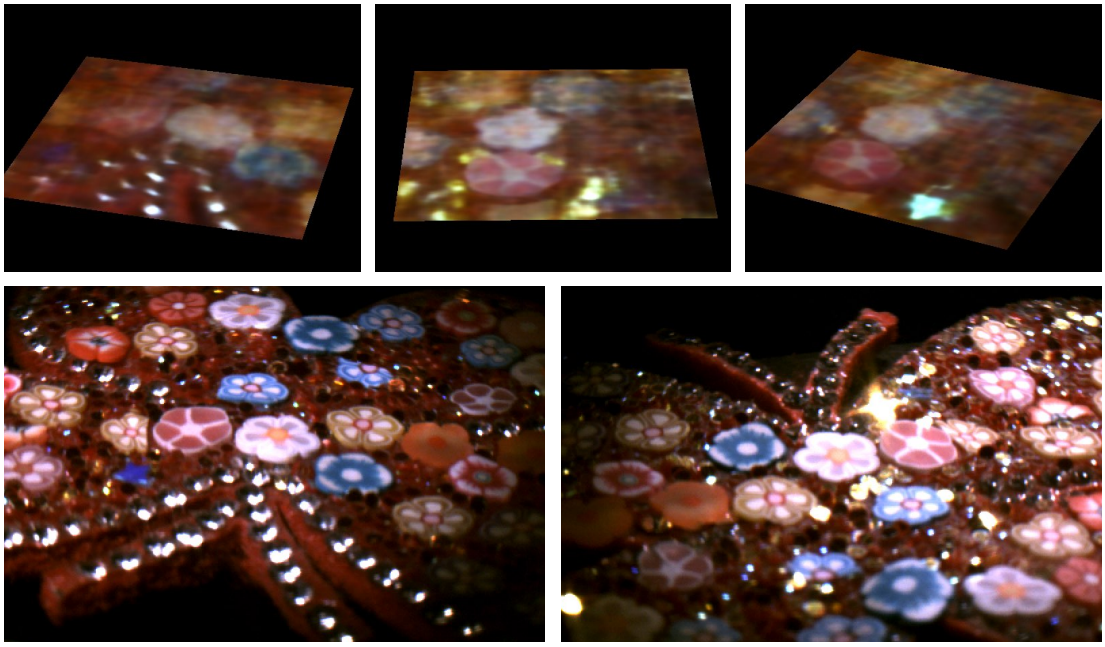

Figure 7.10: **Limitations due to narrow depth of field.** We render the *Butterfly* (top) and reference images of the clip (bottom). Due to the varied surface height of different elements and the limited depth of field of our imaging system, we are unable to resolve each element in focus with our current technique.

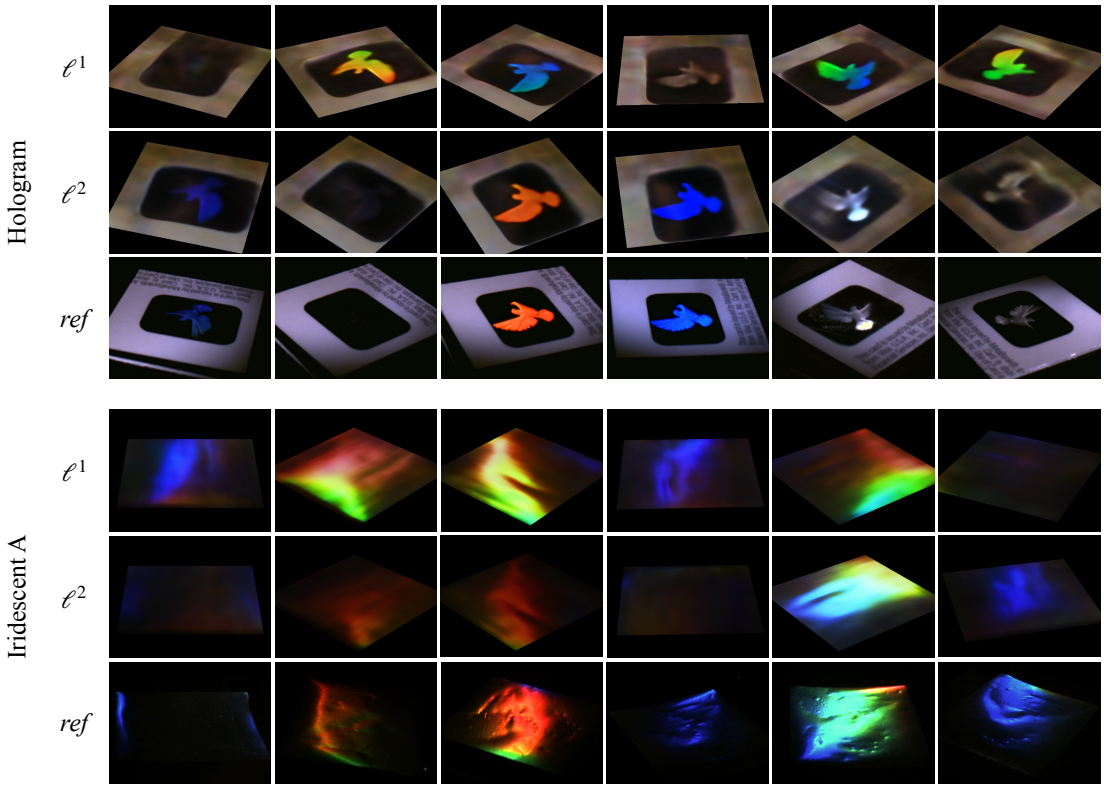

Figure 7.11: **Novel view synthesis.** We show the rendering for the first two of four different material captures under novel viewpoint and illumination conditions. The first and second rows for each object, labelled ℓ_1 and ℓ_2, show two different sets of novel viewpoints under two distinct illumination directions. The third row, labelled *ref*, shows reference captures of the object using a color camera and directional light source. This serves as our visual reference, though the exact observation angles and illuminations of these captures are not necessarily identical to the rendered results as the camera and light source were not calibrated. However, we observe the same iridescent phenomenon in our renderings that are seen in the reference images. In particular, the *Hologram* had sufficient spatial information to nearly align our rendered images with the ground truth; as a result, we produce a matching rendering in row ℓ_2 for each object view in row *ref*.

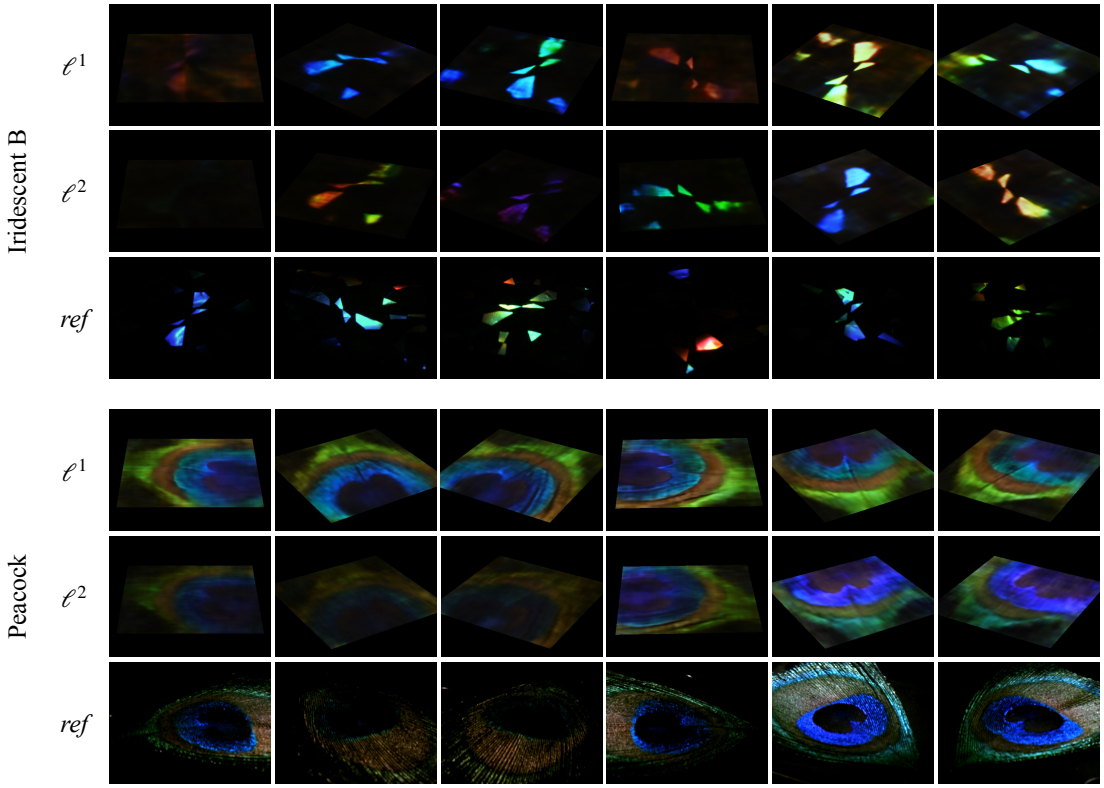

Figure 7.12: **Novel view synthesis.** We show the rendering for the second two of four different material captures under novel viewpoint and illumination conditions. The first and second rows for each object, labelled ℓ_1 and ℓ_2, show two different sets of novel viewpoints under two distinct illumination directions. The third row, labelled *ref*, shows reference captures of the object using a color camera and directional light source. This serves as our visual reference, though the exact observation angles and illuminations of these captures are not necessarily identical to the rendered results as the camera and light source were not calibrated. However, we observe the same iridescent phenomenon in our renderings that are seen in the reference images.

8 Conclusion

8.1 Thesis Contributions

Light fields have a well-established set of applications and benefits, ranging from image-based rendering to full scene reconstruction. When captured fully over the entire gamut of light rays leaving a scene, these tasks are well-defined but data-intensive. Light field cameras capture information with much less data required, but they capture a narrow slice of the full light field. This thesis introduces a novel imaging system for capturing light fields over a broad angular range, which we denote as a wide-baseline light field camera. Our proposed camera retains the data efficiency of prior imaging systems while expanding the measurements of the light field. We paired this camera design with a range of processing algorithms to apply our wide-baseline light field data to challenging tasks such as 3D shape and iridescent reflectance reconstruction.

Overall, we have contributed to light field imaging and applications through the work of this thesis:

- We develop the imaging system for a wide-baseline light field camera and analyze its capabilities. We show how this design improves upon prior light field cameras for the purpose of capturing light fields over a wide angular range. We further present the necessary calibration procedures to utilize this camera in practical applications.
- We provide a geometric parallel to conventional light field processing by producing refocusing, viewpoint synthesis, and 3D shape reconstruction. In particular, we show how wide-baseline light field data enables a novel shape reconstruction algorithm that jointly solves for 3D shape and surface normals.
- We improve further upon 3D shape reconstruction from wide-baseline light field data by utilizing an implicit neural scene representation. This implementation improves our reconstruction of real captured scenes and enables illumination synthesis.

- We apply our wide-baseline light field camera to capturing the high-dimensional space of the spatially-varying bidirectional reflectance distribution function (SVBRDF). We utilize dense illumination control paired with the efficient spatio-angular measurements of our camera to efficiently measure the broad gamut of the SVBRDF. We further develop an implicit neural representation of the SVBRDF to provide a lightweight data-driven reflectance model. In particular, we show that our imaging and representation is able to reconstruct spatially-varying iridescent objects, which represent a highly complex class of materials that have proven challenging to capture and represent.

8.2 Future Work

This thesis lays the groundwork for the imaging and application of wide-baseline light fields. Further work in this space can build upon this thesis in several areas.

8.2.1 Expanded Imaging Area

To promote the application of wide-baseline light fields in more general scenes, the spatial extent of the captured light fields would need to be expanded. This could possibly require the development of alternate mirror geometries that can map a larger area to the same sensor area. Additionally, work to produce low-cost, high-quality ellipsoidal mirrors that are much larger than our setup would enable an expansion of the imaging area.

8.2.2 Reduced Imaging Area

Conversely, we could also imagine aiming to miniaturize the imaging area and improving the spatial resolution for applications in microscopy. Light field microscopy has already shown success using powerful microscope objectives [Levoy *et al.*, 2006], but our setup features a broader angular range and smaller depth of field. These features can be utilized to improve depth resolution of the resulting scans.

8.2.3 Iridescence and 3D Geometry

Our work has shown the wide-baseline light field camera used for both 3D shape reconstruction and spatially-varying iridescent reflectance reconstruction. Combining these two results to be able to reconstruct spatially-varying iridescent objects with arbitrary geometry is a very challenging problem; the high-frequency angular features of iridescent reflectance impact the ability for shape reconstruction algorithms to rely on color consistency as a loss term. Different views of the same scene point may

appear drastically different, which is then indistinguishable from views of separate scene points. However, we have shown success in both capturing and representing these angular effects, which inspires us to believe that this can be achieved. Further work may require clever illumination control schemes, improved neural architectures, or other undiscovered improvements to this work.

8.2.4 Improved Imaging Optics

The optics used in this thesis have inherent limitations; the optical quality of both the ellipsoidal mirror and the Fresnel lens are less than what can be achieved through other optics, like planar mirrors and conventional lenses. However, ellipsoidal mirrors and very large diameter lenses are not practically manufactured to the requirements of our imaging design. Additionally, the custom-assembled light field camera had limitations to the precisions of our alignment and performance of available optics. This would be improved through a light field camera production in a more precise manufacturing environment. In all, further improvement to the production of these optical elements and precision of their construction would improve the achievable results.

8.2.5 Analysis of Etendue Conservation

The etendue of light propagating through space is the product between the area of the source of light and the solid angle of the propagating cone of light [Lerner and Dahlgrenn, 2006]. Optical elements such as lenses and mirrors preserve the etendue of the light passing through the optical system. In many conventional imaging applications, the conservation of etendue concerns the brightness of the imaged light; the optical system can not increase the brightness of the light passing through the system. In our case, we are attempting to measure the light field at the scene over a large angular range and as large as possible spatial extent. However, at our sensor, we are imaging over a fixed spatial extent and with a narrower angular range. This would appear to break the conservation of etendue; however, we do not believe this to be the case. We expect that the non-uniform mapping of the scene light field produces broader spatial extent but narrower measured cones in some areas and narrower spatial exent with wider measured cones in other areas. The overall effect appears to be simultaneously broad in spatial and angular extent, but this non-uniformity leads to a more complex etendue conservation that behaves differently in different local areas. We would expect a theoretical analysis of this etendue conservation for different mirror geometries would yield fascinating results, particular when considering light field imaging where space-angle trade-offs and sparsity of measurements come into consideration as well.

8.2.6 Neural Calibration

Many neural scene reconstruction techniques [Yoonwoo Jeong and Park, 2021] additionally train over the camera poses of the input images. This self-calibration either improves the calibrated poses of the input cameras or works with uncalibrated cameras and can further improve the reconstruction quality beyond that which is attainable using less precise calibration. Our results using the lab prototype are highly dependent on the calibration of our imaging system; as noted in Chapter 4, this is an intensive process and still retains some noise in its completion. The idea of self-calibration is therefore an appealing one: can a neural network learn the correspondences between light field pixels and object space rays directly from input data images? Future work in this space would be fascinating to explore, as the quality of results in both 3D reconstruction and reflectance estimation could be improved. The complexity and non-uniform mapping of our imaging system would pose some difficulties in this process, but either by training over calibration targets or carefully designing network parameters, this research direction could yield fruitful results.

8.2.7 Expanded Measurements via Object Motion

Instead of changing the optics around an object to produce additionally viewpoints, it is common practice to instead move the object itself to generate different views. This requires a calibrated understanding of the translation and rotation that is undertaken by the object between these views, but the range of motion is generally smaller than when moving imaging optics around the object. In our setup, we can rotate our scene to fill in the density and angular range of the captured cone of the light field. For example, rotating a planar sample would fill in the missing cone above the surface normal of that plane without the need of our added Fresnel lens. To achieve this would require an object mounting apparatus that can perform precise and repeatable rotations about the lower focal point. With full and detailed knowledge of the relative motion between data captures, we could transform all measurements to the same coordinate system and consider the expanded angular range and measurement density that we can measure. While this no longer operate in a single-shot setup, we could gain extremely rich light field measurements with very few measurements.

8.2.8 Transparent Objects

Reconstructing the geometry of transparent objects is a particularly challenging problem [Wu *et al.*, 2018]. Typically, this is achieved by imaging a known target image through the transparent object from multiple viewing directions and trying to map the measured ray being refracted through the object. Our

wide-baseline light field camera provides a diverse set of viewpoints in a single image exposure, which could enhance the capabilities of transparent object reconstruction. By placing target images at either a fixed location or at different orientations and positions behind the object, we can utilize our rich set of observation angles being refracted through the object to reconstruct the 3D geometry.

8.2.9 Dynamic Objects

Since our imaging system captures wide-baseline light field measurements in a single image exposure, we can naturally consider dynamic objects for study. One motivating example of this would be studying iridescent structural coloration that is found in living creatures such as cephalopods [Mäthger *et al.*, 2009]. Rather than requiring multiple images with different perspectives or illumination, during which a living sample would move, we can instead capture a rich set of data in a single exposure without concern of motion.

Another possible application of dynamic capture would be stress-induced birefringence [Hagen *et al.*, 2003]. When certain materials are under stress, they exhibit wavelength-dependent iridescent birefringence based on the stress experienced by different points on the object. In particular, this occurs due to different refraction paths through certain transparent materials under stress when observed in different directions or under different light polarization conditions. This distribution of stress is of concern for manufacturing and designing components for high-stress applications; our camera could provide an efficient and effective way to measure this stress by characterizing the measured birefringence over time as stress is applied to an object. This may require the addition of polarization control to our system, either in measurement or illumination control, which adds another control dimension that we could explore. Expanding the applications of our imaging system to dynamic options could open up useful benefits of our setup in these areas.

8.3 Conclusion

We aim for this thesis to provide a starting point for future work with imaging and applying wide-baseline light fields. We have shown that there is a diverse range of possible uses for such data and have developed an imaging system to capture these light fields. Particularly in a time where virtual reality applications have driven a need for digital reconstructions of real-world objects and scenes, we believe a wide-baseline light field camera can provide efficient and high-quality scans of highly-complex scenes.

Bibliography

Byeongjoo Ahn, Ioannis Gkioulekas, and Aswin C Sankaranarayanan. 2021. Kaleidoscopic structured light. *ACM Trans. Graphics* 40, 6 (2021), 1–15.

Simon Baker and Shree K Nayar. 1998. A theory of catadioptric image formation. In *ICCV*.

Emilio Camahort, Apostolos Lerios, and Donald Fussell. 1998. Uniformly Sampled Light Fields. 117–130. https://doi.org/10.1007/978-3-7091-6453-2_11

Kristin J Dana and Jing Wang. 2004. Device for convenient measurement of spatially varying bidirectional reflectance. *JOSA A* 21, 1 (2004), 1–12.

Donald G Dansereau, Glenn Schuster, Joseph Ford, and Gordon Wetzstein. 2017. A wide-field-of-view monocentric light field camera. In *CVPR*.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the Reflectance Field of a Human Face. In *SIGGRAPH*. 145–156.

Yue Dong. 2019. Deep appearance modeling: A survey. *Visual Informatics* 3, 2 (2019), 59–68. https://doi.org/10.1016/j.visinf.2019.07.003

Jirí Filip, Radomír Vávra, and Michal Havlícek. 2014. Effective Acquisition of Dense Anisotropic BRDF. In *2014 22nd International Conference on Pattern Recognition*. 2047–2052. https://doi.org/10.1109/ICPR.2014.357

Todor G Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree K Nayar, and Chintan Intwala. 2006. Spatio-Angular Resolution Tradeoffs in Integral Photography. *Rendering Techniques* 2006, 263-272 (2006), 21.

Andre Gershun. 1939. The light field. *Journal of Mathematics and Physics* 18, 1-4 (1939), 51–151.

Abhijeet Ghosh, Shruthi Achutha, Wolfgang Heidrich, and Matthew O'Toole. 2007. BRDF acquisition with basis illumination. In *ICCV*.

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)* 30, 6 (2011).

Nathan A. Hagen, Derek S. Sabatke, James F. Scholl, Peter A. Jansson, Weinong Wayne Chen, Eustace L. Dereniak, and David T. Sass. 2003. Compact methods for measuring stress birefringence. In *Polarization Science and Remote Sensing*, Joseph A. Shaw and J. Scott Tyo (Eds.), Vol. 5158. International Society for Optics and Photonics, SPIE, 45 – 53. https://doi.org/10.1117/12.509424

John C Hart. 1996. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer* 12, 10 (1996).

Stefan Heber, Rene Ranftl, and Thomas Pock. 2013. Variational shape from light field. In *Energy Minimization Methods in Computer Vision and Pattern Recognition: 9th International Conference, EMMCVPR 2013, Lund, Sweden, August 19-21, 2013. Proceedings 9*. Springer, 66–79.

Zhuo Hui and Aswin Sankaranarayanan. 2015. A Dictionary-Based Approach for Estimating Shape and Spatially-Varying Reflectance. 1–9. https://doi.org/10.1109/ICCPHOT.2015.7168363

Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Godisart, Bart Nabbe, Iain Matthews, *et al.* 2017. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 41, 1 (2017).

Douglas Lanman, Daniel Crispell, and Gabriel Taubin. 2009. Surround structured lighting: 3-D scanning with orthographic illumination. *Computer Vision and Image Understanding* 113, 11 (2009), 1107–1117.

Scott Lerner and Brett Dahlgrenn. 2006. Etendue and optical system design. *Proceedings of SPIE - The International Society for Optical Engineering* 6338 (08 2006). https://doi.org/10.1117/12.685066

Marc Levoy. 2006. Light Fields and Computational Imaging. *IEEE Computer* 39, 8 (2006), 46–55.

Marc Levoy and Pat Hanrahan. 1996. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 31–42. https://doi.org/10.1145/237170.237199

Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. 2006. Light Field Microscopy. *ACM Trans. Graphics* 25, 3 (jul 2006), 924–934.

Hongsong Li, Sing Foo, Kenneth Torrance, and Stephen Westin. 2006. Automated three-axis gonioreflectometer for computer graphics applications. *Optical Engineering* 45 (04 2006). `https://doi.org/10.1117/12.617589`

Xing Lin, Jiamin Wu, Guoan Zheng, and Qionghai Dai. 2015. Camera array based light field microscopy. *Bio. Op. Exp.* 6, 9 (2015), 3179–3189.

Gabriel Lippmann. 1908. Epreuves reversibles donnant la sensation du relief. *J. Phys. Theor. Appl.* 7, 1 (1908), 821–825.

Alkhazur Manakov, John Restrepo, Oliver Klehm, Ramon Hegedus, Elmar Eisemann, Hans-Peter Seidel, and Ivo Ihrke. 2013. A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM Trans. Graphics* 32, 4 (2013), 47–1.

Stephen Marschner, Stephen Westin, Eric Lafortune, and Kenneth Torrance. 2000. Image-Based Bidirectional Reflectance Distribution Function Measurement. *Applied optics* 39 (07 2000), 2592–600. `https://doi.org/10.1364/AO.39.002592`

Lydia M. Mäthger, Eric James Denton, N. Justin Marshall, and Roger T. Hanlon. 2009. Mechanisms and behavioural functions of structural coloration in cephalopods. *Journal of The Royal Society Interface* 6 (2009), S149 – S163. `https://api.semanticscholar.org/CorpusID:2653639`

Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-Driven Reflectance Model. *ACM Trans. Graphics* 22, 3 (July 2003), 759–769.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*.

Yasuhiro Mukaigawa, Kohei Sumino, and Yasushi Yagi. 2007. High-speed measurement of BRDF using an ellipsoidal mirror and a projector. In *CVPR*.

Yasuhiro Mukaigawa, Kohei Sumino, and Yasushi Yagi. 2009. Rapid BRDF measurement using an ellipsoidal mirror and a projector. *IPSJ Transactions on Computer Vision and Applications* 1 (2009), 21–32.

Ren Ng. 2005. Fourier slice photography. *ACM Trans. Graphics* 24, 3 (jul 2005), 735–744.

Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. 2005. *Light field photography with a hand-held plenoptic camera.* Research Report CSTR 2005-02. Stanford university. Stanford University Computer Science Tech Report pages.

Ilya Reshetouski, Alkhazur Manakov, Hans-Peter Seidel, and Ivo Ihrke. 2011. Three-dimensional kaleidoscopic imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Szymon M Rusinkiewicz. 1998. A new change of variables for efficient BRDF representation. *Rendering techniques* 98 (1998), 11–22.

Lu Sang, Björn Häfner, Xingxing Zuo, and Daniel Cremers. 2023. High-Quality RGB-D Reconstruction via Multi-View Uncalibrated Photometric Stereo and Gradient-SDF. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.* 3106–3115.

Christopher Schwartz, Ralf Sarlette, Michael Weinmann, and Reinhard Klein. 2013. DOME II: a parallelized BTF acquisition system. In *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition.*

Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, Vol. 1. IEEE, 519–528.

Vincent Sitzmann. [n.d.]. Awesome Implicit Representations - A curated list of resources on implicit neural representations. `https://github.com/vsitzmann/awesome-implicit-representations`

Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. 2020. Implicit Neural Representations with Periodic Activation Functions. In *Proc. NeurIPS.*

Rosana Montes Soldado and Carlos Ureña Almagro. 2012. An Overview of BRDF Models. `https://api.semanticscholar.org/CorpusID:44900609`

Yuichi Taguchi, Amit Agrawal, Srikumar Ramalingam, and Ashok Veeraraghavan. 2010a. Axial light field for curved mirrors: Reflect your perspective, widen your view. In *CVPR.*

Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. 2010b. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Trans. Graphics* 29, 6 (2010), 172.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *NeurIPS* (2020).

Michael W Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. 2013. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*. 673–680.

Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, W Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, *et al.* 2022. Advances in neural rendering. In *Computer Graphics Forum*, Vol. 41.

Vaibhav Vaish, Marc Levoy, Richard Szeliski, C Lawrence Zitnick, and Sing Bing Kang. 2006. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *CVPR*.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*.

Tun Wang, Hao Sheng, Rongshan Chen, Da Yang, Zhenglong Cui, Sizhe Wang, Ruixuan Cong, and Mingyuan Zhao. 2024. Light field depth estimation: A comprehensive survey from principles to future. *High-Confidence Computing* 4, 1 (2024), 100187. `https://doi.org/10.1016/j.hcc.2023.100187`

Ting-Chun Wang, Manmohan Chandraker, Alexei A. Efros, and Ravi Ramamoorthi. 2016. SVBRDF-Invariant Shape and Reflectance Estimation From Light-Field Cameras. In *CVPR*.

Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. 2005. High Performance Imaging Using Large Camera Arrays. *ACM Trans. Graphics* 24, 3 (jul 2005), 765–776.

Bojian Wu, Yang Zhou, Yiming Qian, Minglun Cong, and Hui Huang. 2018. Full 3D reconstruction of transparent objects. *ACM Trans. Graph.* 37, 4, Article 103 (jul 2018), 11 pages. `https://doi.org/10.1145/3197517.3201286`

Ruilin Xu, Mohit Gupta, and Shree K Nayar. 2018. Trapping light for time of flight. In *CVPR*.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems* (2021).

Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. In *Neural Information Processing Systems (NeurIPS)*.

Christopher Choy Animashree Anandkumar Minsu Cho Yoonwoo Jeong, Seokjun Ahn and Jaesik Park. 2021. Self-Calibrating Neural Radiance Fields. In *ICCV*.

Hang Zhang, Kristin Dana, and Ko Nishino. 2015. Reflectance hashing for material recognition. In *CVPR*.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.

Remo Ziegler, Simon Bucheli, Lukas Ahrenberg, Marcus Magnor, and Markus Gross. 2007. A Bidirectional Light Field-Hologram Transform. In *Computer Graphics Forum*, Vol. 26. Wiley Online Library, 435–446.