Scanning Iridescent Reflectance

Michael De Zeeuw and Aswin C. Sankaranarayanan

Abstract—Acquisition of high-dimensional spatially-varying bidirectional reflectance distribution functions (SVBRDFs) is extremely challenging for scenes that exhibit high-frequency angular effects—in particular, iridescence—that requires dense sampling across spatial and angular dimensions of reflectance. This is further complicated by the need to illuminate and observe the material along grazing angles, where the effects of Iridescence are often most prominent. This paper proposes an imaging system for acquiring the reflectance associated with such iridescent materials. Our system uses an imaging setup, consisting of an ellipsoidal mirror and a light field camera, that can acquire a dense slice of the SVBRDF across spatial and observation angle axes from a single image, including at grazing observation angles. We use an active illumination setup to sparsely sample the incident illumination angles, thereby providing an acquisition setup that only requires a few (light field) photographs. To compensate for the lack of density in the incident angles, we train a neural model that implicitly represents the SVBRDF, creating a light-weight data-driven reflectance function that enables interpolation of the missing measurements. We show that our imaging system can reconstruct textures that exhibit spatially-varying iridescence stemming from diffraction as well as structural coloration.

Index Terms—Spatially-varying reflectance, Neural rendering, Iridescence, Structural coloration

1 Introduction

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, recent research in reflectance acquisition has revolved around faster data capture techniques and improved representations that allow for both reduced memory and input data footprint. 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 scenes with highly complex visible phenomenon such as iridescence.

In this paper, we propose an imaging system capable of efficiently capturing dense measurements in space, observation angle, and illumination angle paired 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 1. The imaging system is built based on the wide-baseline light field camera introduced by [1], which employs an ellipsoidal mirror and light field camera to capture light field images of scenes over a broad set of angles. Each resulting image captures a large solid cone from each scene point, up to rays at grazing angles to the surface of the scene. They argue that this enables shape and normal estimation using modified shape from focus/defocus algorithms, along with a suite of other classic light field processing results. We apply the wide baseline of these light field images to the purpose of reflectance capture, as it enables the capture of spatial information over a broad range of observation angles in a single exposure, as shown on the left in Figure 1. Pairing this capability with dense illumination control, we are able to efficiently capture measurements over the entire gamut of the SVBRDF. This paper proposes such illumination control, along with optics to extend the range of observation angles, to optimize the wide-baseline light field camera for reflectance capture. In addition, owing to the high dimensionality of our captured data, we develop a neural SVBRDF representation to reconstruct the reflectance of highly complex materials.

• *The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213.*

1.1 Contributions

This paper extends 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 improve upon the capabilities of the WBLF camera by adding illumination control and increasing the range of capture angles to allow for efficient SVBRDF capture.
- *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

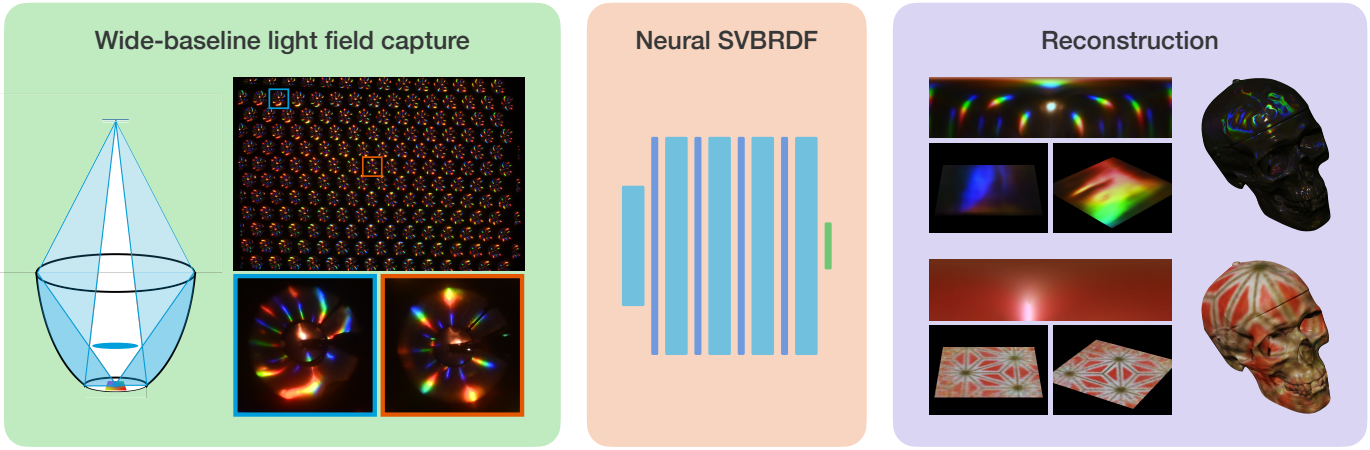

Fig. 1. A summary of the contributions of this paper. We first measure the SVBRDF of various complex materials through wide-baseline light field measurements, optimizing the imaging system for efficient data capture. We then represent this large dataset as a neural SVBRDF, which can be applied to reconstruct the appearance of the captured materials. We show that this system is capable of capturing and reconstructing highly complex and spatially-varying iridescent reflectance.

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. The code and data associated with results can be accessed at the project website [2].

1.2 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 $20\times20\,\mathrm{mm}^2$ 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.

2 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.

2.1 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 [3]. 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 [4], [5], including for anisotropic materials [6]. 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 [7], [8], [9], [10] 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 [11], [12], [13]; 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 [14], [15] and parabolic [16], [17], [18] 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 introduced by De Zeeuw and Sankaranarayanan [1] combines the ellipsoidal mirror with a light field camera to capture a large range of angular and spatial measurements; they apply these measurements to 3D shape reconstruction and other standard light field

processing tasks. This work expands on this design in two distinct ways: first, the incorporation of active illumination to enable reflectance acquisition; and second, the incorporation of additional optical elements to avoid large missing angular cones in the reflected light.

2.2 Neural BRDF Representations

Traditional BRDF representations are usually either model-based [19] or data-driven [4], [20]. 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 [21]. 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 [22], [23] 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. [24] 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. [25] 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. [26] and Wang et al. [27] 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 [24], 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.

3 Imaging Overview

This section outlines the core principles for efficiently measuring the SVBRDF using a wide-baseline light field camera.

3.1 Wide-baseline Light Field Imaging

We utilize the base design proposed by [1] which employs an ellipsoidal mirror to optically map the desired wide-baseline light field at the object to a measurable light field at the sensor. The left side of Figure 2 shows the optical design

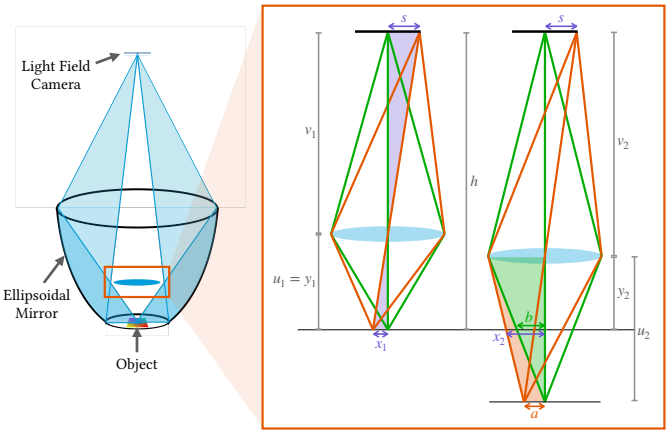

Fig. 2. We use an ellipsoidal mirror to map the light field from the object at one focal point to a light field camera at the other. The light field camera consists of a microlens array mounted in front of an imaging sensor, with an aperture mask aligned to each lenslet to minimize measurement blur. 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.

and how this mapping behaves for a scene point located at the focal point of the ellipsoid. Rays originating at one focal point and reflected off the interior surface will converge at the other focal point. This general focusing applies to points around the focal point as well, though without full convergence. The net effect is the wide-baseline light field near one focal point being mapped to a far narrower light field at the sensor. We measure this mapped light field by placing a microlens array directly over the image sensor. To ensure that each measured pixel corresponds to a single light field ray, an aperture is added over each lenslet. We defer to prior work for a detailed analysis of this design.

3.1.1 *Extending the measurement cone*

This analysis 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 2, 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 subaperture view are shown in Figure 3.

Since the light field camera employs 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 2.

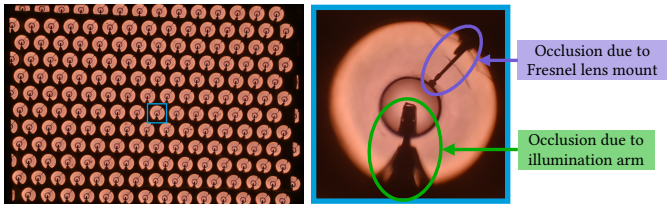

Fig. 3. The Fresnel lens mounting and the illumination arm both introduce occlusion in our measurements. The occlusion from the illumination arm varies as the arm is moved through its range, while the Fresnel lens mounting is fixed. However, this occlusion earns us dense illumination control and many measurements of the central cone.

To get the optimal light collection and limit occlusions, we aim to mount a large-diameter lens in the widest gap between currently measured rays as observed in the figure. This gap occurs at height $73.21\,\mathrm{mm}$ above the object with a width over 5 inches in diameter, beyond the size of most standard lenses. So, we use a Fresnel lens, as the pinhole 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 ($d = 101.6\,\mathrm{mm}$) diameter and a focal length of $71.12\,\mathrm{mm}$.

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 2. Using the focal length equation, we set $f = 71.12\,\mathrm{mm}$, $h = 508.0\,\mathrm{mm}$, and $v_1 = h - u_1$ to find mounting height $y_1 = u_1$ that would bring this lens to focus on the object:

$$y_1 = u_1 = \frac{h - \sqrt{h^2 - 4fh}}{2} = 85.52\ \mathrm{mm} \tag{1}$$

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 $s = 17.5\,\mathrm{mm}$) magnified down to the object plane (producing total width $w_1 = 2x_1$):

$$x_1 = \frac{u_1}{v_1} s, \tag{2}$$

$$w_1 = 2x_1 = 7.08\ \mathrm{mm} \tag{3}$$

This spatial extent is smaller than the extent achieved through the ellipsoidal mirror, where we aim for a target object size of $20{\times}20\,\mathrm{mm}^2$. 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 $y_2 = 73.21\,\mathrm{mm}$, as shown on the right in Figure 2. 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:

$$b = \frac{u_2 - y_2}{u_2}\frac{d}{2}, \quad a = \frac{u_2}{v_2} s, \tag{4}$$

$$x_2 = \frac{y_2}{u_2} a + b = \frac{y_2}{v_2} s + \frac{u_2 - y_2}{u_2}\frac{d}{2}, \tag{5}$$

$$w_2 = 2x_2 = 20.02\ \mathrm{mm}. \tag{6}$$

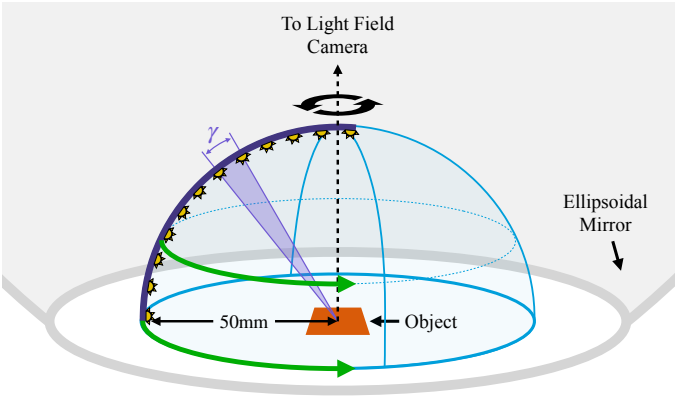

Fig. 4. We mount an LED strip over the object within the ellipsoidal mirror to provide dense illumination control. This provides discrete illumination control in elevation with spacing $\gamma \approx 8°$ and continuous illumination control in azimuth via a rotation stage. The LEDs are $50\,\mathrm{mm}$ away from the sample, so they must be calibrated as nearby point light sources.

This provides a similar spatial extent as the ellipsoidal mapping, which allows us to utilize this new information in a similar way. We move forward with this configuration and use it throughout our experiments.

3.1.2 *Illumination control*

The previous optical design, even considering the new lens addition, is only capable of measuring different observation angles and has no illumination control. Thus, we extend the measuring capabilities by mounting a small strip of LEDs over the object inside of the mirror, as shown in Figure 4. We mount 12 LEDs in a circular arc $50\,\mathrm{mm}$ away from the sample, which gives us $\gamma \approx 8°$ spacing in elevation. The LEDs are mounted on a rotation stage to provide full illumination control in azimuth revolving around the primary optical axis. As with the mounting of the lens, this comes with the cost of a small amount of occlusion, observed in Figure 3. 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. The individual lighting directions for each spatial location can then be computed from these points.

3.2 Analysis of Measured Light Field

We show that the coverage that we can achieve of the desired observation and illumination directions is very near to the desired full-hemisphere coverage. To visualize this, we place a color pattern of size 20mm × 20mm at the focal point of the ellipsoidal mirror. Figure 5 shows a visualization of the resulting light field as observed by the camera. 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

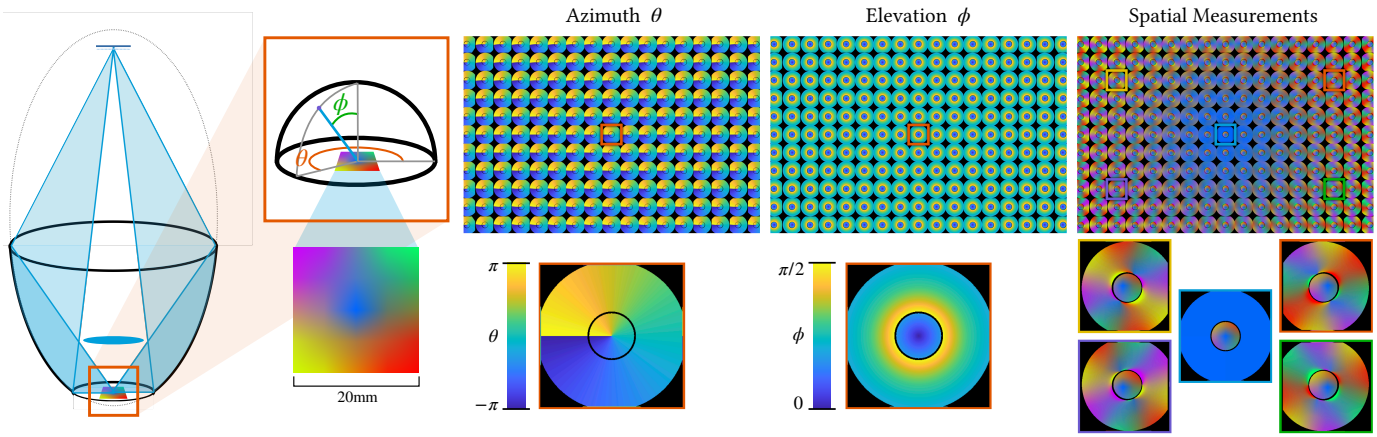

Fig. 5. Visualization of the captured light field. We place a color-coded plane at the object focal point and capture a wide-baseline light field image. The far right column shows this captured image, which highlights the spatial extent of our measurements. The middle two columns visualize the angular 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. 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.

closer to the lower focal point. The range of angles that we observe is large as well, from $0 < \theta < 2\pi$ in azimuth and $0 < \phi < \pi/2$ in elevation. This angular range paired with the spatial mapping indicates a large span of measurements of the object-space light field. The effect of the added fresnel lens is apparent as well, as the central set of pixels in each lenslet image corresponds to the mapping done by the lens.

Additionally, we consider our measurements on a point-by-point basis. Figure 6 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. The darker patches within each 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 cover the entire visible region. We observe that while the measurement distribution changes and reduces as we move away from the focal point, we are still capturing a broad set of angles from each point. This includes measurements through the Fresnel lens, which 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.

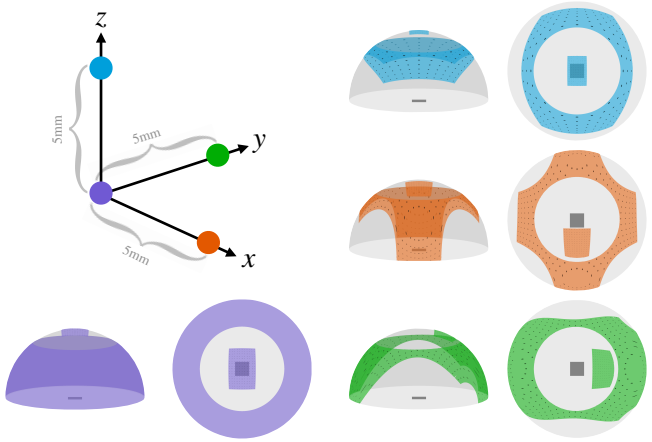

Fig. 6. Given a full hemisphere of possible observation directions from a single point, we visualize the subset of this space that can actually be measured using the imaging system that we propose. We analyze this region for one point at the object focal point and three points shifted by 5mm away from 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.

4 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.

4.1 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. The network consists of four fully-connected layers of size 512, comparable to the rendering network employed in Implicit Differentiable Rendering (IDR) [24]. 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.

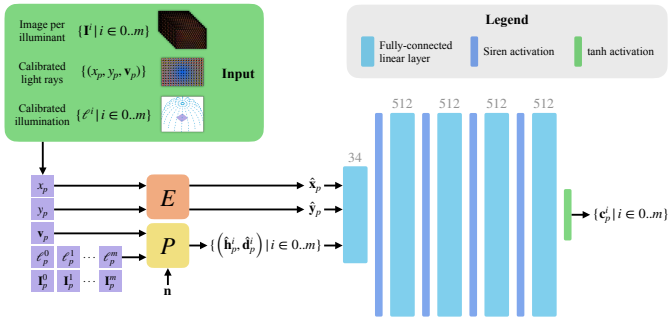

Fig. 7. Overview of the network design to represent the SVBRDF. We take as input the captured image data, one image per illumination direction, along with the calibrated light rays for each light field pixel and the illumination direction for each pixel p and illuminant i. The observation and illumination 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 $[-1, 1]$).

4.2 Encoding High-Frequency Features

The fully-connected network should be able to model the complexity of the desired BRDFs, but as Sitzmann et al. [28] and Tancik et al. [29] 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 [28], 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 6.5 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.

4.3 Input Parameterization

To further improve our ability to model the SVBRDF, we use the half-angle reparameterization of [30], 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 $\mathbf{d}$:

$$\mathbf{h} = \frac{\mathbf{w}_o + \mathbf{w}_i}{||\mathbf{w}_o + \mathbf{w}_i||}, \quad \mathbf{d} = \text{rot}_{\widehat{\mathbf{b}}, -\theta_h} \text{rot}_{\widehat{\mathbf{n}}, -\phi_h} \mathbf{w}_i, \tag{7}$$

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 $\widehat{\mathbf{n}}$).

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:

$$\widehat{\mathbf{h}} = [\sin\theta_h, \cos\theta_h, \sin\phi_h, \cos\phi_h], \tag{8}$$

$$\widehat{\mathbf{d}} = [\sin\theta_d, \cos\theta_d, \sin\phi_d, \cos\phi_d]. \tag{9}$$

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.

5 Implementation

We show the wide-baseline light field camera prototype used in this work in Figure 8. The primary components consist of the ellipsoidal mirror and the light field camera, which is constructed by stacking a microlens array and aperture mask on the sensor of a Nikon Z6 camera. The sensor is placed at the upper focal point of the ellipsoid, with room to place the scene to be imaged at the lower focal point. A 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 in Section 3.1.1, 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.

5.1 Calibration

The need for individually estimated rays for each light field pixel is straightforward in simulation, where the mirror shape and alignment are known. However, practical difficulties in manufacturing ellipsoidal mirrors at the desired scale limit the reliability of its ellipsoidal shape. Therefore, we designed a novel calibration procedure to map directly between light field pixels and object-space light rays.

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

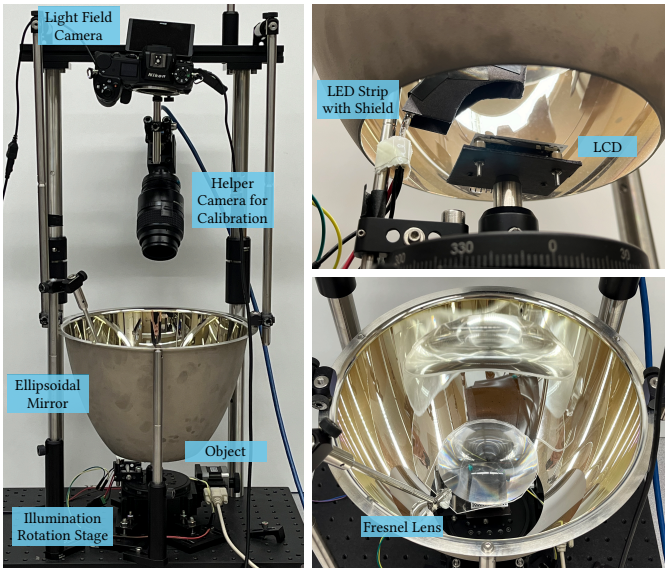

Fig. 8. Optical prototype of the wide-baseline light field camera for measuring SVBRDFs. The ellipsoidal mirror is positioned with the object at the lower focal point and the light field camera at the upper focal point. A calibrated helper camera and LCD aid in calibration and plane detection. On the 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).

determine the location of the screen in camera coordinates using standard techniques. We then display a series of 8-bit graycode patterns on the display, capturing eight images on the light field camera for LCD rows and eight images for columns. Based on the sequence captured by a given light field pixel, we determine the corresponding pixel on the LCD. This, in turn, corresponds to a 3D point defined in camera coordinates. Through shifting the LCD through a series of different orientations and locations, we produce a set of 3D points for each pixel on the light field camera. Since each pixel maps to a ray in object-space, these points should fall along this ray. Fitting a ray to these 3D points produces the desired pixel-to-ray mapping. Given the common presence of outliers introduced by different sources of errors in our measurements, we perform this fitting using an iterative algorithm based on the Huber loss function, which appropriately fits to inlier points.

Once we have calibrated each light field pixel to a corresponding object-space ray, we project these rays to the measurement plane. This plane is determined by once again displaying a checkerboard on the LCD and capturing its location with the helper camera. The materials that we scan are placed on this display so that the object plane is known. The measured rays are propagated 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 BRDF network. The calibrated plane also means the surface normal $\mathbf{n}$ is known and passed to the network as input as well.

The final input required is the illumination direction ℓ for each captured image. Since we are using nearby point light sources, we need to calibrate the precise 3D location of

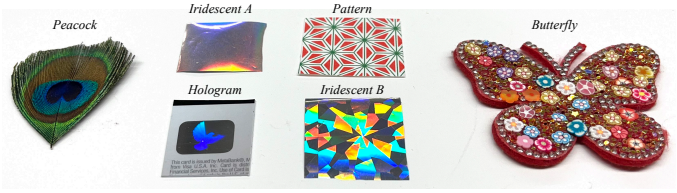

Fig. 9. Objects used for evaluation: peacock feather (*Peacock*), uniform iridescent material (*Iridescent A*), patterned origami paper (*Pattern*), holographic dove from a credit card (*Hologram*), spatially-varying iridescent material (*Iridescent B*), and butterfly hair clip (*Butterfly*).

each LED. We perform this calibration by imaging a reflective sphere of known diameter through the helper camera. The observed reflection of each light source through this sphere provides a ray that observes the desired 3D point. By repeating this process for each light source and with several sphere positions, we produce 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. The illumination direction for each observed ray can be computed by the difference between the intersection point of the observed ray with the object plane and the light source location. Any measured rays occluded by the illumination arm are masked and omitted from training for that particular lighting position.

6 Results

We show a range of results for reflectance with high-frequency variation in space and/or angle using the objects shown in Figure 9. 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.

6.1 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 10 shows these results for five different material captures under three different novel illumination directions. We observe that we capture the different ways that iridescent 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.

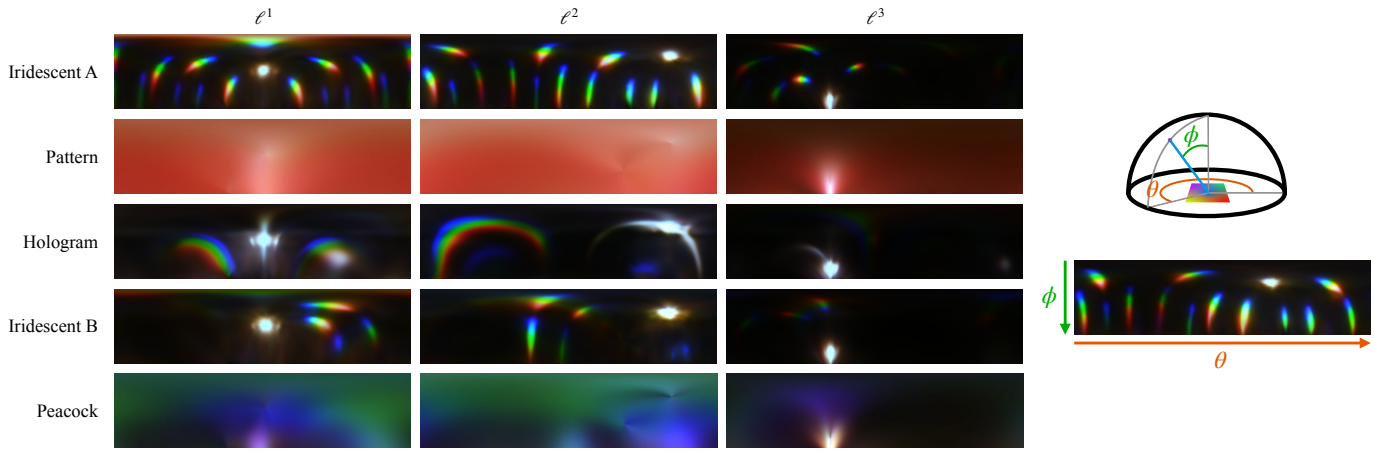

Fig. 10. Visualization of the SVBRDF learned for various input materials. We fix the spatial coordinates at $(x, y) = (0, 0)$ and plot the observed color at viewing direction $\mathbf{v}$ by parameterizing $\mathbf{v}$ according to its azimuth angle θ and elevation angle ϕ, as defined in the illustration on the right. Each column is generated with a different fixed illumination direction ℓ^i.

Fig. 11. Renderings 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.

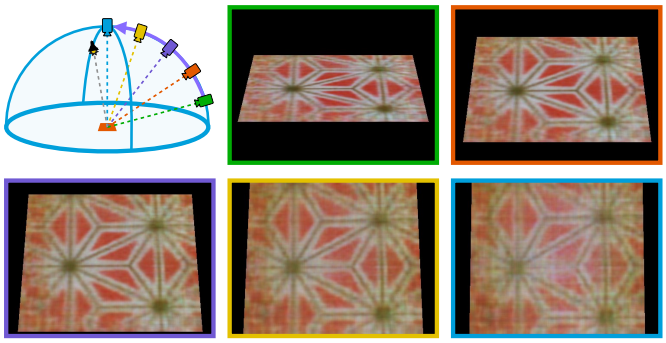

Fig. 12. Renderings with a fixed illumination direction while moving the viewpoint in an arc in elevation (see diagram). We note the increase in artifacts as we move closer to a head-on view; this is due to the relatively limited number of observations from these angles and the optical quality of the Fresnel lens. However, we only achieve these viewpoints at all due to the addition of the Fresnel lens.

6.2 Novel Renderings

We next consider our SVBRDF representation by producing novel renderings of the captured material planes. Figure 17 shows a set of renderings at different viewpoints with two different illumination directions. We render a $20\text{mm} \times 20\text{mm}$ 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.

We further render the *Hologram* under a fixed viewpoint with varying illumination. Figure 11 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 mentioned in Section 3.1.1 when determining the spatial extent. The quality of the reconstruction is also dependent on the observation elevation angle. Figure 12 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 6.5. Overall, the central region reproduces the captured material faithfully and the full area is captured with some artifacts.

6.3 Material Transfer

Finally, we demonstrate the applicability of our captured SVBRDFs by rendering them as BRDF texture maps on different geometry. Figure 13 shows each captured material

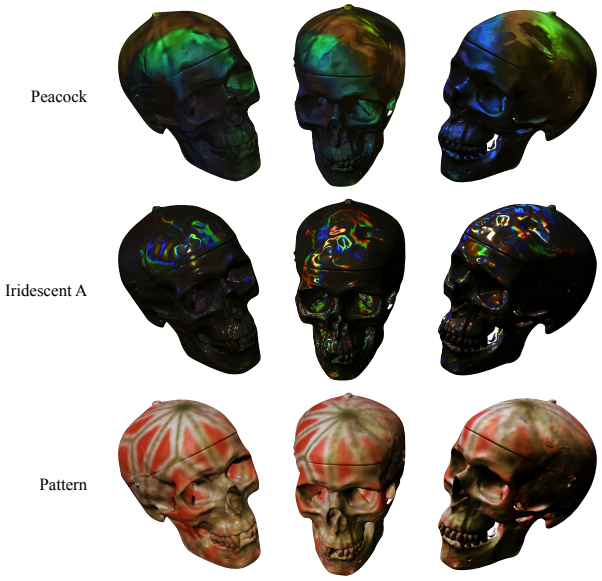

Fig. 13. We render pre-existing geometric models using our captured SVBRDFs as texture maps. This tests our reconstruction over a range of different surface normals and observation angles, and we see both the iridescent effects and the spatial textures successfully rendered.

rendered on the skull geometry from a pre-trained IDR network [24], 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.

6.4 Validation

We validate our learned SVBRDF model by leaving captured illuminations out of training and then reconstructing the original data. Figure 14 shows this process performed with different quantities of training data. We first consider a data set of the full 198 captured images under different illumination conditions. Using 197 of these images, we train the SVBRDF model and apply this model to reconstruct the missing data image. We analyze the reconstruction quality in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

The figure shows this experiment performed with 198, 27, and 9 images, with one image in each case left out for validation. We visualize the reconstructed SVBRDF in the same way as Figure 10 along with two renderings under novel viewpoint and illumination conditions. We further show the ground truth and reconstructed data for the left-out data image, with a close-up provided of one sub-aperture view. This reveals the quality at which we are able to reconstruct the high-frequency features present in this missing data. Even when training with only 26 images, we are able to reconstruct most of the image features present in our measurements. The PSNR and SSIM values back up these observations, as we note high values for each until a drop-off once we utilize only 8 images in training.

6.5 Ablation Study

These results exemplify the capability of our proposed imaging system and neural SVBRDF representation. We consider the impact of particular design elements on our reconstructions, shown in Figure 15.

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 reconstruction 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.

7 Conclusion

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).

7.1 Limitations and Future Work

We conclude by considering the limitations of our approach and outlining avenues for future work. We noted at the start of this paper 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 could overcome limitations in imaging optics.

Our technique is also affected by a loss in spatial resolution that is endemic to light field acquisition using lenslet arrays. We use a sparse pinhole array on the light field camera with a pitch of approximately 2.2mm; one approach to increase resolution to move the camera by small amounts to measure light in the gaps between the pinholes (similar to the Jitter camera [31]). Another approach, borrowing ideas from hybrid imaging [32], is to take a high resolution 2D image of the material and using it to regularize the reconstruction.

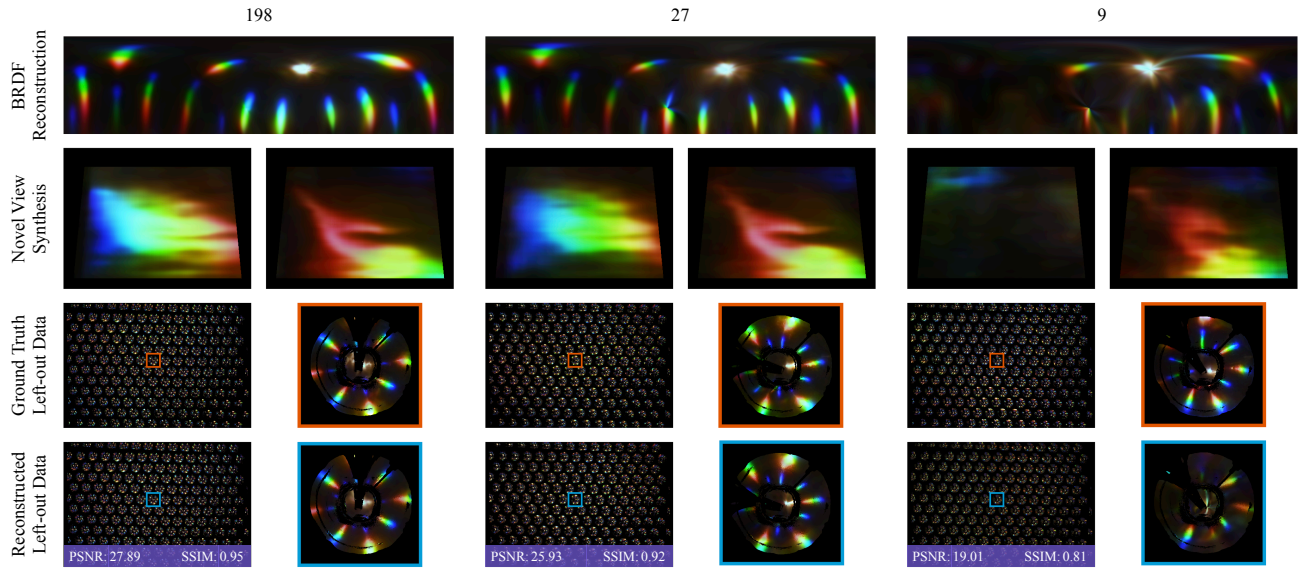

Fig. 14. Validation of the learned SVBRDF model. We train our model with different numbers of illumination conditions: 198 (left), 27 (center), and 9 (right). In each case, we consider illumination conditions from the LED spaced fairly evenly over the hemisphere of observations. For each experiment, we train our model with one data image left out for validation. We visualize the reconstructed SVBRDF (top) and novel view synthesis (second row) along with the ground truth and reconstruction of the left-out data image, with a focus on one sub-aperture view to highlight the differences. We observe high-quality reconstruction with the full dataset, and fairly high-quality reconstruction even when reducing to 27 illumination observations. Our results degrade beyond that, as applying only 9 illumination conditions does not provide the expressiveness that our model requires to interpolate between the known observations. These quality observations are confirmed through peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) calculations between the reconstructed and ground truth left-out data images.

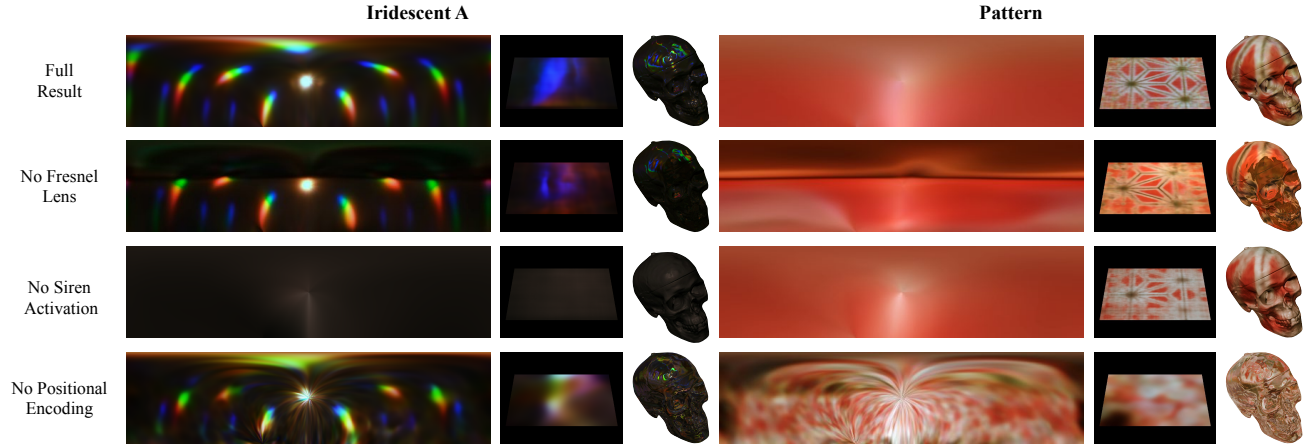

Fig. 15. The impact of major design decisions on reconstruction results. We show a BRDF plot, rendering, and material transfer for two materials (*Iridescent A* and *Pattern*) both with and without the added Fresnel lens, positional encoding, and Siren activation. The addition of the lens fills in the 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.

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 16 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 not on the level with the flowers 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. Future work in this space could develop the ability to model geometry variation along with reflectance for improved modeling of objects like this.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under the grant 2008464, and a Sony Faculty Innovation award.

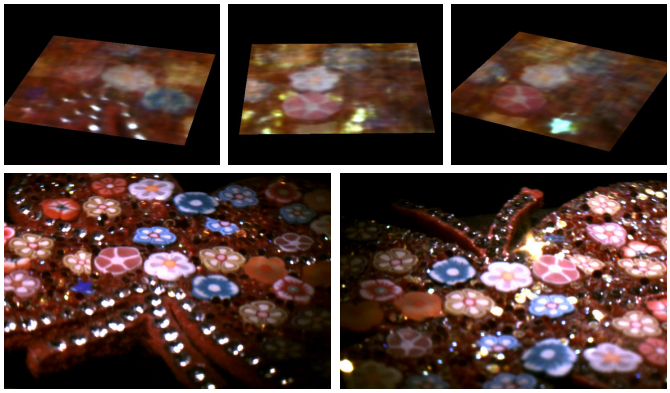

Fig. 16. Renderings of *Butterfly* (top) and reference images of the clip (bottom). Due to the varied surface height of different elements and the limited depth of field of our imaging system, we are unable to resolve each element in focus with our current technique.

REFERENCES

[1] M. De Zeeuw and A. C. Sankaranarayanan, "Wide-baseline light fields using ellipsoidal mirrors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[2] ——, "Project website," https://github.com/Image-Science-Lab-cmu/iridescent_iccp2024.

[3] H. Li, S. Foo, K. Torrance, and S. Westin, "Automated three-axis gonioreflectometer for computer graphics applications," *Optical Engineering*, vol. 45, 04 2006.

[4] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 759–769, Jul. 2003.

[5] S. Marschner, S. Westin, E. Lafortune, and K. Torrance, "Image-based bidirectional reflectance distribution function measurement," *Applied optics*, vol. 39, pp. 2592–600, 07 2000.

[6] J. Filip, R. Vávra, and M. Havlícek, "Effective acquisition of dense anisotropic brdf," in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 2047–2052.

[7] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *SIGGRAPH*, 2000, p. 145–156.

[8] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec, "Multiview face capture using polarized spherical gradient illumination," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, 2011.

[9] C. Schwartz, R. Sarlette, M. Weinmann, and R. Klein, "Dome ii: a parallelized btf acquisition system," in *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*, 2013.

[10] H. Joo, T. Simon, X. Li, H. Liu, L. Tan, L. Gui, S. Banerjee, T. Godisart, B. Nabbe, I. Matthews *et al.*, "Panoptic studio: A massively multiview system for social interaction capture," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 41, no. 1, 2017.

[11] I. Reshetouski, A. Manakov, H.-P. Seidel, and I. Ihrke, "Three-dimensional kaleidoscopic imaging," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[12] R. Xu, M. Gupta, and S. K. Nayar, "Trapping light for time of flight," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018.

[13] B. Ahn, I. Gkioulekas, and A. C. Sankaranarayanan, "Kaleidoscopic structured light," *ACM Trans. Graphics*, vol. 40, no. 6, pp. 1–15, 2021.

[14] Y. Mukaigawa, K. Sumino, and Y. Yagi, "Rapid BRDF measurement using an ellipsoidal mirror and a projector," *IPSJ Transactions on Computer Vision and Applications*, vol. 1, pp. 21–32, 2009.

[15] ——, "High-speed measurement of brdf using an ellipsoidal mirror and a projector," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2007.

[16] H. Zhang, K. Dana, and K. Nishino, "Reflectance hashing for material recognition," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015.

[17] K. J. Dana and J. Wang, "Device for convenient measurement of spatially varying bidirectional reflectance," *JOSA A*, vol. 21, no. 1, pp. 1–12, 2004.

[18] A. Ghosh, S. Achutha, W. Heidrich, and M. O'Toole, "BRDF acquisition with basis illumination," in *ICCV*, 2007.

[19] R. M. Soldado and C. U. Almagro, "An overview of brdf models," Tech. Rep., 2012.

[20] Z. Hui and A. Sankaranarayanan, "A dictionary-based approach for estimating shape and spatially-varying reflectance," in *IEEE Intl. Conf. Computational Photography (ICCP)*, 2015.

[21] Y. Dong, "Deep appearance modeling: A survey," *Visual Informatics*, vol. 3, no. 2, pp. 59–68, 2019.

[22] V. Sitzmann, "Awesome Implicit Representations - A curated list of resources on implicit neural representations." [Online]. Available: https://github.com/vsitzmann/awesome-implicit-representations

[23] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, W. Yifan, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi *et al.*, "Advances in neural rendering," in *Computer Graphics Forum*, vol. 41, no. 2, 2022.

[24] L. Yariv, Y. Kasten, D. Moran, M. Galun, M. Atzmon, B. Ronen, and Y. Lipman, "Multiview neural surface reconstruction by disentangling geometry and appearance," in *Neural Info. Proc. Systems*, 2020.

[25] X. Zhang, P. P. Srinivasan, B. Deng, P. Debevec, W. T. Freeman, and J. T. Barron, "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–18, 2021.

[26] L. Yariv, J. Gu, Y. Kasten, and Y. Lipman, "Volume rendering of neural implicit surfaces," in *Neural Info. Proc. Systems*, 2021.

[27] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," in *Neural Info. Proc. Systems*, 2021.

[28] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Neural Info. Proc. Systems*, 2020.

[29] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," in *Neural Info. Proc. Systems*, 2020.

[30] S. M. Rusinkiewicz, "A new change of variables for efficient brdf representation," *Rendering techniques*, vol. 98, pp. 11–22, 1998.

[31] M. Ben-Ezra, A. Zomet, and S. K. Nayar, "Jitter camera: High resolution video from a low resolution detector," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[32] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *IEEE Intl. Conf. Computational Photography (ICCP)*, 2014.

Michael De Zeeuw received his B.S. in Electrical and Computer Engineering from Calvin University in 2018 and his Ph.D. in ECE at Carnegie Mellon University in 2024. He received the Pittsburgh Chapter Award from the Achievement Rewards for College Scientists Foundation in 2018. His research interests include capturing and analyzing light fields, efficient shape estimation, and reflectance capture and modeling.

Aswin C. Sankaranarayanan is a professor in the ECE department at CMU, where he is the PI of the Image Science Lab. His research interests are broadly in compressive sensing, computational photography, signal processing and machine vision. His doctoral research was in the University of Maryland where his dissertation won the distinguished dissertation award from the ECE department in 2009. Aswin is the recipient of multiple best paper awards including at SIGGRAPH 2023 and CVPR 2019 , the NSF CAREER award, and the Eta Kappa Nu (CMU Chapter) Excellence in Teaching award.

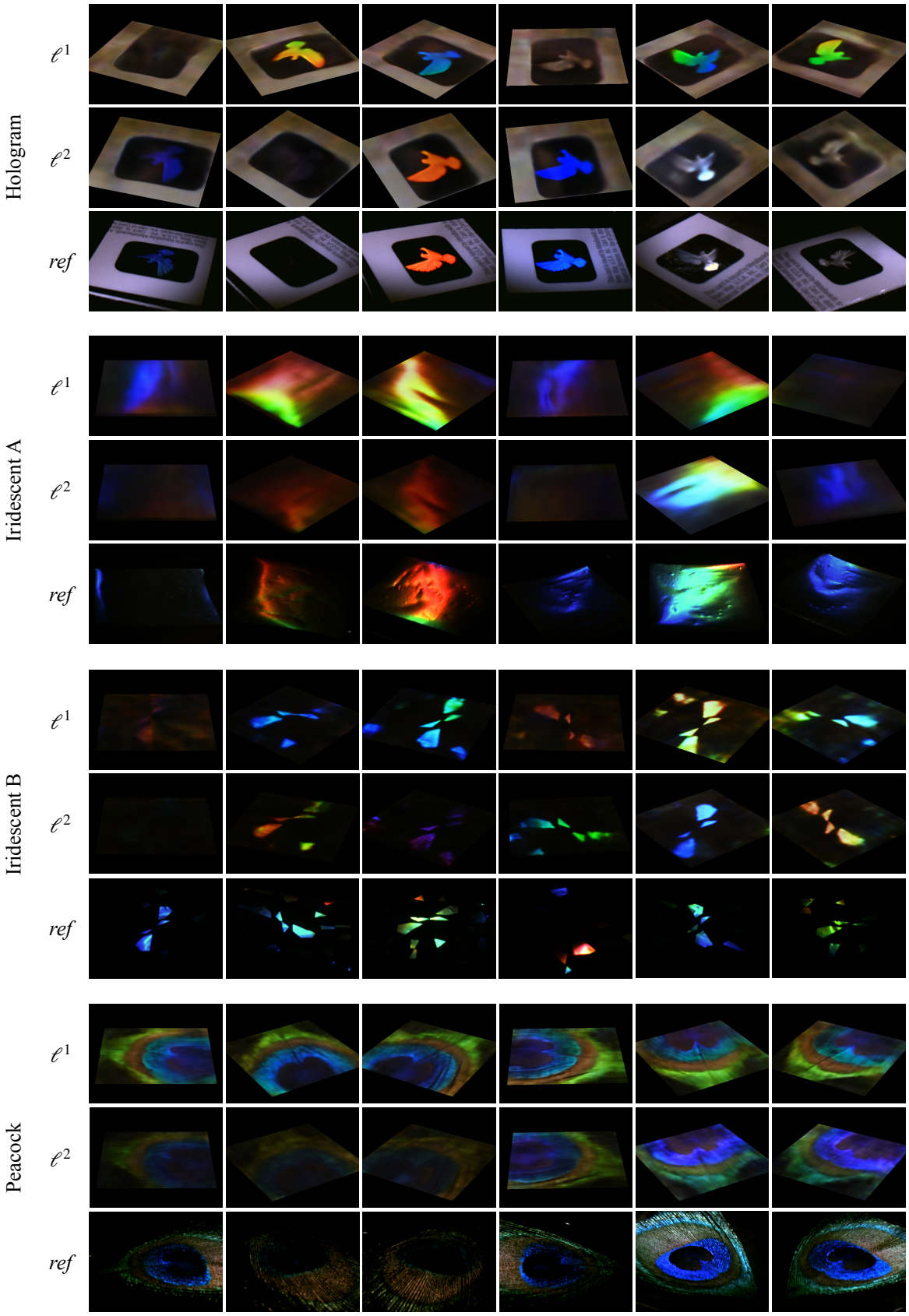

Fig. 17. Renderings of four different material captures under novel viewpoint and illumination conditions. The first and second rows for each object, labelled ℓ_1 and ℓ_2, show two different sets of novel viewpoints under two distinct illumination directions. The third row, labelled *ref*, shows reference 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 for each object view.