Reconstructing Realistic and Relightable Eyes

Wesley Khademi
Oregon State University
Corvallis, OR, USA
khademiw@oregonstate.edu

Jogendra Kundu
Meta
Menlo Park, CA, USA
jogendrak@meta.com

Yatong An
Meta
Menlo Park, CA, USA
yatong@meta.com

Alexander Fix
Meta
Menlo Park, CA, USA
alexander.fix@meta.com

David Colmenares
Meta
Menlo Park, CA, USA
dcol@meta.com

Abstract

Accurately modeling the eye is a challenging task as it exhibits refraction and reflection at the cornea, complex iris texture, and self-occlusion and shadowing due to eyelids and eyelashes. To address these challenges, we present a system for learning a hybrid relightable eye model which can be relit under near-field point lights. Our hybrid model leverages an eyeball mesh for explicitly representing the cornea surface and the reflections on it while learning the geometry and light transport of the periocular region and eye interior implicitly. To account for refraction, we explicitly handle the refraction of camera rays using Snell's law and predict the refraction of incident light rays using a neural network. Furthermore, we propose an extension of our method which enables us to relight the eye using a fringe projector to simulate structured light. Through experiments, we demonstrate that our method results in higher fidelity rendering under novel viewpoint and lighting conditions, improves learned iris geometry, and more accurately simulates structured light fringe patterns on the eye.

1. Introduction

Eyes convey a wealth of information, providing cues about our attention, intentions, and emotions. These cues are critical in Augmented and Virtual Reality (AR/VR) as we aim to enable natural social interaction through virtual avatars and to understand where a human's attention is focused in a scene through eye tracking. Because of this, building a model of the eye and periocular region which can be rendered has the potential to bring significant benefits to creating digital humans [24, 46] or generating synthetic data to train machine learning based eye tracking methods on [27, 50, 58, 61]. However, modeling the geometry and appearance of the eye presents many unique challenges. The transparent cornea causes both refraction and reflection, the iris exhibits fine-grained texture which can appear distorted due to refraction, and self-occlusion and shadowing can occur due to eyelids and eyelashes.

Despite the importance of eyes, most of the computer graphics literature has adopted overly simplistic representations, such as using a sphere-on-sphere model for the eye and a planar disk or cone to represent the iris [30, 44]. Bérard et al. [4] was one of the first works to propose a way to recover the shape and texture of all visible parts of the eye (i.e., cornea, sclera, iris) from multi-view imagery, but did not reconstruct the periocular region such as eyelids and eyelashes. Schwartz et al. [48] later proposed a way to model the entire face by registering an eyeball model into a face mesh, where both the geometry and texture of the eye and face were optimized through differentiable rendering. However, the eye appearance of their method lacks photorealism and cannot be relit.

In recent years, Neural Radiance Fields (NeRF) [39] has emerged as a promising technique for rendering photorealistic images. NeRF works by learning a continuous volumetric representation of a scene's geometry and appearance through differentiable volume rendering. However, a major limitation of NeRF is that it does not explicitly model how light interacts with an object's geometry and material properties, preventing the ability to relight the scene under novel illumination. Because of this, many efforts have been made to enable relighting with NeRF through inverse rendering [6, 11, 12, 25, 51] or by learning to model the light transport function [47, 54, 62]. In relation to modeling and relighting the eye, EyeNeRF [31] introduced a hybrid representation where an explicit mesh model is used to represent the eyeball surface while a NeRF model represents the periocular region and eye interior. There representation enabled

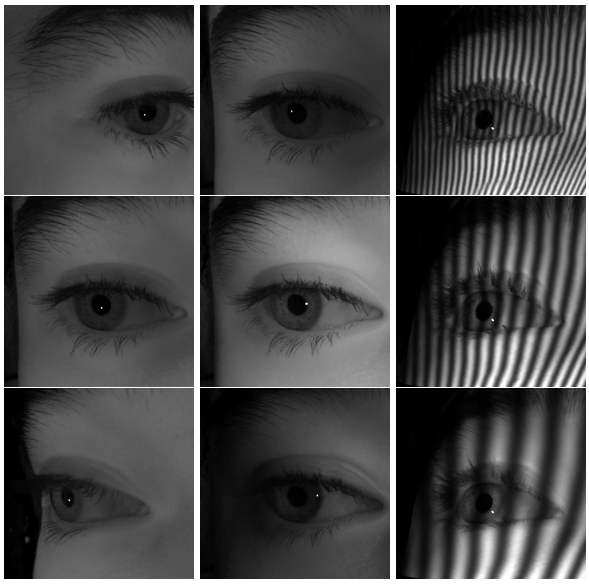


Figure 1. Our model can render the eye under novel viewpoint, relight the eye under arbitrary point light sources, and simulate structured light fringe patterns on the eye.

them to account for some of the challenges of the eye like refraction and reflection at the cornea, while also supporting relighting using environment maps.

Our work shares many similarities to EyeNeRF [31]. However, instead of relighting using an environment map, we focus on relighting under two commonly used illumination sources in eye tracking: near-field point lights and fringe projectors. These two illumination sources explore different avenues for eye tracking. Near-field point light sources provide sparse surface information by producing sparse specular reflections on the cornea, known as glints, which can be used alongside other eye features such as the pupil to recover gaze [16, 21, 38]. Alternatively, fringe projection-based structured light provides dense depth information about eye features such as the sclera and iris, which can be used to enrich traditional computer vision methods and machine learning methods for eye tracking [57, 59, 70]. By supporting relighting under both illumination sources, our method provides a viable path forward for generating synthetic data for eye tracking exploration.

To accomplish building a model of the eye that captures all of its intricacies and allows for relighting, we design a hybrid relightable model that can be learned from a light stage consisting of a sparse of set of cameras and point lights. Inspired by EyeNeRF [31], our method leverages an explicit eyeball mesh for representing the surface of the cornea while implicitly learning the periocular region and eye interior as a NeRF. For relighting the eye interior and face, our method learns a light transport field which describes the relationship between incoming and outgoing light at a point in space. This allows us to approximate complex lighting effects (e.g., direct reflections, interreflections, self-shadowing) learned from data captured by our light stage instead of performing expensive simulation of light transport. To handle accurately relighting the eye interior, we must also account for light ray refraction. Finally, we extend the rendering equation from handling point lights to relighting using fringe projectors. We demonstrate that our approach models the appearance of the eye with high fidelity and allows for relighting using either point light sources or fringe projectors.

In summary, our contributions are:

- A hybrid relightable eye model that handles both refraction and reflection and can be relit under near-field illumination point light sources.
- We propose a way to recover the direct illumination ray in the presence of refraction, which is needed to accurately model both light transport and structured light on surfaces lying behind the refractive cornea.
- We introduce an extension of our method that enables us to relight the eye using a fringe projector.

2. Related Work

2.1. Image-based relighting

Image-based relighting aims to model the complex lighting effects present within a scene without the need for performing an explicit simulation of light transport. Early work by Debevec et al. [17] showed that it was possible to model the light transport function under distant illumination from a dense set of OLAT images. Using their reconstructed light transport function and the fact that light is additive, they were able to relight a subject's face under novel illumination by interpolating these dense samples. In more recent years, the use of neural networks have been explored to model the light transport function for general relighting [19, 43, 64], single image portrait relighting [41, 52], and multi-view face relighting [36, 53]. Most closely related, are methods that learn a continuous volumetric light transport field to relight the face in distant illumination [54] or near-field illumination [47, 62] settings. However, these methods do not account for refraction during volume rendering, making them unsuitable for handling surfaces like the cornea.

To enable image-based relighting, special efforts have been made to build light stages which consist of a dense set of controllable cameras and light sources [7, 22, 36, 37]. Improving practicality, LitNeRF [47] demonstrated it was possible to learn to relight the face from a sparse light stage containing an order of magnitude less viewpoints and light sources. Similar to LitNeRF, we build a sparse light stage for capturing images of the eye under varying illumination.

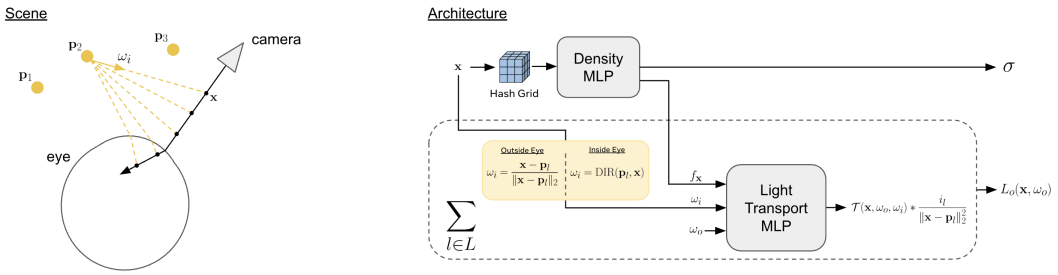


Figure 2. Overview of our Relightable Instant NGP architecture.

2.2. Neural rendering

Neural Radiance Fields (NeRF) [39] has demonstrated impressive abilities to render photo-realistic images of a scene, but only supports rendering from novel views under a fixed illumination. Bi et al. [6] was the first to extend NeRF for relighting by learning a continuous volumetric field of BRDF parameters, enabling relighting under a single collocated light source. Since then many works have extended NeRF for inverse rendering while enabling relighting under global illumination [11, 12, 25, 51, 67, 68]. While these works have exhibited excellent performance on general relighting, none of these methods address relighting in the presence of refraction which is the major focus of our work.

While NeRF is capable of modeling view-dependent appearance, it struggles to represent the specular reflections of glossy surfaces. To address this, some works have focused on reparameterizing NeRF's view-dependent appearance to simplify the mapping that the color MLP needs to learn [34, 55, 56]. Other works have focused on reflections in specific settings, such as planar mirrors and windows. NeRFReN [23] models the transmitted and reflected rays of a reflective planar surface with separate radiance fields, while MS-NeRF [65] learns multiple virtual subspaces of a scene containing mirrors. Unlike these works, we first recover a mesh of the cornea and then explicitly trace reflected light rays to shade glints on the cornea.

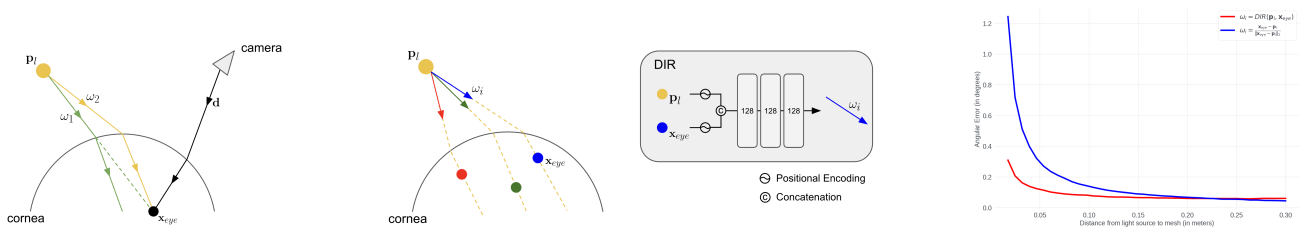
Aside from struggling to model reflections, NeRF is unable to properly handle refractive objects/media. To model transparent objects, some works try to learn refraction through differentiable rendering by warping rays [28, 42] or predicting where along the ray the transparent surface exists to perform refraction using Snell's law [66]. Chen et al. [15] estimate the transparent object's shape through a differentiable shape-from-silhouette pipeline and then perform explicit ray tracing through the scene, handling refraction using Snell's law. Similar to our work, they recover geometry first and account for all instances of light refraction between camera and light source; however, their method assumes the object they are modeling is fully transparent which does not hold for the eye.

Finally, [49] propose a way to learn high-fidelity depth maps by optimizing a NeRF from raw structured light images. Our method differs from there's as we focus on learning a relightable model from infrared images which can then be relit under a fringe projector to simulate structured light.

2.3. Modeling the face and eyes

Modeling the face has a been a long standing problem of interest. Previous works have explored learning the 3D geometry of the face using photometric stereo [14, 60], multi-view stereo [1, 3, 13], and morphable models [8–10, 33]. More recently, works have focused on learning both appearance and geometry of the face through neural rendering [18, 20, 24, 26, 29, 35, 46, 63, 69, 71]. However, these methods produce low fidelity eye reconstructions due to the eye's spatial resolution being low in the images they train on and not considering the effects of corneal refraction.

Substantially less effort has been made towards accurately reconstructing the geometry and appearance of the eye. [4, 5] propose ways to reconstruct all visible parts of the eye (i.e., sclera, cornea, and iris) from multi-view imagery. Differing from [4, 5], we not only reconstruct the geometry and appearence of the eye but also the periocular region. Schwartz et al. [48] propose a hybrid representation where an explicit eye mesh model is registered into a predicted face mesh; however, their iris appearance lacks detail and the subject cannot be relit. GazeNeRF [45] and ShellNeRF [32] enable NeRF to perform gaze redirection on a reconstructed subject, but neither of these methods are relightable. Most closely related to our work, EyeNeRF [31] models the geometry of the transparent cornea explicitly with an eye mesh and implicitly learns the appearance and geometry of the face and eye interior as a NeRF. Their method accounts for the refraction and reflection of camera rays, can perform gaze redirection, and can be relit using environment maps. Unlike EyeNeRF, our method additionally models the refraction of light rays used to relight the eye interior that resides behind the refractive cornea. Furthermore, we provide an extension of our method to relighting under fringe projectors for simulating structured light.



(a) (b) (c)

Figure 3. (a) For points inside the eye, the direct illumination ray is no longer the straight line between light position $\mathbf{p}_l$ and point $\mathbf{x}_{eye}$ due to refraction. (b) Our Direct Illumination Ray (DIR) MLP predicts the correct light ray that will intersect a point behind the refractive cornea. (c) The direct illumination ray under refraction predicted by our network is more accurate than the straight line assumption.

3. Method

In this paper, we build a system which can reconstruct and relight the eye from a sparse set of multi-view images with varying illumination. We begin by introducing a relightable architecture that models appearance with a data-driven light transport function which describes the relationship between a point light source and outgoing radiance (Section 3.1). To account for the geometry of the transparent cornea, we leverage a hybrid eye representation where an explicit eye mesh is registered into our neural volume (Section 3.2). Using our hybrid eye representation, we model glints on the cornea with an explicit specularity shader and handle refraction of rays using Snell's law. To properly condition our light transport function on incoming light rays in the presence of refraction, we introduce a Direct Illumination Ray network which predicts the light ray direction that will pass through a sampled point after undergoing refraction (Section 3.3). Finally, we describe how our pipeline can be extended to relighting with fringe projectors to simulate structured light (Section 3.4).

3.1. Relightable Instant NGP

Instant NGP [40] is a neural rendering method which can achieve high quality view synthesis while benefitting from significantly faster training and rendering time than the original NeRF [39]. However, similar to NeRF, Instant NGP cannot perform relighting. In this section, we describe our Relightable Instant NGP, which learns to model both the geometry and light transport of the eye and periocular region from images captured under varying illumination.

Ignoring emission, the outgoing radiance at a point $\mathbf{x}$ in the direction ω_o can be defined by the light transport equation as:

$$L_o(\mathbf{x}, \omega_o) = \int_{\Omega} T(\mathbf{x}, \omega_o, \omega_i) * L_i(\mathbf{x}, \omega_i)\, d\omega_i \qquad (1)$$

where light transport $T(\mathbf{x}, \omega_o, \omega_i)$ describes the relationship between incoming and outgoing light at different points in space and $L_i(\mathbf{x}, \omega_i)$ is the incident radiance at point $\mathbf{x}$ from direction ω_i.

We choose to approximate the real underlying light transport T as an MLP $\mathcal{T}$ which is learned in a data-driven manner. The total outgoing radiance at a point $\mathbf{x}$ in the direction ω_o illuminated by a set of point lights L is then:

$$L_o(\mathbf{x}, \omega_o) = \sum_{l \in L} \mathcal{T}(\mathbf{x}, \omega_o, \omega_i) * \frac{i_l}{||\mathbf{x} - \mathbf{p}_l||_2^2} \qquad (2)$$

where p_l and i_l are the position and intensity of point light source l, respectively. To model near-field illumination effects, we consider a spatially-varying incident light direction $\omega_i = \frac{\mathbf{x}-\mathbf{p_l}}{\|\mathbf{x}-\mathbf{p_l}\|_2}$. Furthermore, we scale our intensity i_l by $\frac{1}{\|\mathbf{x}-\mathbf{p_l}\|_2^2}$ to model the inverse-square law of a point light.

We base our neural field's architecture on Instant NGP. For a point $\mathbf{x}$, we first query a feature from a multi-resolution hash encoding $\mathcal{H}$ before using an MLP $\mathcal{D}$ to predict a geometry feature and density $(\mathbf{f_x}, \sigma(\mathbf{x})) = \mathcal{D}(\mathcal{H}(\mathbf{x}))$. Using our light transport MLP $\mathcal{T}$ and Equation 2, we then compute our outgoing radiance $L_o(\mathbf{x}, \omega_o)$. In practice, we condition $\mathcal{T}$ on geometry feature $\mathbf{f_x}$ instead of on point $\mathbf{x}$.

Volume rendering can then be achieved by replacing the predicted radiance in NeRF with Equation 2. We define a camera ray as $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, where $\mathbf{o}$ is the ray origin and $\mathbf{d}$ is the ray direction. Letting $\omega_o = -\mathbf{d}$, we then render a pixel as:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i L_o(\mathbf{r}(t_i), \omega_o) \qquad (3)$$

$$\text{where} \quad \alpha_i = 1 - exp(-\sigma_i \delta_i)$$

$$T_i = \prod_{j=1}^{i-1} exp(-\sigma_j \delta_j)$$

where $\sigma_i = \sigma(\mathbf{r}(t_i))$ and $\delta_i = t_{i+1} - t_i$.

Our training loss is then defined as:

$$L = \lambda_{color} L_{color} + \lambda_{dist} L_{dist} \qquad (4)$$

where L_{color} is the $L1$ loss between predicted pixel color $\hat{C}(\mathbf{r})$ and ground truth pixel color $C(\mathbf{r})$ and L_{dist} is Mip-NeRF 360's distortion loss [2]. We set $\lambda_{color} = 1$ and $\lambda_{dist} = 0.03$ for all our experiments.

3.2. Hybrid Eye Representation

While our architecture enables training from images with varying illumination and relighting under arbitrary near-field point light sources, it is not suitable for modeling the geometry and appearance of the eye due to the cornea. In particular, the transparency of the cornea makes learning its geometry difficult. Additionally, light sources cause sharp specularities on the cornea, known as glints. These specularities are difficult to learn as their position and shape change rapidly as a function of viewing angle, light source position, and cornea shape. Finally, refraction distorts the appearance of the iris and pupil breaking the multi-view consistency assumption of NeRF.

To address these challenges, we leverage a hybrid eye representation similar to Li et al. [31], where prior to training our NeRF model we extract an explicit mesh of the eye and register it into our NeRF volume. This allows us to represent the surface of the transparent cornea in an explicit manner while our NeRF model learns the geometry of the face and interior of the eye implicitly. Moreover, it provides us with a way to properly model the reflection and refraction of rays that hit the cornea.

3.2.1. Explicit eye mesh model

We first reconstruct a mesh of the eye using an eye scanner from Transfolio. The eye pose needed to register the mesh into our NeRF training volume is then estimated from our reconstructed eye mesh and our light stage. Using the known light positions and camera poses, we optimize for an eye pose which minimizes both pupil and glint reprojection errors against the detected pupils and glints captured in our light stage images. Unlike [31], our method does not require a manual initialization of the eye pose and shape in Blender, making it more scalable.

3.2.2. Glint rendering

From our sparse set of point lights, it is difficult for our model to learn that the cornea produces mirror-like specularities. Instead of trying to learn these specularities implicitly, we explicitly shade glints onto the cornea. To do so, we make the assumption that the cornea acts as a perfect mirror and treat our point light source as a disk with center $\mathbf{p_l}$, radius r_{disk}, and color s.

Given a camera ray $\mathbf{r}$, we first find where the ray intersects our eye mesh using a ray tracer. We denote the intersection point as $\mathbf{x}_{mesh}$, which exists a distance of t_{mesh} along the ray $\mathbf{r}$. We further obtain surface normal $\mathbf{n}$ using barycentric interpolation of the vertex normals of the intersected triangle face. The reflected ray is then defined as:

$$\begin{aligned} \mathbf{r}_r(t) &= \mathbf{x}_{mesh} + t\mathbf{d}_r \\ \text{where } \mathbf{d}_r &= \mathbf{d} - 2(\mathbf{d}\cdot\mathbf{n})\mathbf{n} \end{aligned} \tag{5}$$

Using reflected ray $\mathbf{r}_r$, we then find where along the ray it intersects the z-plane of our light disk and denote this distance as t_l. Finally, the glint c_{glint} is produced as:

$$c_{glint} = \begin{cases} s, & \text{if } \|\mathbf{r}_r(t_l) - \mathbf{p}_l\|_2 < r_{disk} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

For rendering from our training cameras, we set point light size $r_{disk} = 1.25$cm and light color to $s = 255$ (i.e., white).

3.2.3. Camera ray refraction

Aside from mirror-like specularities, the cornea also exhibits refraction which distorts the appearance of the iris and pupil across camera views. To account for this distorted appearance, we choose to explicitly account for refraction by bending rays. In particular, from our previously computed ray-mesh intersection, we additionally compute a refracted ray $\mathbf{r}_d(t) = \mathbf{x}_{mesh} + t\mathbf{d}_d$, where $\mathbf{d}_d$ is the refracted ray direction computed using Snell's law. This allows us to split our rendering into two parts, rendering outside the eye using pre-mesh interesection ray $\mathbf{r}$ to produce color c_{outer} using Equation 3, and similarly rendering inside the eye along refracted ray $\mathbf{r}_d$ to produce color c_{inner}. Letting $w_i = T_i\alpha_i$, we can then produce the final pixel color as a composition of our rendered colors:

$$\hat{C}(\mathbf{r}) = c_{outer} + (1 - \sum_{i=0}^{N_{outer}} w_i) * (c_{glint} + c_{inner}) \tag{7}$$

where N_{outer} is the number of samples along the ray between the pre-mesh intersection ray's sampling bounds t_{near} and t_{mesh}.

3.3. Light ray refraction

Our light transport MLP $\mathcal{T}$ is conditioned on the light ray direction which directly illuminates a 3D point. For points outside the eye, this light ray direction can be defined as $\omega_i = \frac{\mathbf{x}-\mathbf{p_l}}{\|\mathbf{x}-\mathbf{p_l}\|_2}$. However, as shown in Figure 3a, for points inside the eye this light ray direction is no longer accurate as the cornea causes the light ray to refract.

To address this, we introduce a Direct Illumination Ray network that predicts the light ray which directly illuminates a point under refraction, as shown in Figure 3b. In particular, given a light source l with position $\mathbf{p}_l$ and a sampled point inside the eye $\mathbf{x}_{eye}$, our Direct Illumination Ray network predicts the light ray direction ω_i that is used to condition our light transport MLP $\mathcal{T}$ in Equation 2:

$$\omega_i = \text{DIR}(\mathbf{p}_l, \mathbf{x}_{eye}) \tag{8}$$

We model our Direction Illumination Ray network as an MLP and use a fourier feature encoding to encode both light position and sampled point inside the eye. We train this network separately from our Relightable Instant NGP model and supervise our training with the ground truth light ray

	PSNR ↑	SSIM ↑	LPIPS ↓
Instant NGP [40]	26.78	0.7797	**0.6121**
Ours	**29.43**	**0.8278**	0.6260

Table 1. Quantitative comparison on a holdout camera viewpoint.

	View Synthesis / Relighting			Iris Geometry
	PSNR ↑	SSIM ↑	LPIPS ↓	MAE ↓
NR	28.55	0.8086	0.6384	8.65
CRR	28.94	0.8184	0.6262	**5.09**
CRR + LRR	**29.43**	**0.8278**	**0.6260**	5.12

Table 2. Ablation of our method with no refraction (NR), camera ray refraction (CRR), and both camera ray refraction + light ray refraction (CRR + LRR).

direction ω_{gt} using a cosine similarity loss. We refer readers to our supplemental for a description of how we generate our training data pairs $\{(\mathbf{p}_l, \mathbf{x}_{eye}), \omega_{gt}\}$.

3.4. Relighting with fringe projectors

We now introduce a way to relight the eye under a fringe projector to simulate structured light. In order to do so, we make an assumption that the projector acts like a pinhole camera and follows perspective projection.

Given a fringe projector with intrinsics $\mathbf{K}$ and extrinsics $\mathbf{M} = [\mathbf{R}, \mathbf{t}]$, we define the illumination source of the projector as its center $\mathbf{c} = -\mathbf{R}^T\mathbf{t}$ with a maximal intensity i. Note that the true intensity emitted into the scene is dependent on the chosen fringe pattern. In our experiments, we use a vertical sinusoidal grating as our fringe pattern. We first generate a sinusoidal fringe pattern image as $I_{atten}(u, v) = \sin(2\pi f \frac{u}{W} + \phi)$, where W is the width of the image and f and ϕ are the frequency and phase of the sine wave, respectively. Then we model the true intensity emitted in a direction as the maximal intensity i times an attenuation that is looked up in I_{atten}. To simulate fringe projection in our renderings, we introduce an attenuation factor $A(\omega_i)$ into the outgoing radiance of Equation 2:

$$L_o(\mathbf{x}, \omega_o) = \mathcal{T}(\mathbf{x}, \omega_o, \omega_i) * \frac{i}{||\mathbf{x} - \mathbf{c}||_2^2} * A(\omega_i) \quad (9)$$

$$A(\omega_i) = \mathcal{B}(I_{atten}, p(\omega_i)) \quad (10)$$

$$p(\omega_i) = \mathbf{K}(\mathbf{R}(\mathbf{c} + \omega_i) + \mathbf{t}) \quad (11)$$

where $\mathcal{B}(I_{atten}, p(\omega_i))$ denotes bilinear interpolation of I_{atten} at the projected point $p(\omega_i)$. We note that we use the direct illumination ray ω_i for looking up attention values instead of sampled point $\mathbf{x}$, which allows us to account for the refraction of the fringe pattern caused by the cornea.

4. Experiments

4.1. Implementation details

Our Direct Illumination Ray network is modeled as a 3-layer MLP with a hidden dimension of 128 at each layer. We set the number of frequencies for our fourier feature encoding to 10. The model is trained per subject for 100k iterations, which takes about 1 hour on a NVIDIA A100.

Our Relightable Instant NGP is based off of Instant NGP [40]. We leverage a multi-resolution hash encoding where we vary the hash grid resolution from 16 to 2048 and set our hash map to contain 2^{21} entries. We model the density network as 3-layer MLP with a hidden dimension of 256 at each layer and the light transport network as a 5-layer MLP with a hidden dimension of 256 at each layer. View directions and illumination directions are encoded as 4th order spherical harmonics. We train our model per subject for 100k iterations, which takes 6 hours on a NVIDIA A100.

4.2. Datasets

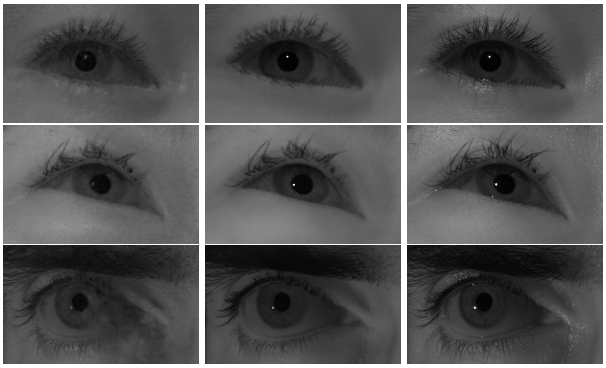
For training and evaluation, we build a dataset containing high quality captures of 11 different subjects. Our light stage consists of 18 IR cameras spread along the frontal hemisphere such that their visual hull contains half of the subject's face. Cameras are positioned roughly 30 cm from the face and capture 2048×2048 images focused on the eye in view. To illuminate the face, we use 16 point light sources. Each point light can be controlled individually and turning on all the point lights simultaneously provides roughly uniform illumination of the face. Cameras extrinsic and intrinsics are calibrated using a target board and light positions are calibrated using a mirror sphere. During capture, we ask the subject to fixate on an illuminated LED on a target board and capture a 35 frame burst of images under varying illumination. We refer readers to our supplemental for an example capture of a subject.

4.3. Results

In this section, we present quantitative and qualitative results of our method for novel view synthesis, relighting, iris geometry, and structured light simulation. We refer readers to our supplemental material for more results.

4.3.1. Novel view synthesis

We begin by demonstrating that the original Instant NGP [40] is not suitable for training on OLAT images. Since the architecture is not relightable, Instant NGP must be trained from the multi-view images of a single light source, and therefore lighting is baked in and we can only render from novel viewpoints. On the other hand, with our architecture we are able to leverage more data for training by using different lighting conditions. For experiments on our model, we choose to hold out the point light source that we train



Instant NGP [40] Ours Ground Truth

Figure 4. Qualitative comparison of renderings from a novel viewpoint (and novel lighting condition for our method).

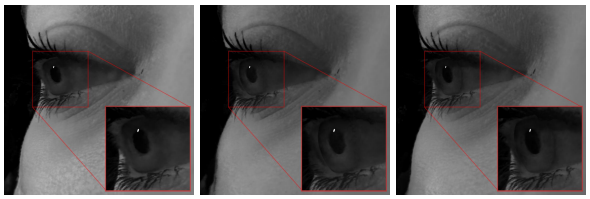
Instant NGP with and train our model with all our other point light sources. In this way, we can demonstrate that our method outperforms our baseline even when additionally generalizing to novel lighting conditions.

We quantitatively demonstrate the superiority of our method when generalizing to novel viewpoints (and novel light source for our method) in Table 1. In particular, we observe a 10% relative improvement in PSNR and 6% relative improvement in SSIM over our baseline, despite our baseline being trained on the specific lighting condition we evaluate on while our method was not.

In Figure 4, we present a qualitative comparison of our method against our baseline. Instant NGP struggles to represent glints on the cornea, producing only a faint cloudy white speckle on the eye. Alternatively, our method is able to recover the glints on the cornea caused by a light source thanks to our explicit glint shading. Furthermore, we observe that our method can more faithfully model the appearance of the eye. In the first two rows, we observe that Instant NGP produces an iris that appears flat and too bright, while our method models shading of the iris much better. Finally, we find that Instant NGP can suffer from floaters and blurriness (right side of eye in bottom row). This is likely due to poor convergence when training on images that contain non-uniform illumination, as occlusions of the single point light source can cause shadows which heavily alter the appearance of the eye and face from different viewpoints. Our method is less likely to suffer from these issues as appearance changes from shadows can be encoded by our transport field, and we can also train on more lighting conditions which ensure all points of the face are illuminated.

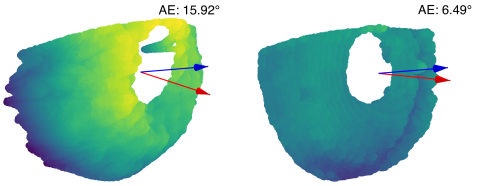
4.3.2. Relighting

To understand the importance of certain parts of our architecture, we evaluate different variants of our method when rendering under novel viewpoint and lighting conditions. In Table 2, we evaluate our Relightable Instant NGP model



NR CRR CRR + LRR

Figure 5. Qualitative comparison of renderings from a novel viewpoint and novel lighting condition.



NR CRR+LRR

Figure 6. Visualization of our reconstructed iris, estimated optic axis (red arrow), and a reference optic axis produced by our eye mesh pose fitting (blue arrow).

without accounting for refraction (NR), accounting for just camera ray refraction (CRR), and accounting for both camera and light ray refraction (CRR+LRR). We observe improvements across all metrics as we include camera ray refraction via Snell's law into our rendering and again see further improvements across all metrics as we account for light ray refraction with our Direct Illumination Ray prediction network. We further qualitatively evaluate these model variants in Figure 5. Without any refraction, the geometry of the cornea is not well observed and the iris appears slightly distorted to account for the lack of refraction. Upon adding camera ray refraction, the shape of cornea becomes more visible and iris appearance improves. Finally, with adding in light ray refraction the transition between iris and sclera, known as the limbus, becomes more apparent and shadowing on the eye due to occlusion by the upper eyelid and eyelashes improves.

4.3.3. Iris geometry

While we do not have ground truth iris geometry to compare against, we are still interested in understanding whether accounting for refraction leads to better iris geometry. When modeling the geometry of the eye, a common simplification is to treat the iris as a planar disk whose normal is aligned with the optic axis. We take inspiration from this and evaluate how well we can recover the subject's gaze direction from our reconstructed iris as a proxy for geometry evaluation. In particular, we fit a plane to our reconstructed iris and measure the angular error (AE) between the normal of the estimated plane and the reference optic axis we have from fitting the pose of our eye mesh. In Table 2, we

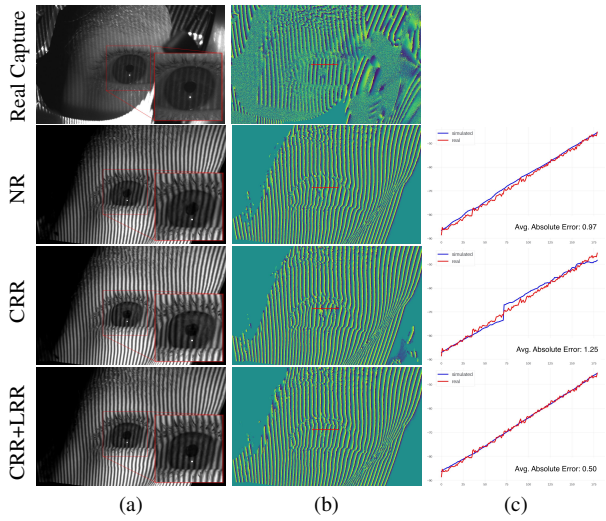


Figure 7. (a) Comparison of real and simulated structured light. (b) Wrapped phase computed using a 3-step PSP method. (c) Cross section comparison of the real and simulated wrapped phase.

	NR	CRR	CRR+LRR
Absolute Error	1.08	1.48	**0.98**

Table 3. Absolute error between simulated and real phase averaged over different cross sections of the iris.

show that adding in refraction improves gaze estimation by $3.5°$. Furthermore, in Figure 6, we show an example reconstructed iris with and without refraction, where we visualize the depth of the iris points along the optic axis by color. We find that without refraction, the iris is significantly more curved and takes the shape of the cornea, while the iris reconstruction when accounting for refraction is much flatter. Both improved gaze estimation and flatter iris reconstruction suggest that modeling refraction does improve learning geometry behind the cornea.

4.3.4. Structured light

To evaluate our simulation of structured light, we capture the same subject using both our light stage and a structured light rig. We calibrate the camera of our structured light rig and use the calibration for simulating renders with our learned reconstruction. In Figure 7a, we show a comparison between real and simulated structured light. Looking at the zoomed insets of the eye, we see that training our model without refraction (NR) or just camera ray refraction (CRR) leads to blurred and warped fringes. On the other hand, when using our full pipeline (CRR+LRR), we find that our method recovers the fringe pattern with minimal distortions even in the presence of the refractive cornea. To quanitatively verify this, we compute the wrapped phase (Figure 7b) using a 3-step phase-shift profilometry (PSP) method and compute the absolute error between a cross section of the real and simulated wrapped phase (Figure 7c). While modeling camera ray refraction (CRR) improves iris geometry over the no refraction (NR) case, in Figure 7c and Table 3, we show that the simulated phase is worse as we do not accurately model a light ray's path from the surface back to the projector. However, adding in light ray refraction (CRR+LRR) corrects for this as our Direction Illumination Ray network predicts a light ray's correct path back to the projector, leading to the best wrapped phase simulation.

4.3.5. Light ray prediction

We demonstrate the importance of our Direct Illumination Ray (DIR) network by comparing the accuracy of the straight line approximation, $\omega_i = \frac{\mathbf{x}_{eye} - \mathbf{p}_l}{||\mathbf{x}_{eye} - \mathbf{p}_l||_2}$, of a direct light ray under refraction and our predicted light ray from our Direct Illumination Ray network. In Figure 3c, we plot the angular errors of these predictions against the ground truth direct light ray ω_{gt} that originates from a point source $\mathbf{p}_l$ and intersects the point $\mathbf{x}_{eye}$ inside the eye after undergoing refraction. The straight line assumption becomes increasingly less accurate as the point light source moves towards the eye. While the angular error of our predictions follow a similar trend, we see that it occurs at a much slower rate. In particular, we find that our method obtains a $3-4\times$ improvement in angular error over the straight line assumption in the near-field illumination setting (< 0.05m). This reduction in angular error is important as accurately modeling fringes for structured light requires properly tracing a ray from the surface back to a light source.

5. Conclusion

In this paper, we build a system which can reconstruct and relight the eye from a sparse set of multi-view images captured under varying illumination. By resolving the geometry of the cornea ahead of time, we are able to correctly model the reflection and refraction caused by the cornea, which we demonstrate leads to better appearance and geometry. We further introduce a Direct Illumination Ray network which allows us to model direct reflections of surfaces behind the cornea. We show that accurately modeling the refraction of light rays improves relighting quality and enables us to simulate structured light on the iris which lies behind the refractive cornea. Despite our method being relightable, we still struggle to represent highly specular reflections on the skin and high frequency shadows on the eye (e.g., eyelash strand shadows). We believe addressing these issues are interesting lines of future work which can lead to achieving better photo-realism.

References

[1] Ziqian Bai, Zhaopeng Cui, Jamal Ahmed Rahim, Xiaoming Liu, and Ping Tan. Deep facial non-rigid multi-view stereo.

In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 5850–5860, 2020. 3

[2] 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. 4

[3] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. High-quality single-shot capture of facial geometry. In *ACM SIGGRAPH 2010 papers*, pages 1–9, 2010. 3

[4] Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. High-quality capture of eyes. *ACM Trans. Graph.*, 33(6):223–1, 2014. 1, 3

[5] Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. Lightweight eye capture using a parametric model. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016. 3

[6] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020. 1, 3

[7] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM Transactions on Graphics (ToG)*, 40(4):1–15, 2021. 2

[8] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 157–164, 2023. 3

[9] James Booth, Anastasios Roussos, Stefanos Zafeiriou, Allan Ponniah, and David Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5543–5552, 2016.

[10] James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. Large scale 3d morphable models. *International Journal of Computer Vision*, 126(2): 233–254, 2018. 3

[11] 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. 1, 3

[12] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems*, 34: 10691–10704, 2021. 1, 3

[13] Derek Bradley, Wolfgang Heidrich, Tiberiu Popa, and Alla Sheffer. High resolution passive facial performance capture. In *ACM SIGGRAPH 2010 papers*, pages 1–10, 2010. 3

[14] Xuan Cao, Zhang Chen, Anpei Chen, Xin Chen, Shiying Li, and Jingyi Yu. Sparse photometric 3d face reconstruction guided by morphable models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4635–4644, 2018. 3

[15] Xiaoxue Chen, Junchen Liu, Hao Zhao, Guyue Zhou, and Ya-Qin Zhang. Nerrf: 3d reconstruction and view synthesis for transparent and specular objects with neural refractive-reflective fields. *arXiv preprint arXiv:2309.13039*, 2023. 3

[16] Flávio Luiz Coutinho and Carlos Hitoshi Morimoto. Free head motion eye gaze tracking using a single camera and multiple light sources. In *2006 19th Brazilian Symposium on Computer Graphics and Image Processing*, pages 171–178. IEEE, 2006. 2

[17] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000. 2

[18] Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Nießner. Dynamic neural radiance fields for monocular 4d facial avatar reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8649–8658, 2021. 3

[19] Duan Gao, Guojun Chen, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. Deferred neural lighting: free-viewpoint relighting from unstructured photographs. *ACM Transactions on Graphics (TOG)*, 39(6):1–15, 2020. 2

[20] Philip-William Grassal, Malte Prinzler, Titus Leistner, Carsten Rother, Matthias Nießner, and Justus Thies. Neural head avatars from monocular rgb videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18653–18664, 2022. 3

[21] Elias Daniel Guestrin and Moshe Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on biomedical engineering*, 53(6):1124–1133, 2006. 2

[22] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)*, 38(6):1–19, 2019. 2

[23] 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, 2022. 3

[24] Yuxuan Han, Junfeng Lyu, and Feng Xu. High-quality facial geometry and appearance capture at home. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 697–707, 2024. 1, 3

[25] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2023. 1, 3

[26] Kacper Kania, Stephan J Garbin, Andrea Tagliasacchi, Virginia Estellers, Kwang Moo Yi, Julien Valentin, Tomasz Trzciński, and Marek Kowalski. Blendfields: Few-shot example-driven facial modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 404–415, 2023. 3

[27] Harsimran Kaur and Roberto Manduchi. Eyegan: Gaze-preserving, mask-mediated eye image synthesis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 310–319, 2020. 1

[28] Wooseok Kim, Taiki Fukiage, and Takeshi Oishi. Ref ^{2}-nerf: Reflection and refraction aware neural radiance field. *arXiv preprint arXiv:2311.17116*, 2023. 3

[29] Tobias Kirschstein, Shenhan Qian, Simon Giebenhain, Tim Walter, and Matthias Nießner. Nersemble: Multi-view radiance field reconstruction of human heads. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023. 3

[30] Aaron Lefohn, Brian Budge, Peter Shirley, Richard Caruso, and Erik Reinhard. An ocularist's approach to human iris synthesis. *IEEE Computer Graphics and Applications*, 23 (6):70–75, 2003. 1

[31] Gengyan Li, Abhimitra Meka, Franziska Mueller, Marcel C Buehler, Otmar Hilliges, and Thabo Beeler. Eyenerf: a hybrid representation for photorealistic synthesis, animation and relighting of human eyes. *ACM Transactions on Graphics (ToG)*, 41(4):1–16, 2022. 1, 2, 3, 5

[32] Gengyan Li, Kripasindhu Sarkar, Abhimitra Meka, Marcel Buehler, Franziska Mueller, Paulo Gotardo, Otmar Hilliges, and Thabo Beeler. Shellnerf: Learning a controllable high-resolution model of the eye and periocular region. In *Computer Graphics Forum*, page e15041. Wiley Online Library, 2024. 3

[33] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017. 3

[34] Li Ma, Vasu Agrawal, Haithem Turki, Changil Kim, Chen Gao, Pedro Sander, Michael Zollhöfer, and Christian Richardt. Specnerf: Gaussian directional encoding for specular reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21188–21198, 2024. 3

[35] Zhiyuan Ma, Xiangyu Zhu, Guo-Jun Qi, Zhen Lei, and Lei Zhang. Otavatar: One-shot talking face avatar with controllable tri-plane rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16901–16910, 2023. 3

[36] Abhimitra Meka, Christian Haene, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, et al. Deep reflectance fields: high-quality facial reflectance field inference from color gradient illumination. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 2

[37] Abhimitra Meka, Rohit Pandey, Christian Haene, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. Deep relightable textures: volumetric performance capture with neural rendering. *ACM Transactions on Graphics (TOG)*, 39(6):1–21, 2020. 2

[38] Clara Mestre, Josselin Gautier, and Jaume Pujol. Robust eye tracking based on multiple corneal reflections for clinical applications. *Journal of biomedical optics*, 23(3):035001–035001, 2018. 2

[39] 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 ACM*, 65(1):99–106, 2021. 1, 3, 4

[40] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 4, 6, 7

[41] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5124–5133, 2020. 2

[42] Jen-I Pan, Jheng-Wei Su, Kai-Wen Hsiao, Ting-Yu Yen, and Hung-Kuo Chu. Sampling neural radiance fields for refractive objects. In *SIGGRAPH Asia 2022 Technical Communications*, pages 1–4, 2022. 3

[43] Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. Image based relighting using neural networks. *ACM Transactions on Graphics (ToG)*, 34(4):1–12, 2015. 2

[44] Kerstin Ruhland, Sean Andrist, Jeremy Badler, Christopher Peters, Norman Badler, Michael Gleicher, Bilge Mutlu, and Rachel Mcdonnell. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics 2014-State of the Art Reports*, pages 69–91, 2014. 1

[45] Alessandro Ruzzi, Xiangwei Shi, Xi Wang, Gengyan Li, Shalini De Mello, Hyung Jin Chang, Xucong Zhang, and Otmar Hilliges. Gazenerf: 3d-aware gaze redirection with neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9676–9685, 2023. 3

[46] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2024. 1, 3

[47] Kripasindhu Sarkar, Marcel C Bühler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, Paulo Gotardo, Thabo Beeler, et al. Litnerf: Intrinsic radiance decomposition for high-quality view synthesis and relighting of faces. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023. 1, 2

[48] Gabriel Schwartz, Shih-En Wei, Te-Li Wang, Stephen Lombardi, Tomas Simon, Jason Saragih, and Yaser Sheikh. The eyes have it: An integrated eye and face model for photorealistic facial animation. *ACM Transactions on Graphics (TOG)*, 39(4):91–1, 2020. 1, 3

[49] Aarrushi Shandilya, Benjamin Attal, Christian Richardt, James Tompkin, and Matthew O'toole. Neural fields for structured lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3512–3522, 2023. 3

[50] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2107–2116, 2017. 1

[51] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron.

Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 1, 3

[52] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 2

[53] Tiancheng Sun, Zexiang Xu, Xiuming Zhang, Sean Fanello, Christoph Rhemann, Paul Debevec, Yun-Ta Tsai, Jonathan T Barron, and Ravi Ramamoorthi. Light stage super-resolution: continuous high-frequency relighting. *ACM Transactions on Graphics (TOG)*, 39(6):1–12, 2020. 2

[54] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. *arXiv preprint arXiv:2107.12351*, 2021. 1, 2

[55] 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. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 3

[56] Dor Verbin, Pratul P Srinivasan, Peter Hedman, Ben Mildenhall, Benjamin Attal, Richard Szeliski, and Jonathan T Barron. Nerf-casting: Improved view-dependent appearance with consistent reflections. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–10, 2024. 3

[57] Jiazhang Wang, Tianfu Wang, Bingjie Xu, Oliver Cossairt, and Florian Willomitzer. Accurate eye tracking from dense 3d surface reconstructions using single-shot deflectometry. *arXiv preprint arXiv:2308.07298*, 2023. 2

[58] Kang Wang, Rui Zhao, and Qiang Ji. A hierarchical generative model for eye image synthesis and eye gaze estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 440–448, 2018. 1

[59] Tianfu Wang, Jiazhang Wang, Oliver Cossairt, and Florian Willomitzer. Optimization-based eye tracking using deflectometric information. *IEEE Transactions on Computational Imaging*, 2024. 2

[60] Xueying Wang, Yudong Guo, Bailin Deng, and Juyong Zhang. Lightweight photometric stereo for facial details recovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 740–749, 2020. 3

[61] Erroll Wood, Tadas Baltrusaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 3756–3764, 2015. 1

[62] Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. Renerf: Relightable neural radiance fields with nearfield lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22581–22591, 2023. 1, 2

[63] Yingyan Xu, Prashanth Chandran, Sebastian Weiss, Markus Gross, Gaspard Zoss, and Derek Bradley. Artist-friendly relightable and animatable neural heads. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2457–2467, 2024. 3

[64] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (ToG)*, 37 (4):1–13, 2018. 2

[65] Ze-Xin Yin, Jiaxiong Qiu, Ming-Ming Cheng, and Bo Ren. Multi-space neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12407–12416, 2023. 3

[66] Yifan Zhan, Shohei Nobuhara, Ko Nishino, and Yinqiang Zheng. Nerfrac: Neural radiance fields through refractive surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18402–18412, 2023. 3

[67] 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. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 3

[68] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022. 3

[69] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C Bühler, Xu Chen, Michael J Black, and Otmar Hilliges. Im avatar: Implicit morphable head avatars from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13545–13555, 2022. 3

[70] Yi Zheng, Qing Chao, Yatong An, Seth Hirsh, and Alexander Fix. Fringe projection-based single-shot 3d eye tracking using deep learning and computer graphics. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) IV*, pages 265–275. SPIE, 2023. 2

[71] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Instant volumetric head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4574–4584, 2023. 3