ORCa: Glossy Objects as Radiance-Field Cameras

Kushagra Tiwary[1*],   kshat Dave[2*], Nikhil Behari[1], Tzofi Klinghoffer[1],
shok Veeraraghavan[2], Ramesh Raskar[1]

[1]Massachusetts Institute of Technology, [2]Rice University

`ktiw ry,beh ri,tzofi,r sk r}@mit.edu,    ksh t,v shok}@rice.edu`

(a) Scene: Glossy Elephant in a room   (b) Input: images with distorted reflections   (c) Using objects to image beyond camera's field-of-view   (d) Recovered and Undistorted RGBs
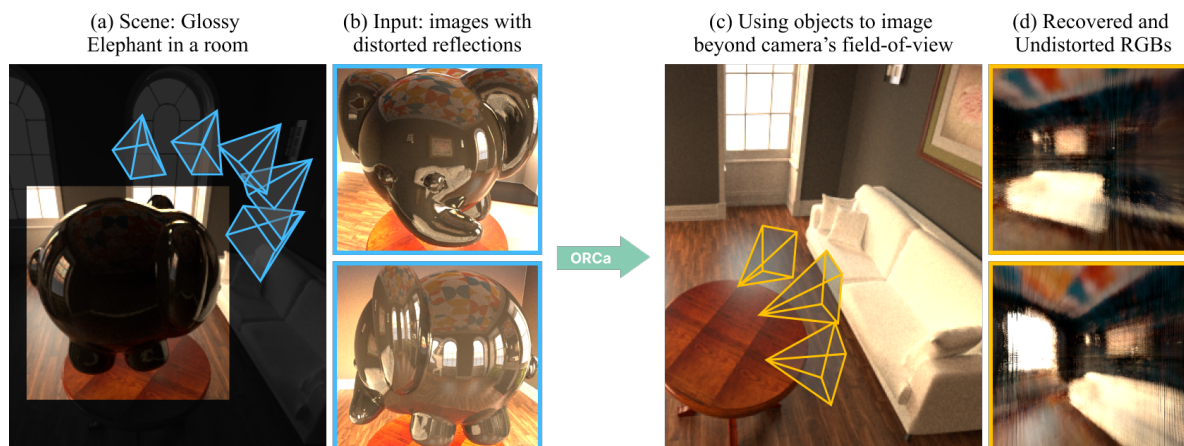
Figure 1. **Objects as radiance-field cameras**. We convert everyday objects with unknown geometry (a) into radiance-field cameras by modeling multi-view reflections (b) as projections of the 5D radiance field of the environment. We convert the object surface into a virtual sensor to capture this radiance field (c), which enables depth and radiance estimation of the surrounding environment. We can then query this radiance field to perform beyond field-of-view novel view synthesis of the environment (d).

bstract

Reflections on glossy objects contain valuable and hidden information about the surrounding environment. By converting these objects into cameras, we can unlock exciting applications, including imaging beyond the camera's field-of-view and from seemingly impossible vantage points, e.g. from reflections on the human eye. However, this task is challenging because reflections depend jointly on object geometry, material properties, the 3D environment, and the observer viewing direction. Our approach converts glossy objects with unknown geometry into radiance-field cameras to image the world from the object's perspective. Our key insight is to convert the object surface into a virtual sensor that captures cast reflections as a 2D projection of the 5D environment radiance field visible to the object. We show that recovering the environment radiance fields enables depth and radiance estimation from the object to its surroundings in addition to beyond field-of-view novel-view synthesis, i.e. rendering of novel views that are only directly-visible to the glossy object present in the scene, but

not the observer. Moreover, using the radiance field we can image around occluders caused by close-by objects in the scene. Our method is trained end-to-end on multi-view images of the object and jointly estimates object geometry, diffuse radiance, and the 5D environment radiance field. For more information, visit our website.

1. Introduction

Imagine that you're driving down a city street that is packed with lines of parked cars on both sides. Inspection of the cars' glass windshields, glossy paint and plastic reveals sharp, but faint and distorted views of the surroundings that might be otherwise hidden from you. Humans can infer depth and semantic cues about the occluded areas in the environment by processing reflections visible on reflective objects, internally decomposing the object geometry and radiance from the specular radiance being reflected onto it. Our aim is to decompose the object from its reflections to "see" the world from the object's perspective, effectively turning the object into a camera which images its environment. However, reflections pose a long-standing challenge

1

(a) Learning Neural Implicit Surfaces [Sec. 3.2]   (b) Objects Surface as Virtual Sensor [Sec. 3.3]   (c) Learning Environment Radiance Fields [Sec. 3.4]
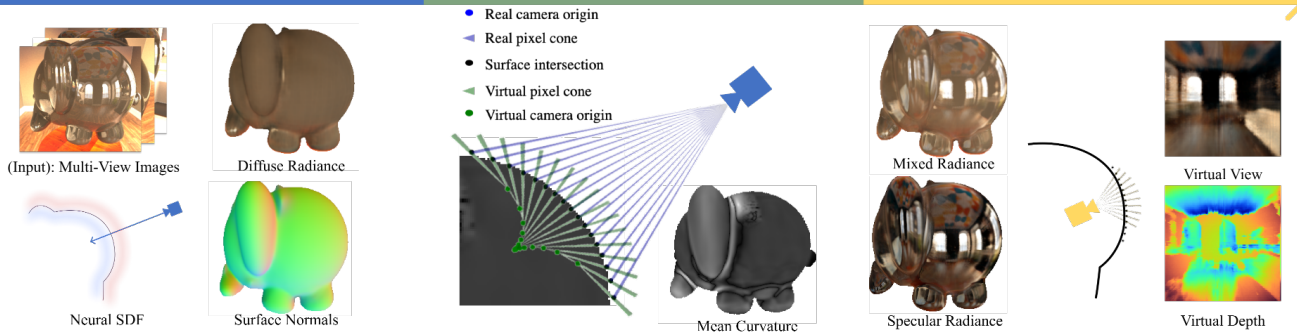
Figure 2. **ORCa Overview**. We jointly estimate the glossy object's geometry and diffuse along with the environment radiance field estimation through a three-step approach. First, we model the object as a neural implicit surface (a). We model the reflections as probing the environment on virtual viewpoints (b) estimated analytically from surface properties. We model the environment as a radiance field queried on these viewpoints (c). Both neural implicit surface and environment radiance field are trained jointly on multi-view images of the object using a photometric loss.

in computer vision as the reflections are a 2D projection of an unknown 3D environment that is distorted based on the shape of the reflector.

To capture the 3D world from the object's perspective, we model the object's surface as a virtual sensor that captures the 2D projection of a 5D environment radiance field surrounding the object. This environment radiance field consists largely of areas only visible to the observer through the object's reflections. Our use of environment radiance fields not only enables depth and radiance estimation from the object to its surroundings, but also enables beyond *field-of-view* novel-view synthesis, i.e. rendering of novel views that are only directly visible to the glossy object present in the scene, but not the observer. Unlike conventional approaches that model the environment as a 2D map, our approach models it as a 5D field without assuming the scene is infinitely far away. Moreover, by sampling the 5D radiance field, instead of a 2D map, we can capture depth and images around occluders, such as close-by objects in the scene, as shown in Fig. 3. These applications cannot be done from a 2D environment map.

We aim to decompose reflections on the object's surface, from its surface and exploit those reflections to construct a radiance field surrounding the object, therefore capturing the 3D world in the process. This is a challenging task because the reflections are extremely sensitive to local object geometry, viewing direction and inter-reflections due to the object's surface. To capture this radiance field, we convert glossy objects with unknown geometry and texture into radiance-field cameras. Specifically, we exploit neural rendering to estimate the local surface of the object viewed from each pixel of the real-camera. We then convert this local surface into a virtual pixel that captures radiance from
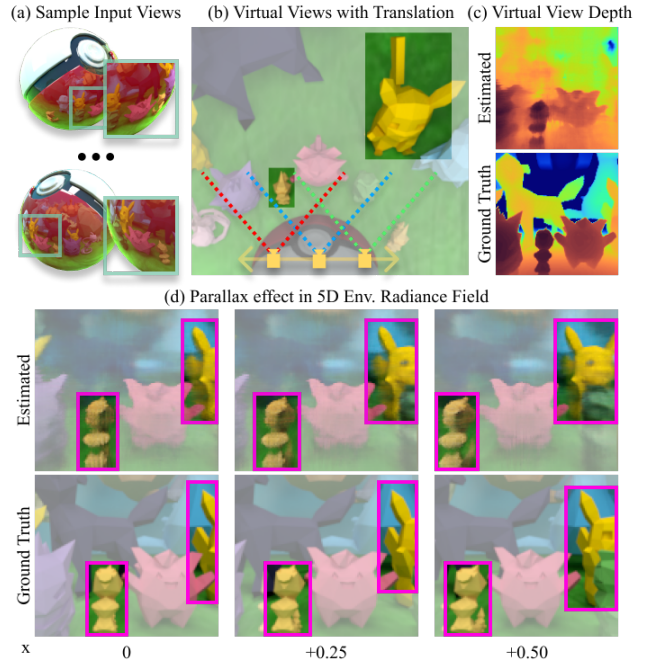

Figure 3. **dvantages of 5D environment radiance field**. Modeling reflections on object surfaces (a) as a 5D env. radiance field enables beyond *field-of-view* novel-view synthesis, including rendering of the environment from translated virtual camera views (b). Depth (c) and environment radiance of translated and parallax views can further enable imaging behind occluders, for example revealing the tails behind the primary Pokemon occluders (d).

the environment. This virtual pixel captures the environment radiance as shown in Fig 5. We estimate the outgoing frustum from the virtual pixel as a cone that samples

2

the scene. By sampling the scene from many virtual pixels on the object surface, we construct an environment radiance field that can be queried independently of the object surface, enabling beyond *field-of-view* novel-view synthesis from previously unsampled viewpoints.

Our approach jointly estimates object geometry, diffuse radiance, and the environment radiance field from multi-view images of glossy objects with unknown geometry and diffuse texture in three steps. First, we use neural signed distance functions (SDF) and an MLP to model the glossy object's geometry as a neural implicit surface and diffuse radiance, respectively, similar to P NDOR [8]. Then, for every pixel on the observer's camera, we estimate the virtual pixels on the object's surface based on the estimated local geometry from the neural SDF. We analytically compute parameters of the virtual cone through the virtual pixel. Lastly, we use the cone formulation in MipNeRF [4] to cast virtual cones from the virtual camera to recover the environment radiance .

To summarize, we make the following contributions:

- We present a method to convert implicit surfaces into virtual sensors that can image their surroundings using virtual cones. (Sec. 3.3)

- We jointly estimate object geometry, diffuse radiance, and estimate the 5D environment radiance field surrounding the object. (Fig. 7 & 8)

- We show that the environment radiance field can be queried to perform *beyond-field-of-view* novel viewpoint synthesis, i.e render views only visible to the object in the scene (Section 3.4)

Scope. We only model glossy objects with low roughness as such specular reflections tend to have a low signal-to-noise ratio, therefore are a blurrier estimate of environment radiance field. However, we note that the virtual cone computation can be extended to model the cone radius as a function of surface roughness. Deblurring approaches can further improve resolution of estimated environment. In addition, we approximate the local curvature using mean curvature, which fails for objects with varying radius of curvature along the tangent space. We explain how our virtual cone curvature estimation can be extended to handle general shape operators in the supplementary material. Lastly, similar to other multi-view approaches, our approach relies on a sufficient virtual baseline between virtual viewpoints to recover the environment radiance field.

2. Related Work

2.1. Modeling reflections

Catadioptric imaging systems aim to expand the field of view of conventional cameras using reflective mirrors [1]

Approach	Input	Scene Geometry	Algorithm Type	Environment Dimension
Neural Illum	Single RGB	Normals	Supervised	2D
Lighthouse	Stereo RGB pair	Multi-plane Image	Supervised	5D
NeRFFactor	Multi-view RGB	Volumetric Albedo	Semi-supervised	2D
RefNeRF	Multi-view RGB	Volumetric Albedo	Self-supervised	2D
PANDORA	Multi-view RGB+Pol.	Neural SDF	Self-supervised	2D
ORCa (Ours)	Multi-view RGB	Neural SDF	Self-supervised	5D

Figure 4. **Comparison of environment estimation approaches**. pproaches such Neural Illum [26], Lighthouse [27], NeRFFactor [37] train on datasets of natural illumination maps to regularize ill-posedness of environment estimation. P NDOR [8] and RefNeRF [31] exploit multi-view reflections on object but approximate surrounding environment is infinitely far away and model it with a flat 2D map. From multi-view reflections, ORCa converts the object surface into virtual sensor and extracts 5D radiance field of the environment.

[19]. Recent work in catadioptric imaging proposes using ellipsoidal mirrors to increase the baseline of a camera, such that more of the light is observed [9] and novel view synthesis from a single capture [34]. These works assume the geometry of the reflecting surface is known or calibrated. In contrast to these methods, we create a catadioptric imaging system from everyday glossy objects of unknown geometry.

Recent progress in neural radiance fields (NeRF) has enabled impressive novel view rendering and geometry reconstruction from multi-view images [17]. NeRF does this by sampling the 5D light field of the scene and learning a representation that is consistent with the training images. MipNeRF [4] demonstrates better novel view synthesis by modeling outgoing rays as cones to enable anti-aliasing. However MipNeRF fails to model sharp view dependencies of reflections. RefNeRF [31] shows improved novel view synthesis on reflections using Integrated-Directional Encoding to explicitly separate diffuse and specular radiance. Similarly, NeRFRN [13] separates diffuse and specular radiance by using separate neural networks. Neural Catacaustics [14] propose a neural warping method to improve novel view synthesis of reflections by learning the caustics of the surface. Comparatively, while all such works improve the quality of novel-view synthesis from the scene to the primary camera, we perform view synthesis that is beyond the line-of-sight of the primary camera, i.e. rendering views only visible to the objects present in the scene, while jointly estimating object geometry and separating diffuse and specular radiance. We perform *beyond line-of-sight* view synthesis by extracting a 5D environment radiance field from the target object.

2.2. Environment Estimation

Recovering underlying scene properties from multiple images is inherently ill-posed [24], but can be regularized using the natural statistics of scene properties as a prior [25] [3]. Recent works exploit this prior through deep neural networks and demonstrate inverse rendering of indoor scenes from a single image [10] [15] [33] [38]. However, these techniques typically recover only coarse representations of lighting and cannot reconstruct fine details of the environment. Lombardi *et al.* [16] recover environment and reflectance, assuming the scene is composed of known geometry and uniform material. Georgolis *et al.* [11] recover the environment map behind the camera from a single image of a glossy object, assuming the object is composed of textureless materials and using ground truth segmentation masks. Song *et al.* [26] estimate plausible environment maps by mapping reflections in the image and inpainting unmapped regions. Srinivasan *et al.* [27] capture stereo image pairs and estimate plausible spatially-coherent environment maps. NeRD [6], NeRFactor [37] and NeuralPIL [7] employ data-driven priors for lighting and BRDF in a NeRF-based approach for radiance decomposition from multi-view images.

While the above approaches, which rely on scene priors, can generate realistic environment maps suitable for virtual object insertion and re-lighting, the actual environment might consist of occlusions. Other imaging modalities and properties of light can aid in extracting information about the surrounding environment. Park *et al.* [23] use RGB-D videos to estimate environment map. Swedish *et al.* [29] recover high-frequency illumination map from the shadows of an object with known geometry. PhySG [36] and Munkberg *et al.* [18] perform inverse rendering from multi-view images by modeling the surface as signed distance functions. P NDOR [8] performs radiance decomposition from polarized RGB images.

3. Learning environment radiance fields from multi-view reflections

3.1. Overview

Reflections on glossy objects offer a glimpse into the surrounding environment beyond the camera's field-of-view. From multi-view images of a glossy object with unknown geometry and albedo, we aim to recover the 5D radiance field of the surrounding environment. The mapping from images captured by the observer to the surrounding environment depends on the glossy object's surface properties, in particular, the surface normals and curvature. We first cast a cone from the observer camera's center-of-projection through each pixel viewing the scene. When the cone intersects the object surface, it reflects, causing the cone to be transformed (Fig. 5). The transformed cone, referred to
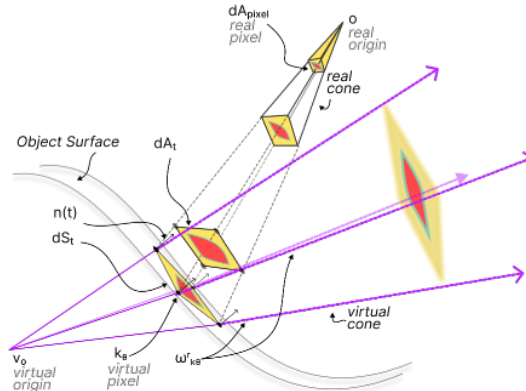

Figure 5. **Virtual Sensor**. We image the world through the object by modeling each pixel's specular radiance as a projection of the 5D radiance field of the environment onto the object's surface. We capture the radiance field by treating the surface area on the object that the pixel views, d_t, as a single-pixel virtual camera with its center-of-projection at v. We cast virtual cones through the virtual sensor to capture the 5D radiance field of the environment.

as a virtual cone, samples the environment and is primarily responsible for the specular radiance observed on the glossy object. Our key insight is that the reflections captured by the observer's camera can be modeled as a projection of the environment radiance field on to the object surface. By modeling the reflected rays as a cone and computing the parameters of the cone, we can more accurately estimate the projected environment radiance field onto the pbject surface, as shown in Fig. 9.

ORCa is composed of three steps: modeling the object's geometry as a neural implicit surface (Sec. 3.2), converting the object's surface into a virtual sensor (Sec. 3.3), and modeling the environment radiance field as a projection along these virtual cones (Sec. 3.4). The learned environment radiance field can then be queried on novel viewpoints to show occluded areas in the scene. Fig. 2 depicts our output for each component on a scene rendered with a complex glossy object and 3D environment. Fig. 6 shows our system architecture. Next, we describe each step in detail.

3.2. Learning Neural implicit Surfaces

Neural Signed Distance Function We model the object geometry as a neural signed distance function (SDF). f : $^3 \rightarrow$. SDFs provide a helpful inductive bias for learning smooth surface geometry [35] [32] [22] that assists downstream tasks in our pipeline. Moreover, the surface properties crucial for our framework, surface normals and curvature, can be conveniently computed from SDFs in a differentiable manner. Consider the 3D spatial coordinates, x, in the scene. The glossy object surface, S is then represented

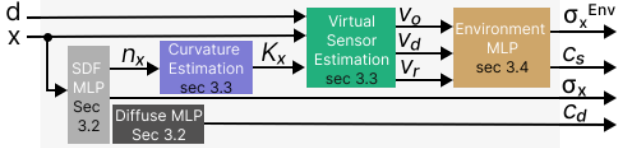Figure 6. **Overview of our proposed architecture**.

by the zero-level set of the SDF

$$\mathcal{S} = \{f_{\mathcal{S}}(\mathbf{x}) = 0 | \mathbf{x} \in {}^3\} \qquad (1)$$

Similar to Yariv *et al.* [35], we model the SDF $f_{\mathcal{S}}$ as a coordinate-based MLP.

Surface Normals Gradients of the SDF at the zero level set point $\mathcal{S}$ towards the surface normals $\mathcal{S}$,

$$\mathbf{n}(\mathbf{x}) = \frac{\nabla_{\mathbf{x}} f_{\mathcal{S}}(\mathbf{x})}{\|\nabla_{\mathbf{x}} f_{\mathcal{S}}(\mathbf{x})\|} \quad \mathbf{x} \in \mathcal{S} \qquad (2)$$

Surface Curvature We employ differential geometry techniques developed by Novello *et al.* [21] to estimate curvature for neural implicit surfaces. In particular, we estimate the mean curvature $K(\mathbf{x})$ for the implicit surface from the divergence, of the surface normals

$$K(\mathbf{x}) = \frac{\cdot \mathbf{n}(\mathbf{x})}{2} \qquad (3)$$

Mean curvature approximates the surface with an osculating sphere. Our approach also works for more generalized notions of curvature through the shape operator, at the cost of higher computational complexity. We refer our readers to the supplement for the general case.

Diffuse Radiance We sepreate the captured radiance at the obsever camera with diffuse radiance, that depends on the glossy object's albedo, and specular radiance that depends on the environment radiance. The diffuse radiance does not have any view dependance and only depends on surface point $\mathbf{x}$. We denote the diffuse radiance as f_d model it using a coordinate-based MLP (Fig. 6).

Volume Rendering s proposed in [35], we perform volumetric rendering on the SDF. We define the volume density $\sigma(\mathbf{x})$ as the cumulative distribution function (CDF), denoted as (s), applied to $f_{\mathcal{S}}$:

$$\sigma(\mathbf{x}) = (f_{\mathcal{S}}) \qquad (4)$$

In contrast to [35], however, we only aim to recover the diffuse radiance of the object along a particular ray. We define a function f_d that estimates the diffuse radiance at each point, $\mathbf{x}$, along the ray. To get the final diffuse radiance along a given primary ray, $\mathbf{r}_p(t)$, we perform volumetric rendering:

$$\hat{\mathbf{c}}_d(\mathbf{r}) = \int f_d(\mathbf{r}(t), f_{\mathcal{S}}^k(\mathbf{r}(t))\tau(t) dt \qquad (5)$$
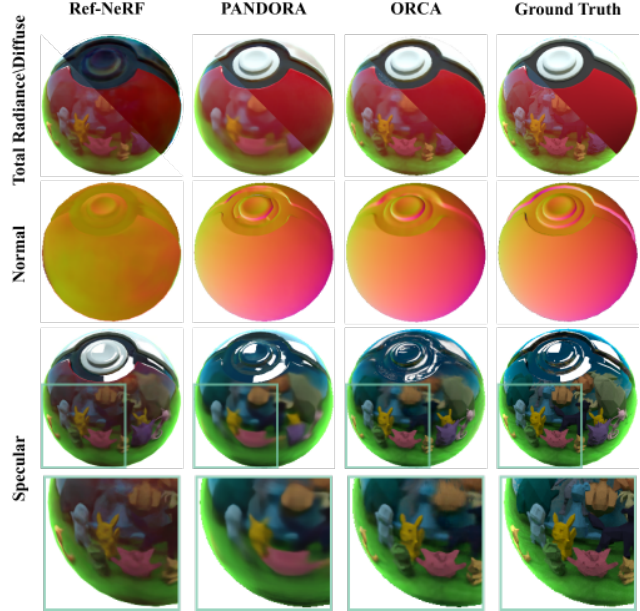

Figure 7. **Qualitative comparisons of diffuse-specular separation and geometry estimation on rendered dataset**. The environment contains nearby objects with complex occlusions when seen through reflections on the glossy object. RefNeRF fails to perform accurate diffuse-specular separation and P NDOR blurs the nearby objects in the specular map. ORC can model the complex specular reflections through environment radiance field.

Note that there is no view dependence in Eq. 5 and intermediate features, $f_{\mathcal{S}}^k$, are used as input. $\tau(t)$ is the accumulated transmittance along the ray.

3.3. Objects Surface as Virtual Sensor

Each pixel, $\mathbf{p}$, with a finite surface area, $\mathbf{d}_p$, on the real-camera sensor views the surface of the object through a frustum originating at that pixel. The object then samples the environment radiance field through this finite surface converting the finite surface into a virtual pixel with surface area, $\mathbf{dS}$. Through this model, we can interpret the object surface as a virtual sensor consisting of many virtual pixels that sample radiance from the environment field based on geometry of the object and observer viewing direction. We now formulate a virtual pixel based on real camera post and implicit surface geometry. Please refer to Fig. 5 for a visualzation of the virtual sensor.

Consider a real camera origin as $\mathbf{o}$ and a pixel on the real sensor $\mathbf{p}_{i,j}$ that corresponds to ray direction $\mathbf{d}$. The primary ray for pixel $\mathbf{p}_{i,j}$ is parameterized with ray length t as $\mathbf{r}_{\mathbf{p}}(t) = \mathbf{o} + t\mathbf{d}$

Casting Real Cones We can approximate the outgoing conical frustum from pixel $\mathbf{p}_{i,j}$ as a cone originating at $\mathbf{o}$ with axis-of-direction $\mathbf{d}$ and radius $\dot{r}$, equivalent to half the distance of the pixel in the x and y directions. We represent the

real-cone as parametric volume,

$$\mathbf{r}_{cone}(\dot{r}, s, \theta) = \dot{r}s\cos(\theta)\hat{\mathbf{e}}_u + \dot{r}s\sin(\theta)\hat{\mathbf{e}}_v + \dot{r}s\mathbf{d}, \quad (6)$$

where $\hat{\mathbf{e}}_u$ and $\hat{\mathbf{e}}_v$ are basis vectors in the plane perpendicular to $\mathbf{d}$, $\theta \in [0, \pi]$ and $s \in [0, t_{max}]$

Virtual Pixel Virtual pixels are characterized by the intersection of the real cone with the object surface. In Sec. 3.2, we model local surface properties using mean curvature which enable efficient analytical computations for the virtual pixel parameters even though our approach works with general shape operators. For a sampled point t_i along the ray, we have the surface normals $\mathbf{n}(t_i)$ from Eq. 2 and estimated mean curvature $K(t_i)$ from Eq. 3. The local object surface at t_i, can be approximated with an osculating sphere, $\mathcal{O}(t_i)$, centered at $\hat{\mathbf{o}}\ (t_i)$ with radius, $R(t_i)$ as follows:

$$R(t_i) = \frac{2}{K_{t_i}}$$
$$\hat{\mathbf{o}}\ (t_i) = \mathbf{r}_p(t_i) + R(t_i) \cdot \hat{\mathbf{n}}(t_i)$$

Note that for concave surfaces $K_{t_i} < 0$, so $\hat{\mathbf{o}}$ will lie outside the object and, for $K_{t_i} > 0$, $\hat{\mathbf{o}}$ will lie inside the object.

The edges of the virtual pixel for $\mathbf{r_p}(t_i)$ would lie at the intersection of the osculating sphere $\mathcal{O}(t_i)$ and the primary cone given by $\mathbf{r}_{cone}$. Computing exact cone-sphere intersections are computationally expensive so we approximate the cone-sphere intersection using rays bound cone-sphere interectional surface $\mathbf{dS}$. We consider four rays that bound the cone and sample them at $\theta_j \in \{0, \pi/2, \pi, 3\pi/2\}$ with Eq. 6. We perform intersections of the corresponding bounding rays with the osculating sphere $\mathcal{O}(t_i)$ to get corners of the virtual pixel $\mathbf{ds}_j$. These ray sphere intersections can be computed analytically in an efficient manner.

Virtual Cone Origin With an estimate of the virtual pixel surface area, we can now compute the virtual cone that samples the environment. We first compute normal vectors at virtual pixel corners $\mathbf{ds_j}$ from the center of osculating sphere $\hat{\mathbf{o}}$

$$\hat{\mathbf{n}_j} = \frac{\mathbf{ds_j}\quad\hat{\mathbf{o}}}{||\mathbf{ds_j}\quad\hat{\mathbf{o}}||} \quad (7)$$

t each virtual pixel corner, we compute the reflected ray directions, ω_j^r, by computing the dot product between the incoming ray directions, ω_k^i, and the normals, $\hat{\mathbf{n}}_k$, where ω^r is the primary ray's reflected vector.

$$\omega^r = \mathbf{d}\quad\mathbf{d} \cdot \hat{\mathbf{n}}(t_i))\hat{\mathbf{n}}(t_i) \quad (8)$$
$$\omega_j^r = \mathbf{d}_j\quad\mathbf{d}_j \cdot \hat{\mathbf{n}}_j(k))\hat{\mathbf{n}}_j(k) \quad (9)$$

$\mathbf{d}_j$ are the incident directions to the virtual pixel corners $\mathbf{ds}_j$. The virtual cone origin is the intersection of these reflected rays at the pixel corners and pixel center. However, these rays might not intersect at a single point so we approximate a virtual origin to be the point that minimizes the sum of distances to the reflected rays ω_j.

$$\mathbf{v_o} = \text{argmin}_{\mathbf{v}} \sum_j |(\mathbf{v}\quad\mathbf{ds_j}) \times \omega_j^r| \quad (10)$$

We pose this as a linear least squares problem and compute the psuedo-inverse to efficiently compute the virtual cone origin.

Virtual Cones Direction. The reflected ray at the center of the virtual pixel reflects the object surface along the direction ω^r from Eq. 8. We consider this as the direction-of-axis of the virtual cone.

$$\hat{\mathbf{v}}_d = \omega^r \quad (11)$$

Virtual Cone Radius. We compute the radius of the cone by treating the reflection vectors of the bounding rays as the neighboring "pixel" directions. Similar to [4], we can compute the distance between $\{\omega_k^r\}_{\theta=}^{2\pi}$ and the primary reflected ray ω^r in the (x, y) components (omitted below for clarity).

$$\hat{\mathbf{v}}_{\dot{r}} = \|\{\omega_k^r\}_{\theta=}^{2\pi}\quad\omega^r\| \quad (12)$$

Finally, for each sampled point t_i, we can characterize our single-pixel virtual sensor located at the object surface $\mathbf{dS}$ as a virtual cone with $\hat{\mathbf{v}}_o$ as its apex, $\hat{\mathbf{v}}_d$ as axis-direction, $\hat{\mathbf{v}}_{\dot{r}}$ as the radius.

Connections to caustics. Our work takes inspiration from Catadiopritc Imaging systems. To covert objects into cameras, we esentially compute the surface and find a corresponding center-of-projection for this surface-as-sensor. However, unlike conventional perspective cameras, objects don't have a fixed center-of-projection, other than in a few special configurations [2], but a locus of viewpoints that vary with object geometry and viewing direction. These viewpoins lie on the "caustic surface" of the object. While typical works in catadioptric imaging use an analytical equation for the caustic surface by assuming known geometry [12] [28], or making assumptions about placement of the observer [30], our formulation approximates the caustic surface of unknonw geometry through intersection of reflected rays on virtual pixels. We emperically show in supplementary that as the surface area of the virtual pixel goes to 0, $d\mathbf{S} \to 0$, our method estimates the true caustic of object without assuming geometry. Our method also has applications in estimating the caustic surface of the unknown geometry.

3.4. Environment Radiance Fields

Our goal is to capture a 5D environment radiance field of the scene by imaging the world through these single-pixel virtual sensors located at the object's surface. We use our formulation of virtual cones to recover 5D environment radiance fields. We define an environment radiance field as $f_{\mathcal{E}} : (\hat{\mathbf{v}}_o, \hat{\mathbf{v}}_d) \to (\sigma^{Env}, c_s)$,

where $f_{\mathcal{E}}$ outputs opacity and radiance along sampled virtual cones. We note that this view dependent radiance is equivalent to the specular radiance at point t_i sampled along the primary-camera ray $\mathbf{r_p}(t)$. We can render the final specular radiance at pixel $\mathbf{p}_{i,j}$ as follows:

$$\hat{\mathbf{c}}_s(\mathbf{r}) = \int f_{\mathcal{E}}(\hat{\mathbf{v}}_o, \hat{\mathbf{v}}_d)\tau(t)dt$$

$$\hat{\mathbf{c}} = \hat{\mathbf{c_d}} + \hat{\mathbf{c_s}}$$

Intuitively, $f_{\mathcal{E}}$ learns the 5D radiance field by sampling single-pixel virtual sensors from the object surface area, and must learn geometry and environment radiance that is consistent with multiple views from the object's reflections. Moreover, we can query $f_{\mathcal{E}}$ to render novel viewpoints and associated depths that are beyond field-of-view of the real camera. We volume render each virtual cones by dividing them into conical frustums using Integrated-Positional Encoding as proposed in MipNeRF [4]. Our formulation of virtual cones works well with Mip-Nerf's rays-as-cones method.

4. Experiments

Our experiments study the ability of our method to recover 5D environment radiance fields (assessed through quality of predicted surface normals, diffuse radiance, specular radiance, and 3D environment maps) from objects of varying complexity, both in simulation (Fig. 7) and the real-world (Fig. 8). Quantative results are provided in Table [1].

s in prior works on novel view synthesis, we report PSNR and SSIM to evaluate estimated diffuse, specular, and mixed radiance, and report mean angular error (M E) to evaluate estimated surface normals.

4.1. Implementation Details

s in P NDOR , we parameterize $f_{\mathcal{S}}$ with an 8-layer MLP to estimate the surface, and, as in MipNeRF, f_d with 4-layer MLP with input geometric features of size 512 from $f_{\mathcal{S}}$. We follow the sdf-to-opacity conversion and the iterative sampling of the ray proposed in [35]. To aid the network to learn the geometry quickly, we also train $f_{\mathcal{S}}$ with a mask-net as proposed in [8]. We use five losses in our architecture: photometric loss, mask loss [8], normal loss [31], eikonal loss [35], and distortion loss [5]. dditional training details are discussed in the supplementary materials.

| Scene | pproach | Diffuse Radiance | | Specular Radiance | | Mixed Radiance | | Normals |
		PSNR ↑ (dB)	SSIM ↑	PSNR ↑ (dB)	SSIM ↑	PSNR ↑ (dB)	SSIM ↑	M E ↓ (°)
D1	Ref-NeRF	**17.59**	**0.7217**	14.88	0.4750	**19.58**	**0.7956**	62.45
	P NDOR	13.23	0.4759	15.12	**0.5231**	12.87	0.4607	2.387
	ORC	13.29	0.4683	**16.64**	0.5148	18.23	0.5745	**1.873**
D2	Ref-NeRF	11.86	0.6090	15.28	**0.7059**	21.80	**0.8643**	33.92
	P NDOR	22.53	0.8689	17.76	0.6326	**22.73**	0.7787	3.693
	ORC	**23.47**	**0.8954**	**18.98**	0.6954	22.31	0.8107	**3.568**

Table 1. **Quantitative evaluation of rendered scenes**. We compare ORCa to other neural rendering techniques that model reflections, including Ref-NeRF and P NDOR , on the globe (D1) and Pokemon (D2) datasets. ORCa is competitive with the comparison methods in accurate diffuse and specular separation, and provides consistent improvement in geometry and specular radiance estimation.

4.2. Datasets

We conduct experiments on both simulated and real-world datasets. Simulated datasets are rendered in Mitsuba2 [20]. Simulated datasets contain a range of increasingly complex object geometries (elephant, Pokeball, and orca) and scenes (living room and Pokemon). We train with 200 views for simulated datasets. We also show results for a real-world dataset [8] capturing a glossy cup with a black vase sitting atop it using 35 views. ll datasets will be publically released upon publication.

4.3. Comparisons with Baselines

We compare our method to other neural rendering techniques that model reflections, Ref-NeRF and P NDOR .

We first discuss results in Fig. 3 which show the advantages of recovering a 5D radiance field with close-by objects as they often cause occlusions which cannot be modeled by 2D environment maps. By estimating the radiance field, we can image behind occluders through sampling novel viewpoints such as the translated viewpoints shown in Fig. 3. Moreover, we can also show depth to surroundings from these vriutal viewpoints. We provide additional examples of depth and beyond *field-of-view* novel-view synthsis in the supplementary materials.

While Ref-NeRF and P NDOR learn 2D environment radiance fields, ORCa recovers a 5D environment radiance field. s shown in Fig. 7 and 8, ORCa estimates more accurate surface normals than other methods. While the total radiance predicted by Ref-NeRF and P NDOR are visually similar, the surface normals are less smooth than ORCa. We also observe that ORCa is able to achieve better diffuse and specular radiance separation than P NDOR , which is evident in the Pokeball surface normals Fig. 7. In these examples, P NDOR recovers blurry specular radiance. We see that ORCa's predicted depth is highly interpretable and matches the underlying geometry of the environment, as shown in Fig. 3. Even on cylinderical real-world datasets, such as the black vase in Fig. 8, the nearby hallway is vis-
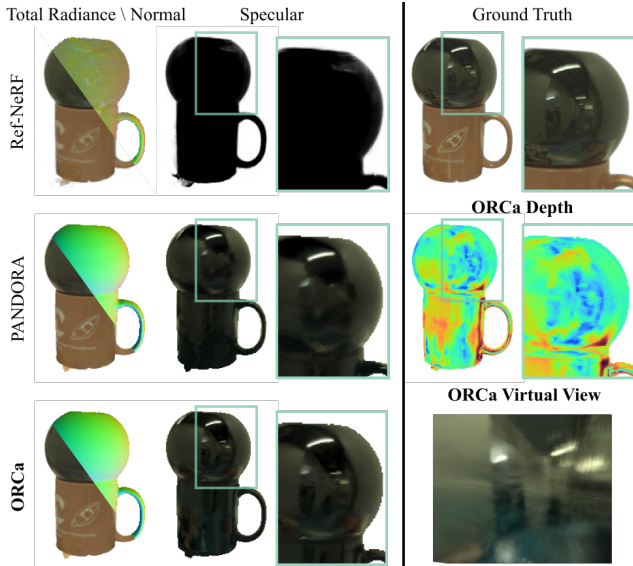
Figure 8. **Comparisons on real dataset**. Using a real-world dataset with only 35 views, ORCa can model the sharp specularities on the ball arriving from regions of the nearby scene, such as the table, and far away scene regions, such as the hallway, to learn an environment radiance field. We query the radiance field for depth of the far hallway (blue) and the nearby objects, such as the table (red). We also render novel viewpoints that are beyond the field-of-view of the observer camera and show that ORCa interpolates well between those views.

Figure 9. **blation on virtual cone formulation**. We study the impact of our proposed virtual cone formulation for recovering the 5D radiance fields compared to a naive formulation where the emitted cones' origins are placed on the object surface. In contrast, we place the emitted cones' origins within the object based on caustics and compute the cone radius. We see that this leads to smoother estimated surface normals (left) and more accurate estimated specular radiance (right).

ible in both the virtual view and depth, despite never being in the field of view of the primary camera. Unlike Ref-NeRF, our primary objective is not to perform novel-view synthesis, but instead to capture the environment radiance field from the object surface.

s shown in Table 1, ORCa is competitive with both Ref-NeRF and P NDOR in estimating diffuse radiance, specular radiance, mixed radiance, and normals. lthough the comparison methods slightly outperform ORCa in full, mixed-radiance scene rendering, ORCa consistently provides better specular radiance and object geometry estimation across scenes and viewpoints. This is again indicative of a key strength of ORCa; whereas existing approaches aim to perform novel-view synthesis on reflective objects, ORCa specifically focuses on accurate specular reflection retrieval for environment radiance field modeling. This is achieved through accurate object geometry modeling, which enables high-accuracy specular radiance estimation, thereby aiding in beyond *field-of-view* novel-view synthesis.

4.4. Impact of Correct Virtual Cones

We base our method on a physically accurate formulation by modeling ray-cone intersections and using the surface as a virtual sensor, as described in Sec 3.3. Naively, the
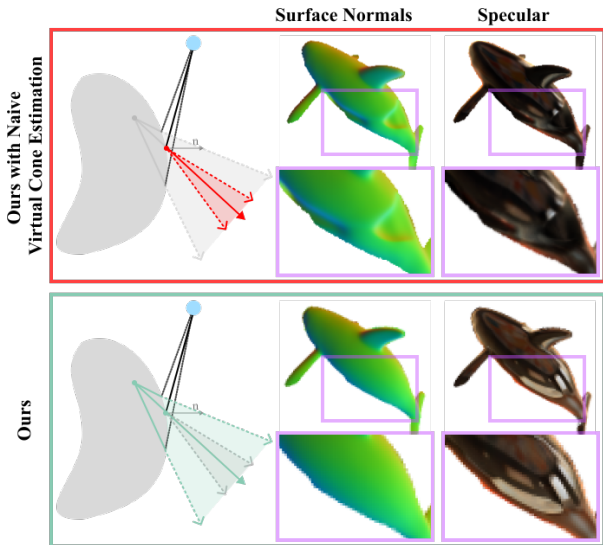
origins of the virtual cones could instead be placed at the intersection of the primary camera ray and surface, in essence placing a Mip-NeRF at each intersection point. This alternative formulation would not be physically accurate and Fig. 9 shows the improvement that we achieve, underscoring the importance of correctly modeling virtual cones.

5. Conclusion

In conclusion, we present a method to convert glossy objects with unknown geometry and texture into radiance-field cameras that capture the environment radiance field around them. Our method recovers object geometry and diffuse radiance, in addition to capturing the depth and radiance of the object's surroundings from its perspective. Our modeling of environment as a radiance field is effective in recovering close-by objects (Fig 7), in addition to being occlusion aware (Fig 3). Moreover, by recovering the environment radiance field we can perform beyond *field-of-view* novel-view synthesis. Our work can unleash applications in virtual object insertion and 3D perception, e.g. inferring information beyond the line-of-sight of the camera using predicted virtual views and depth.

Our formulation of the radiance field beyond the conventional direct-line-of-sight radiance field can enable further areas of research that aim to to extract more information from the environment and the objects present in it.

References

[1] S. Baker and S.K. Nayar. theory of catadioptric image formation. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 35–42, 1998. 3

[2] Simon Baker and Shree K. Nayar. theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35:175–196, 2004. 6

[3] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. 4

[4] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: multiscale representation for anti-aliasing neural radiance fields. *ICCV*, 2021. 3, 6, 7

[5] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 7

[6] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021. 4

[7] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *dvances in Neural Information Processing Systems*, 34:10691–10704, 2021. 4

[8] kshat Dave, Yongyi Zhao, and shok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. *arXiv preprint arXiv:2203.13458*, 2022. 3, 4, 7

[9] Michael De Zeeuw and swin C Sankaranarayanan. Wide-baseline light fields using ellipsoidal mirrors. *IEEE Transactions on Pattern nalysis and Machine Intelligence*, 2022. 3

[10] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6908–6917, 2019. 4

[11] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Tinne Tuytelaars, and Luc Van Gool. What is around the camera? In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5170–5178, 2017. 4

[12] J. Gluckman and S.K. Nayar. Planar catadioptric stereo: geometry and calibration. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 22–28 Vol. 1, 1999. 6

[13] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18409–18418, June 2022. 3

[14] Georgios Kopanas, Thomas Leimkühler, Gilles Rainer, Clément Jambon, and George Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *CM Transactions on Graphics*, 41(6): rticle–201, 2022. 3

[15] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2475–2484, 2020. 4

[16] Stephen Lombardi and Ko Nishino. Reflectance and natural illumination from a single image. In *European Conference on Computer Vision*, pages 582–595. Springer, 2012. 4

[17] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the CM*, 65(1):99–106, 2021. 3

[18] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, lex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8280–8290, 2022. 4

[19] Shree K Nayar and Simon Baker. Catadioptric image formation. In *Proceedings of the 1997 D RP Image Understanding Workshop*, pages 1431–1437, 1997. 3

[20] Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: retargetable forward and inverse renderer. *CM Transactions on Graphics (TOG)*, 38(6):1–17, 2019. 7

[21] Tiago Novello, Guilherme Schardong, Luiz Schirmer, Vinicius da Silva, Helio Lopes, and Luiz Velho. Exploring differential geometry in neural implicits, 2022. 5

[22] Michael Oechsle, Songyou Peng, and ndreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021. 4

[23] Jeong Joon Park, leksander Holynski, and Steven M Seitz. Seeing the world in a bag of chips. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1417–1427, 2020. 4

[24] Ravi Ramamoorthi and Pat Hanrahan. signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 117–128, 2001. 4

[25] Fabiano Romeiro and Todd Zickler. Blind reflectometry. In *European conference on computer vision*, pages 45–58. Springer, 2010. 4

[26] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6918–6926, 2019. 3, 4

[27] Pratul P Srinivasan, Ben Mildenhall, Matthew Tancik, Jonathan T Barron, Richard Tucker, and Noah Snavely. Lighthouse: Predicting lighting volumes for spatially-

coherent illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8080–8089, 2020. 3, 4

[28] R. Swaminathan, M.D. Grossberg, and S.K. Nayar. Caustics of catadioptric cameras. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 2–9 vol.2, 2001. 6

[29] Tristan Swedish, Connor Henley, and Ramesh Raskar. Objects as cameras: Estimating high-frequency illumination from shadows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2593–2602, 2021. 4

[30] Yuichi Taguchi, mit grawal, shok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. xial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *CM Trans. Graph.*, 29(6), dec 2010. 6

[31] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. *CVPR*, 2022. 3, 7

[32] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 4

[33] Zian Wang, Jonah Philion, Sanja Fidler, and Jan Kautz. Learning indoor inverse rendering with 3d spatially-varying lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12538–12547, 2021. 4

[34] Ziyu Wang, Liao Wang, Fuqiang Zhao, Minye Wu, Lan Xu, and Jingyi Yu. Mirrornerf: One-shot neural portrait radiance field from multi-mirror catadioptric imaging. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2021. 3

[35] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. 4, 5, 7

[36] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021. 4

[37] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *CM Transactions on Graphics (TOG)*, 40(6):1–18, 2021. 3, 4

[38] Rui Zhu, Zhengqin Li, Janarbek Matai, Fatih Porikli, and Manmohan Chandraker. Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2822–2831, 2022. 4