Seeing the World through Your Eyes

Hadi Alzayer*   Kevin Zhang*   Brandon Feng   Christopher A. Metzler   Jia-Bin Huang
University of Maryland, College Park
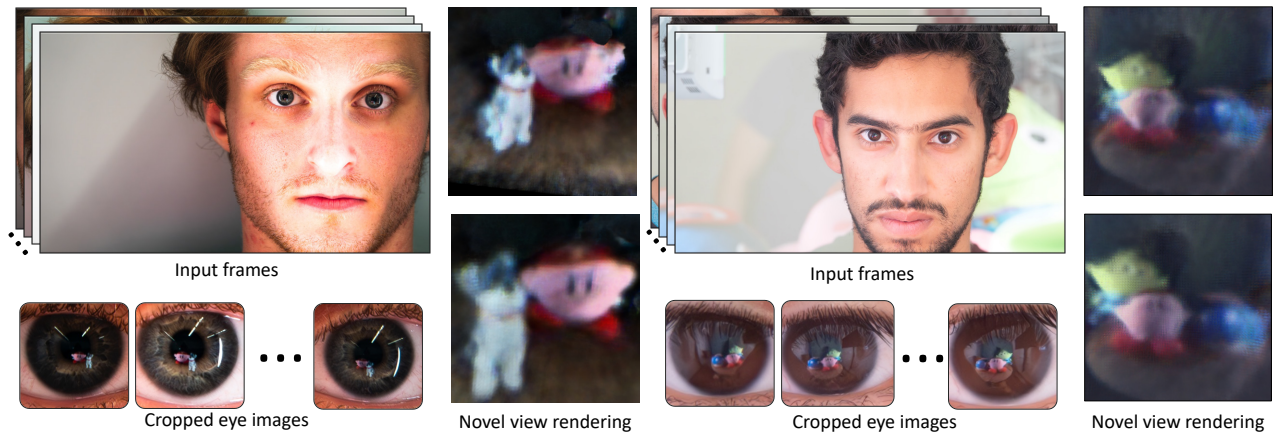https://world-from-eyes.github.io/

Figure 1. **Radiance field reconstruction using eye reflections.** The human eye is highly reflective. We show that from a sequence of frames that capture a moving head, we can reconstruct the radiance field and render the scene of what the person is observing using only the reflections off their eyes.

Abstract

The reflective nature of the human eye is an underappreciated source of information about what the world around us looks like. By imaging the eyes of a moving person, we capture multiple views of a scene outside the camera's direct line of sight through the reflections in the eyes. In this paper, we reconstruct a radiance field beyond the camera's line of sight using portrait images containing eye reflections. This task is challenging due to 1) the difficulty of accurately estimating eye poses and 2) the entangled appearance of the iris textures and the scene reflections. To address these, our method jointly optimizes the cornea poses, the radiance field depicting the scene, and the observer's eye iris texture. We further present a regularization prior on the iris texture to improve scene reconstruction quality. Through various experiments on synthetic and real-world captures featuring people with varied eye colors, and lighting conditions, we demonstrate the feasibility of our approach to recover the radiance field using cornea reflections.

---

*Equal contribution

1. Introduction

The human eye is a remarkable organ that enables vision and holds valuable information about the surrounding world. While we typically use our own eyes as two *lenses* to focus light onto the photosensitive cells composing our retina, we would also capture the light reflected from the cornea if we look at someone else's eyes. When we use a camera to image the eyes of another, we effectively turn their eyes into a pair of *mirrors* in the overall imaging system. Since the light that reflects off the observer's eyes share the same source as the light that reaches their retina, our camera can form images containing information about the surrounding world the observer sees.

Prior studies have explored recovering a panoramic image of the world the observer sees and simple 3D structures like boxes that the observer is looking at from manually specified correspondences from a single image of two eyes [27, 28]. Follow-up works have further explored applications such as personal identification [10, 29], detecting grasp posture [53], focused object estimation [40], illumination estimation [48], and relighting [26]. Given the recent advancements in 3D vision and graphics, we wonder: Can we do more than reconstruct a single panoramic environ-

ment map, simple 3D structures, or recognize patterns? Is it possible to recover the 3D world seen by the observer?

In this paper, we answer these questions by reconstructing a radiance field from a sequence of eye images. We start from the insight that our eyes capture/reflect multi-view information as we naturally move our heads. We draw inspiration from the classical imaging formulation proposed by [27] and integrate it with the recent advances in radiance field reconstruction spearheaded by Neural Radiance Fields (NeRF) [21], which prior work has demonstrated can achieve high-quality view synthesis.

However, while conceptually straightforward, reconstructing a NeRF from eye images is extremely challenging in practice. The first challenge is to separate the cornea reflections from the iris textures of human eyes. Unlike the clear images of the scene typically assumed in standard captures, the eye images we obtain are inherently blended with iris textures. This composition disrupts the pixel correspondence and complicates the reconstruction process. The second challenge is cornea pose estimation. Unlike the standard NeRF capture setup, which requires a *moving camera* to capture multi-view information (often followed by camera pose estimation), our approach employs a *stationary camera* and extracts the multi-view cues from eye images under head movement, as shown in Figure 2. Thus, we need to estimate the 3D positions and orientation of the eyes accurately from image observations. However, this is difficult because of how small the eyes are in portraits.

To address these challenges, we repurpose NeRF for training on eye images by incorporating two crucial components: a) iris texture decomposition, which leverages a simple radial prior to facilitate separating the iris texture from the radiance field for the scene, and b) 6DoF cornea pose refinement, which enhances pose estimation accuracy despite the challenges of the small size of eyes.

To evaluate the performance and effectiveness of our approach, we generate a synthetic dataset of a complex indoor environment with images that capture the reflection from a synthetic cornea with realistic iris texture. We further implement a real-world setup with multiple objects to capture cornea images. We conduct extensive experiments on synthetic and real-world captured eye images to validate several design choices.

Our primary contributions are as follows:

- **Radiance field recovery from eyes**. We present a method for reconstructing radiance fields of the observer's world from eye images, integrating earlier foundational work with the latest advancements in neural rendering.
- **Cornea pose refinement and analysis**. We incorporate a cornea pose refinement to correct the initial noisy pose estimates of the eyes. Using a synthetic dataset, we evaluate the sensitivity to pose errors and the necessity of pose
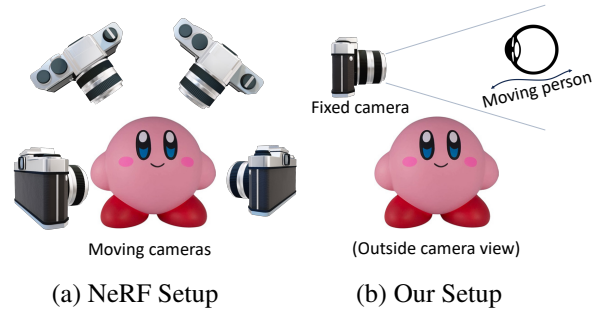

(a) NeRF Setup   (b) Our Setup

Figure 2. **NeRF for non-line-of-sight scene.** The typical NeRF capture setup requires multiple posed images (e.g., captured from a moving camera) for reconstruction. In our setup, we gather multi-view information of the scene through light reflected from the eyes of a moving person.

optimization for improved view synthesis results.
- **Radial prior for irises**. We introduce a radial prior for iris texture decomposition in cornea images, improving the quality of the reconstructed radiance field.

These advancements extend the current capabilities of radiance field reconstruction through neural rendering to handle partially corrupted image observations obtained from eye reflections, opening up new possibilities for research and development in the broader area of accidental imaging [6, 15, 38, 43] to reveal and capture 3D scenes beyond the visible line-of-sight.

2. Related Work

Corneal imaging. Imaging techniques that use the corneas as mirrors fall under the more general category of *catadioptric imaging*. Catadioptric imaging techniques use a combination of lenses and mirrors to capture images. The word catadioptric is derived from *catoptrics* (related to the Greek words for specular and mirrors) and *dioptrics* (related to an Ancient Greek lens-like instrument). In essence, catadioptric imaging seeks to leverage an additional (often curved) mirror to expand a lens-based imaging system's effective field of view. Early studies in catadioptric imaging focused primarily on the design of the mirror profiles and their impact on the final image quality. Prior work [2] studied three design criteria of a catadioptric imaging system: 1) the shape of the mirrors, 2) the resolution of the cameras, and 3) the focus settings of the cameras. Prior work has also demonstrated that it is possible to extract depth information from multiple views captured using a catadioptric system by extending classical depth-from-stereo ideas [24, 25].

Moreover, a creative way to realize an accidental catadioptric imaging system is by treating human eyes as external curved mirrors [27, 28]. The seminal work by Nishino and Nayar [27, 28] shows that it is possible to recover basic 3D structures like box meshes from manually specified

pixel correspondences in the reflections in the eyes in a single image and the epipolar geometry that relates the two eyes. Other applications of using human eyes as part of the imaging system include estimating light direction from the eyes to perform relighting [26, 44], calibration and gaze tracking in head-mounted displays [23, 35], and calibration of display camera systems [30]. Our work advances the prior work on scene reconstruction from eye images by 1) reconstructing a radiance field, which enables new applications such as novel view synthesis and 2) removing the iris texture from scene reconstruction.

Neural radiance field. Neural radiance fields (NeRF) [21] represent a milestone in novel view synthesis. NeRF adopts differentiable volume rendering to represent a 3D scene and uses neural networks to learn the density and color of each scene point. Following the success of NeRF, a plethora of follow-up works have been introduced to improve its rendering quality [3, 4], ability to handle scene dynamics [19, 32, 33, 36], inaccurate camera poses [5, 14, 20, 34, 49], and rendering speed [1, 22, 51]. Our work uses NeRF to parametrize the unknown scene we wish to recover from eye reflections. In particular, we modify the training framework from nerfstudio [41] to implement the NeRF-based scene reconstruction. Our input images are captured at a *fixed viewpoint*, differing from the typical NeRF setup that requires multi-view posed images.

Reflection removal. Removing reflections from captured images is a longstanding computational photography problem. The related literature on this topic can be summarized into two main categories: *multi-frame* and *single-image*. Multi-frame reflection removal methods [8, 17, 18, 39, 50] often exploit the differences of motion patterns between the background and reflection layers and impose various image priors as regularization. Single-image reflection removal methods exploit visual cues available in a single image, such as depth-of-field [13, 47], defocus-disparity [37], or learned image features [54]. More recently, various NeRF-based methods have studied how to accurately model and extract specular reflections from shiny or metallic objects [7, 42, 46, 52]. Nerfren [9] demonstrates that by fitting two NeRFs to model the reflection and transmittance components of the scene separately, reflections from planar surfaces like mirrors can be removed and re-rendered as a separate 3D scene. Our work differs from prior work on reflection removal with NeRF like Nerfren by regularizing the reconstruction of the diffuse color the human eye using a radial symmetry prior. Prior work has also focused on removing reflections, specifically from eyes, by exploiting the fact that the iris texture tends to be constant radially to design a radial autocorrelation prior [48]. Our work also differs from prior work on reflection removal from eyes by using a radial symmetry prior instead of the radial autocorrelation prior.
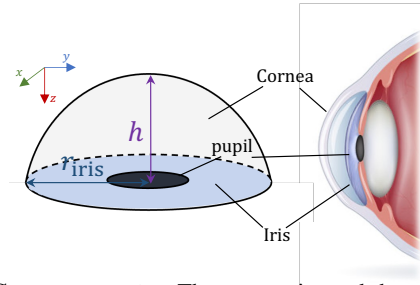

Figure 3. **Cornea geometry.** The cornea size and shape is uniform across all healthy humans. It can be modeled as an ellipsoid, with eccentricity of 0.5, and curvature of 7.8mm. The key fact that we exploit is that radius of the base r_{iris} is approx. 5.5mm, and the height of the cornea h is 2.18mm.

Non-line-of-sight imaging. Non-line-of-sight (NLOS) imaging aims to recover images of objects that are not directly visible by using light reflected off visible surfaces. Active NLOS imaging techniques involve using controlled light sources, such as time-of-flight sensors, to reconstruct the hidden scene [45]. Passive NLOS imaging, on the other hand, exploits natural or ambient light and does not require a controlled light source. Several works analyze the intensities changes on corner regions [6] or a blank wall [38] and reveal information about the hidden scene. Thermal reflections have been used to reconstruct the 3D body pose of non-line-of-sight humans [15]. Orca [42] uses reflections from a glossy object observed in multi-view images to train a NeRF for the surrounding environment. Unlike Orca, which relies on images captured with a moving camera while the "mirror" object is fixed, our method works for a stationary camera and uses the natural movement of the human eye "mirrors", which is visualized in Figure 2.

3. Background: Eye Model

The geometry of the human eye has been extensively studied [31] and is well known. The known eye geometry provides a strong prior to model the camera ray interactions with the eye in closed form. The major components that are visible in the eye are: the sclera; which is the white region of the eye, and the cornea; which includes the iris and the pupil. The cornea is covered by a thin film of tear fluid, making it highly reflective. As noted by Nishino and Nayar [27], since the cornea can act as a mirror, the combination of a camera and the cornea resembles a *catadioptric system*. In our work, we follow the same eye model adopted by Nishino and Nayar [27] for the geometry we assume for the eye.

The eye is modeled as a section of an ellipsoid, as illustrated in Figure 3, which can be described using

$$(1 - e)\, z^2 - 2Rz + r^2 = 0 \qquad (1)$$

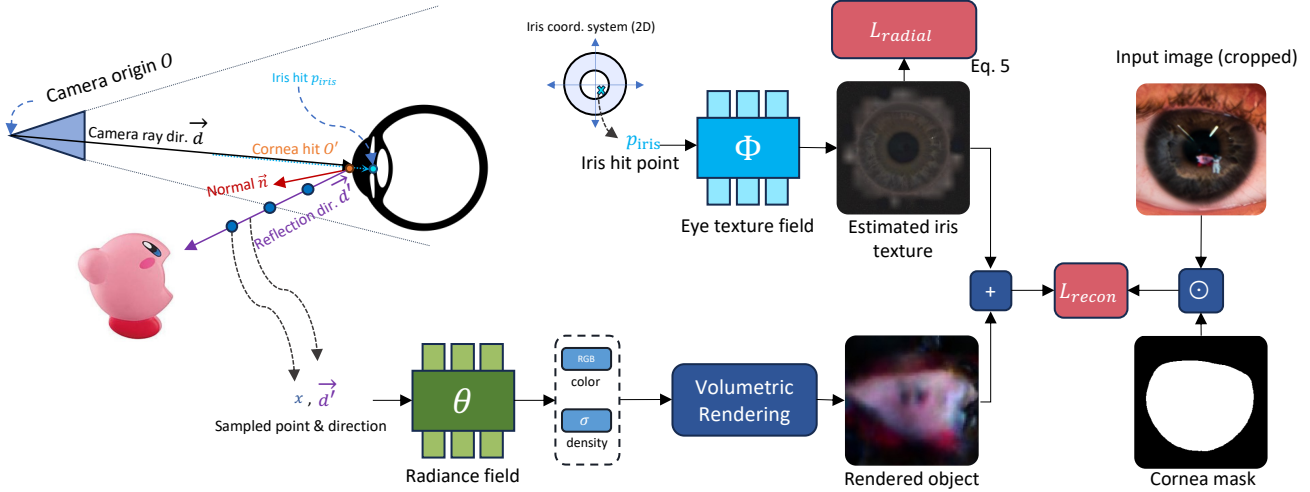where e is the eccentricity, R is the radius of the curvature at

Figure 4. **Joint optimization of radiance field and iris texture.** We trace rays that bounce off the cornea to render the scene using volume rendering. Additionally, we model the iris texture as a 2D neural field Φ whose input is the intersection of the casted ray from the camera and the cornea plane and add it to the rendered color from the radiance field to estimate the cornea measurement and compute the loss against the ground truth cornea measurement to optimize all the neural fields. To ensure that the scene is not absorbed into the iris texture, we regularize the texture field Φ with a radial loss L_{radial} that encourages the estimated texture to maintain a rotational consistency.

(a) Eye localization with GroundingDINO   (b) Ellipse fitting with ELLSeg   (c) Iris segmentation with SAM
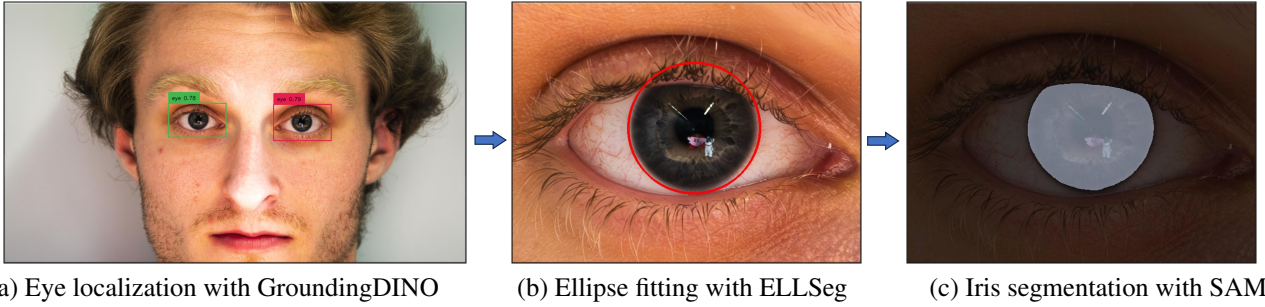
Figure 5. **Data processing pipeline.** To compute the iris ellipse parameters, we first obtain eye bounding boxes using GroundingDINO [16] and then conduct ellipse fitting using ELLSeg [12]. Since we only want to use the visible regions of the cornea in our radiance field optimization, to handle occlusion, we generate a segmentation mask of the iris from the approximated cornea ellipse using SAM [11].

the apex, and $r^2 = x^2 + y^2$. For an adult with healthy eyes, on average the eccentricity e is about 0.5 and curvature R is about 7.8 mm, and the radius at the base of the cornea r_{iris} is approximately 5.5 mm with a height h of 2.18mm. As in the prior work [26], we leverage the fact that humans have a known eye size to compute the average depth of the cornea in 3D by using the camera focal length f and the radius in image space r_{img} in the weak perspective equation:

$$\text{depth}_{\text{avg}} = r_{\text{iris}} \frac{f}{r_{\text{img}}}. \qquad (2)$$

that we use to initialize our estimate of the cornea position in our reconstruction pipeline. The initial pose is further refined, under correct perspective projection, during reconstruction using an optimization procedure that will be discussed more later. To compute the ray reflection direction, we need to compute the cornea's normal at the intersection point. Here the normal can be computed in closed form using ellipsoid equation Eq. 1:

$$\overrightarrow{n}(x, y, z) = \langle 2x, 2y, 2(1-e)z - 2R \rangle, \qquad (3)$$

which we use with the standard reflection equation to reflect the camera rays off the cornea. However, since the cornea is not a perfect mirror, we need to consider the transmitted rays that hit the iris and pupil. For simplicity, we model the iris and pupil as a flat and round plane directly behind the cornea. We can thus compute the hit points onto the iris via the camera ray-iris plane intersection.
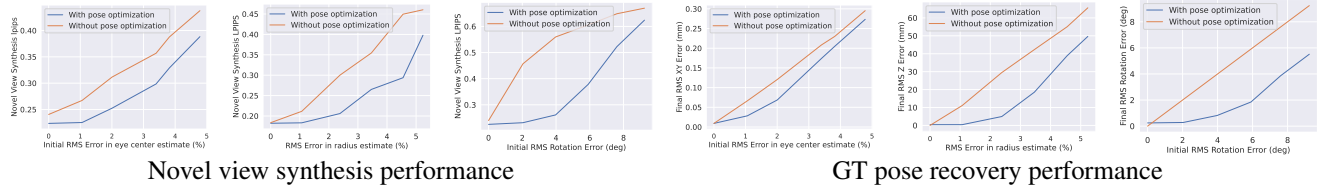
Novel view synthesis performance   GT pose recovery performance

Figure 6. **Pose optimization quantiative eval.** We evaluate the effects of pose optimization in simulation against various noise levels on the quality of synthesized views and the accuracy of recovered poses. For the XY and Z estimation plots, the x-axis represents the percent error in the estimated quantity relative to the true iris radius in pixels.
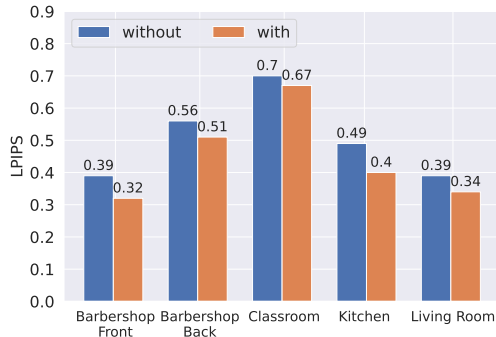

Figure 7. **Texture Decomposition Ablation.** We show that using a neural field to decompose the iris texture from the reflection improves reconstruction performance.

4. Method

Radiance field from reflection. One can train a radiance field by minimizing the photometric loss between ground truth pixel values from captured images and the rendered color. Each pixel color is computed through volume rendering using the color and density values of the sampled points along a camera ray. The ray associated with a pixel starts from the camera origin, denoted by O, and the viewing direction, denoted by $\overrightarrow{d}$.

However, in our setup, what we are interested in reconstructing the radiance field of the scene *reflected* from the person's eyes. In Figure 4, we illustrate how we use the rays reflected from the eye. The reflected ray starts with the origin where the camera ray intersects with the cornea at O', and in the direction of the reflected ray $\overrightarrow{d'}$ instead of using O and $\overrightarrow{d}$. We compute the reflected ray explicitly using the standard reflection equation:

$$\overrightarrow{d'} = \overrightarrow{d} - 2\left(\overrightarrow{n} \cdot \overrightarrow{d}\right)\overrightarrow{n}, \qquad (4)$$

where $\overrightarrow{n}$ is the normal at the hit point O'. Since we model the cornea geometry as an ellipsoid, we directly compute the hit points and normals using closed-form ellipsoid ray intersection formulas.

Iris texture decomposition Since the target images are the scene reflections off the cornea, training NeRF naively leads to poor view synthesis results as the cornea images contain both reflection (scene) and transmittance (iris texture) components. To recover only the scene in the radiance field, we jointly optimize a 2D field Φ to model the iris texture. The iris texture remains constant across the different views while the person moves, while the scene reflections vary. We thus use a texture field shared across input images. Since we model the iris as a flat, round plane behind the cornea, we can directly compute the *2D coordinate* on the iris plane by performing ray-plane intersection in 3D.

However, when a part of the scene does not display considerable motion across the training views (e.g., limited head motion of the subject), parts of the scene can be "absorbed" as part of the iris texture instead of the scene. To resolve this issue, we propose a radial regularization that encourages radial symmetry of the recovered texture. The iris is far from being perfectly rotationally symmetric. However, our observation is that as we rotate the iris, the color variance is small. We use this observation to regularize the colors we learn with the iris texture by penalizing the pixels that are rotationally inconsistent. At each step, for each point p, we empirically estimate the distribution of colors of the *ring*, by computing the average color μ_r and the standard deviation σ_r of the color ring. We then compute the range of the 10th and 90th percentile colors $[c_{10}, c_{90}]$. We only penalize the model to match the mean if a point's color falls outside of $[c_{10}, c_{90}]$ by at least a standard deviation.

$$L_{radial}(p) = \begin{cases} \ell_2(\Phi(p), \mu_r) & c_p - \sigma_p > c_{90} \\ \ell_2(\Phi(p), \mu_r) & c_p + \sigma_p < c_{10} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Cornea pose optimization Due to the small cornea size in the captured images, the cornea pose and normals estimate inevitably have some errors. Training with the erroneous poses significantly affects the radiance field reconstruction's quality. To alleviate the pose errors, we optimize the 6DoF pose of each cornea. For each cornea, we optimize for a transformation matrix $T = [R, t] \in \text{SE}(3)$, where
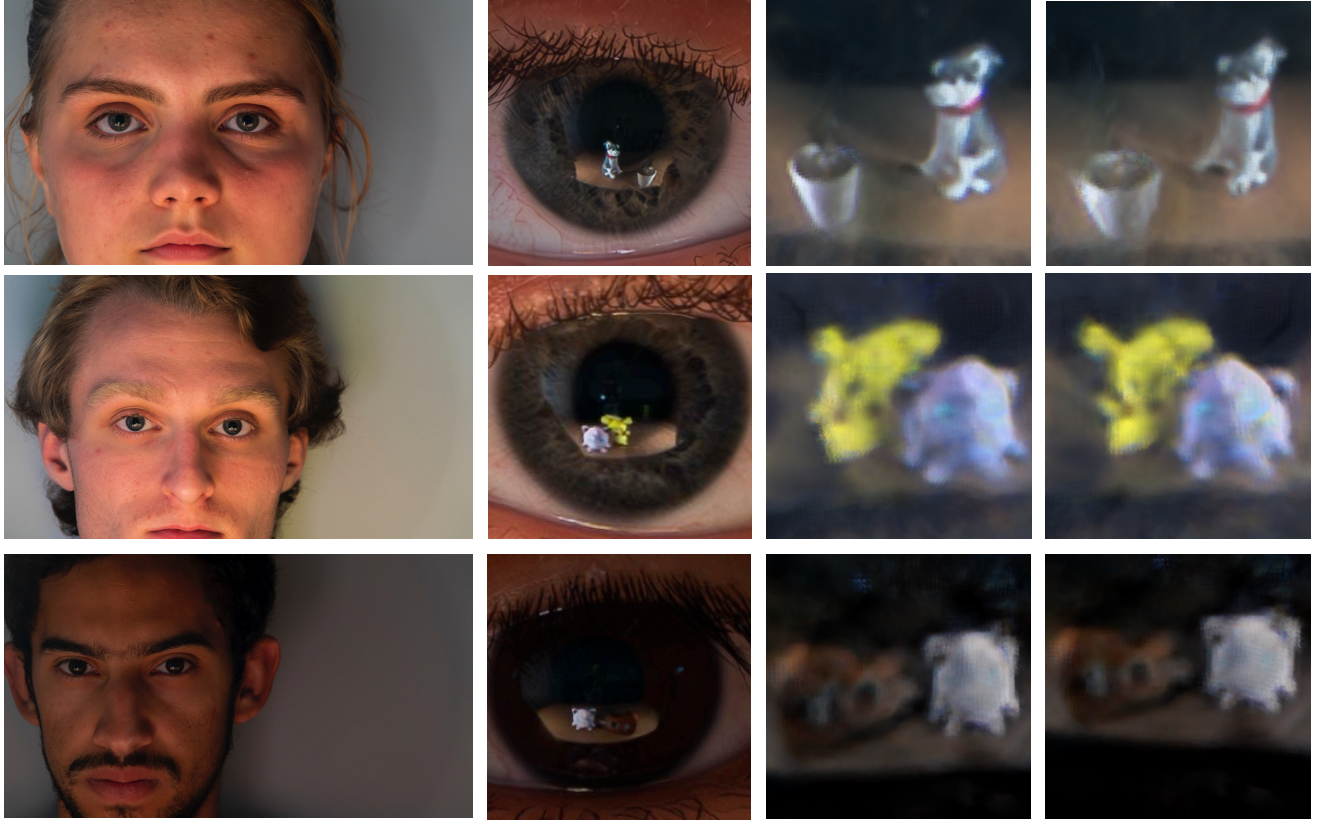
| Sample captured frame | Eye crop | Novel view rendering |

Figure 8. **Additional real results.** We show that our method works in a variety of capture conditions, like smaller objects as in the small plant on the top row, and varying eye colors.

$R \in \mathrm{SO}(3)$ and $t \in \mathbb{R}^3$ denote the rotation and translation, respectively. We optimize the cornea poses during training similar to prior work on training NeRF with noisy poses [14, 20, 49], simultaneously with the iris texture recovery and radiance field reconstruction. We initialize the translation component of the cornea poses using Equation 2 as $Z = \mathrm{depth}_{\mathrm{avg}}$, $X = x(Z/f)$, $Y = y(Z/f)$, where x and y are the normalized image coordinates of the eye center and f is the focal length of the camera. We initialize the rotation component of the cornea poses to the identity matrix. In section 5.1 we analyze how the pose optimization helps improve the novel view synthesis and ground truth eye pose recovery in a synthetic setup, and analyze its sensitivity to varying noise levels.

5. Experiments

5.1. Synthetic data evaluation

We generate synthetic data in Blender with eye models placed in the scene. Since we cannot estimate the cornea poses perfectly in real captures, we evaluate the robustness of our cornea pose optimization to the noise in the estimated cornea radius/position. We study the possible noise sources: 1) the eye center estimation, 2) the iris radius estimation, and 3) the eye rotation initialization separately. To evaluate the sensitivity of our pipeline to the eye center estimation, we corrupt the observed cornea center x coordinate and y coordinate for each imaging by scaling them with varying noise levels, which results in errors in the X and Y coordinates in the 3D poses of the corneas. For iris radius estimation, we corrupt the observed iris radius for each image by scaling the estimated radius with varying noise levels, which results in errors in the Z coordinate in the 3D poses of the corneas. For the eye rotation initialization, we initialization the true rotation of the cornea in simulation with a small amount of rotation noise. We compare the reconstruction quality and pose recovery accuracy at five noise levels for each noise source, across five different scenes, and with and without pose optimization. The quantitative results are summarized in Figure 6. Our experiments show that our reconstruction pipeline can tolerate about 1% error in the initial eye center estimate, 2.5% error in the initial radius estimate, or 5 degrees of rotation error, before quality starts to degrade significantly and pose recovery be-

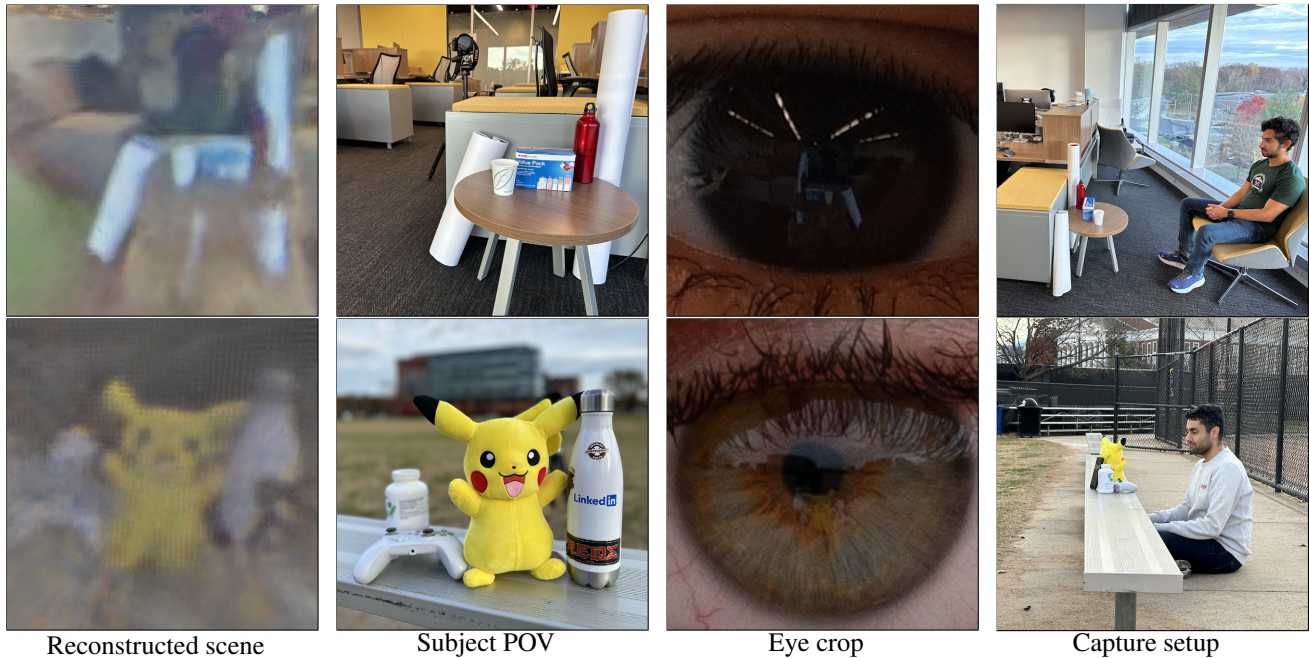| Reconstructed scene | Subject POV | Eye crop | Capture setup |

Figure 9. **Captures without structured lighting.** We show our method on an indoor office space scene and an outdoor scene in a field using natural light, and demonstrate that we can recover the scene from the eye reflection successfully without controlled illumination. Note that the photographer is also visible in the background.

comes intractable. Furthermore, we show quantitative comparisons of our method with and without iris texture decomposition in Figure 7. Our method performs better in terms of SSIM and LPIPS with texture decomposition than without. Notably, we do not compute PSNR because in our setting there is a drastic difference in intensities between the reflection and the scene itself. In the project webpage, we include synthesized novel views of our reconstructions of the synthetic scenes.

5.2. Real-world experiments

We describe capturing real-world images and demonstrate the effectiveness of our method on real captures.

Image capture. To maintain a realistic field of view, we capture images with a field of view that matches a standard portrait capture where the entire head is visible within the frame. We test our method with various illumination conditions, ranging from structured scenes with bright external lights, to unstructured indoor and outdoor settings with natural lighting. We ask the person to move within the camera's field of view and capture 5-15 frames per scene. We captured the images using a Sony RX IV camera as RAW 16-bit photos and post-processed the images using Adobe Lightroom to reduce the noise in the cornea's reflection. We vary the illumination brightness and the reflected object size for a comprehensive evaluation.

5.2.1 Data processing

We estimate the cornea's center and radius on images to get an initial estimate of the cornea's 3D location. Once we have the radius, we can directly approximate the cornea's 3D location using the average depth from Eq. 2 and the camera's focal length, and also compute its surface normals using Eq. 3. To automate the process, we locate the eyes bounding boxes using Grounding Dino [16] and then use ELLSeg [12] to perform ellipse fitting for the iris. While the corneas are typically occluded, we only need the unoccluded regions, so we obtain a segmentation mask for the iris using Segment Anything [11].

5.2.2 Results from real captures

Using our captured images, we show that our method enables the reconstruction of radiance fields from real-world portrait captures despite the cornea location and geometry inaccuracies that may arise in the real world. In Figures 1 and 8, we run our method in a controlled setting with external area lights and object-centric scenes to evaluate the best possible performance by our method and its generalization to the real-world results. In Figure 9 we show that our method can work even without controlled illumination. We find that pose optimization is the most critical compo-
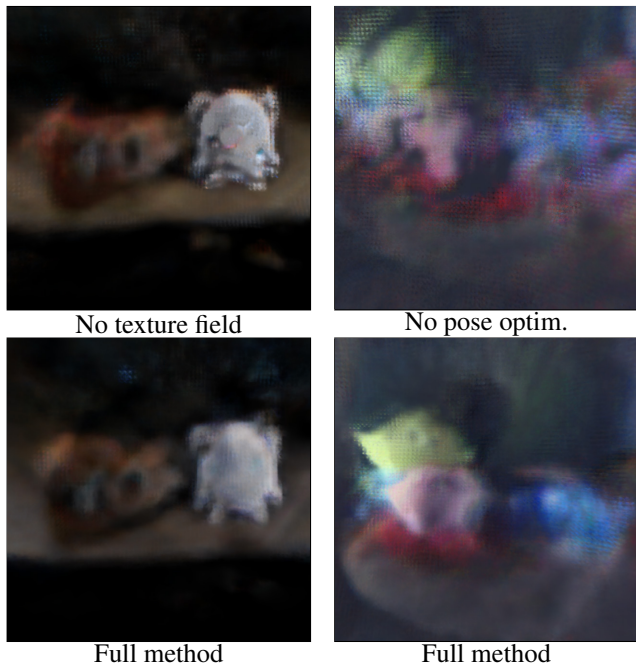
No texture field   No pose optim.

Full method   Full method

Figure 10. **Ablating texture decomposition and cornea pose optimization.** *Left*: not decomposing the iris texture can introduce artifacts in the reconstruction. *Right*: pose optimization is the most critical component for our method to reconstruct the scene.
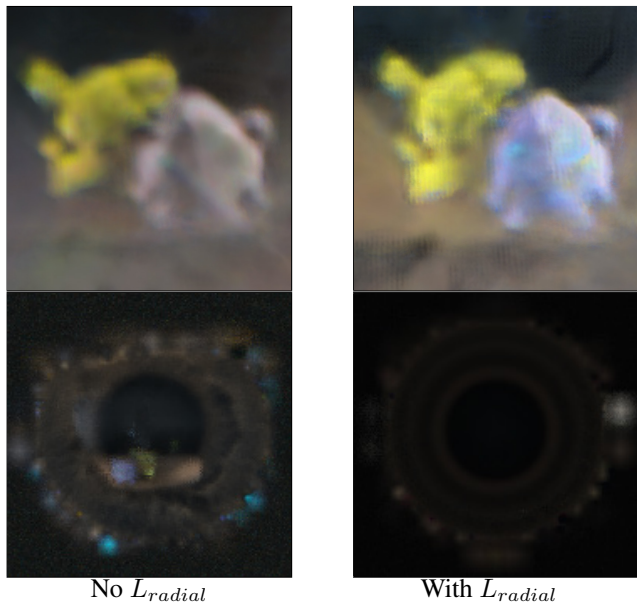

No L_{radial} With L_{radial}

Figure 11. **Ablating radial regularization.** Without radial regularization, the reconstructed iris texture contains parts of the scene with low disparity among observed views.

nent for a coherent reconstruction. Our finding validates sensitivity analysis of error in cornea poses shown in Section 5.1. In Figure 10, by ablating the cornea pose optimization and texture decomposition from our method, we demonstrate that both cornea pose optimization and texture decomposition are necessary for successful scene reconstruction. The initial cornea pose estimation is noisy because the blurriness of the cornea boundary makes it challenging to be localized precisely in the image, as shown in Figure 11. In Figure 10 we show the rendered radiance field with and without the iris texture decomposition. We notice significantly more floaters when not explicitly modeling the texture. Furthermore, Figure 11 demonstrates that the radial regularization improves the reconstruction by preventing scene regions with limited disparity of getting absorbed into the learned iris texture.

5.3. Limitations

Our work demonstrates the feasibility of reconstructing the 3D world only from eye reflections. Our current real-world results are from a "laboratory setup", such as a zoom-in capture of a person's face, and deliberate person's movement. We believe more unconstrained settings remain challenging (e.g., video conferencing with natural head movement) due to lower sensor resolution, compression, and motion blur.

6. Conclusions

By leveraging the reflections of light off human eyes, we develop a method that reconstruct the scene observed by a person using monocular image sequences captured at a fixed camera position. We demonstrate that naively training a radiance field on the observed reflections is insufficient due to several factors: 1) the inherent noise in cornea localization, 2) the complexity of iris textures, and 3) the low-resolution reflections captured in each image. To address these challenges, we introduce cornea pose optimization and iris texture decomposition during training, aided by a radial texture regularization loss based on the nature of the human iris. We validate the design choices with synthetic data and showcase view synthesis results from real-world captures of varying difficulty, including cases with and without external light and outdoor setting. With this work, we hope to inspire future explorations that leverage unexpected, accidental visual signals revealing information about the world around us, broadening the horizons of 3D scene reconstruction.

Acknowledgements

References

[1] Benjamin Attal, Jia-Bin Huang, Michael Zollhöfer, Johannes Kopf, and Changil Kim. Learning neural light fields with ray-space embedding. In *CVPR*, 2022. 3

[2] Simon Baker and Shree K. Nayar. A theory of catadioptric image formation. *ICCV*, 1998. 2

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

[4] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022. 3

[5] Wenjing Bian, Zirui Wang, Kejie Li, Jiawang Bian, and Victor Adrian Prisacariu. Nope-nerf: Optimising neural radiance field with no pose prior. In *CVPR*, 2023. 3

[6] Katherine L. Bouman, Vickie Ye, Adam B. Yedidia, Frédo Durand, Gregory W. Wornell, Antonio Torralba, and William T. Freeman. Turning corners into cameras: Principles and methods. In *ICCV*, 2017. 2, 3

[7] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *ECCV*, 2022. 3

[8] Kun Gai, Zhenwei Shi, and Changshui Zhang. Blind separation of superimposed moving images using image statistics. *IEEE TPAMI*, 2012. 3

[9] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *CVPR*, 2022. 3

[10] Rob Jenkins and Christie Kerr. Identifiable images of bystanders extracted from corneal reflections. *PloS one*, 8(12): e83325, 2013. 1

[11] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 4, 7

[12] Rakshit S Kothari, Aayush K Chaudhary, Reynold J Bailey, Jeff B Pelz, and Gabriel J Diaz. Ellseg: An ellipse segmentation framework for robust gaze tracking. *IEEE Transactions on Visualization and Computer Graphics*, 2020. 4, 7

[13] Yu Li and Michael S. Brown. Single image layer separation using relative smoothness. In *CVPR*, 2014. 3

[14] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *ICCV*, 2021. 3, 6

[15] Ruoshi Liu and Carl Vondrick. Humans as light bulbs: 3d human reconstruction from thermal reflection. In *CVPR*, 2023. 2, 3

[16] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023. 4, 7

[17] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Learning to see through obstructions. In *CVPR*, 2020. 3

[18] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Learning to see through obstructions with layered decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8387–8402, 2021. 3

[19] Yu-Lun Liu, Chen Gao, Andreas Meuleman, Hung-Yu Tseng, Ayush Saraf, Changil Kim, Yung-Yu Chuang, Johannes Kopf, and Jia-Bin Huang. Robust dynamic radiance fields. In *CVPR*, 2023. 3

[20] Andreas Meuleman, Yu-Lun Liu, Chen Gao, Jia-Bin Huang, Changil Kim, Min H. Kim, and Johannes Kopf. Progressively optimized local radiance fields for robust view synthesis. In *CVPR*, 2023. 3, 6

[21] 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. In *ECCV*, 2020. 2, 3

[22] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 2022. 3

[23] Atsushi Nakazawa and Christian Nitschke. Point of Gaze Estimation through Corneal Surface Reflection in an Active Illumination Environment. In *ECCV*, 2012. 3

[24] S. K. Nayar. Sphereo: Determining depth using two specular spheres and a single camera. In *Proceedings of SPIE Optics, Illumination, and Image Sensing for Machine Vision III*, pages 245 – 254, 1989. 2

[25] S.A. Nene and S.K. Nayar. Stereo with mirrors. In *ICCV*, 1998. 2

[26] Ko Nishino and Shree K. Nayar. Eyes for relighting. *SIGGRAPH*, 2004. 1, 3, 4

[27] Ko Nishino and Shree K. Nayar. The World in an Eye. *CVPR*, 2004. 1, 2, 3

[28] Ko Nishino and Shree K. Nayar. Corneal Imaging System: Environment from Eyes. *IJCV*, 2006. 1, 2

[29] Ko Nishino, Peter N. Belhumeur, and Shree K. Nayar. Using eye reflections for face recognition under varying illumination. In *ICCV*, 2005. 1

[30] Christian Nitschke, Atsushi Nakazawa, and Haruo Takemura. Display-camera calibration from eye reflections. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1226–1233, 2009. 3

[31] Anna Pandolfi and Federico Manganiello. A model for the human cornea: Constitutive formulation and numerical analysis. *Biomechanics and Modeling in Mechanobiology*, 5(4), 2006. 3

[32] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *ICCV*, 2021. 3

[33] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. Hypernerf: A higher-

dimensional representation for topologically varying neural radiance fields. *ACM TOG (Proc. SIGGRAPH)*, 2021. 3

[34] Keunhong Park, Philipp Henzler, Ben Mildenhall, Jonathan T. Barron, and Ricardo Martin-Brualla. Camp: Camera preconditioning for neural radiance fields. *ACM Transactions on Graphics*, 2023. 3

[35] Alexander Plopski, Yuta Itoh, Christian Nitschke, Kiyoshi Kiyokawa, Gudrun Klinker, and Haruo Takemura. Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):481–490, 2015. 3

[36] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *CVPR*, 2021. 3

[37] Abhijith Punnappurath and M. S. Brown. Reflection removal using a dual-pixel sensor. In *CVPR*, 2019. 3

[38] Prafull Sharma, Miika Aittala, Yoav Y. Schechner, Antonio Torralba, Gregory W. Wornell, William T. Freeman, and Frédo Durand. What you can learn by staring at a blank wall. In *ICCV*, 2021. 2, 3

[39] Sudipta N. Sinha, Johannes Kopf, Michael Goesele, Daniel Scharstein, and Richard Szeliski. Image-based rendering for scenes with reflections. *ACM Transactions on Graphics*, 2012. 3

[40] Kentaro Takemura, Tomohisa Yamakawa, Jun Takamatsu, and Tsukasa Ogasawara. Estimation of a focused object using a corneal surface image for eye-based interaction. *Journal of eye movement research*, 7(3):1–9, 2014. 1

[41] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Justin Kerr, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, and Angjoo Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *SIGGRAPH*, 2023. 3

[42] Kushagra Tiwary, Akshat Dave, Nikhil Behari, Tzofi Klinghoffer, Ashok Veeraraghavan, and Ramesh Raskar. Orca: Glossy objects as radiance-field cameras. In *CVPR*, 2023. 3

[43] Antonio Torralba and William T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. In *CVPR*, 2012. 2

[44] Norimichi Tsumura, Minh Dang, and Yoichi Miyake. Estimating the directions to light sources using images of eye for reconstructing 3d human face. *International Conference on Communications in Computing*, 2003. 3

[45] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 2012. 3

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

[47] Renjie Wan, Boxin Shi, Ah-Hwee Tan, and Alex Chichung Kot. Depth of field guided reflection removal. In *IEEE International Conference on Image Processing*, 2016. 3

[48] Huiqiong Wang, S. Lin, Xiaopei Liu, and Sing Bing Kang. Separating reflections in human iris images for illumination estimation. In *ICCV*, 2005. 1, 3

[49] Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. NeRF−−: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021. 3, 6

[50] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T. Freeman. A computational approach for obstruction-free photography. *ACM Transactions on Graphics*, 2015. 3

[51] Alex Yu, Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2021. 3

[52] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul E. Debevec, William T. Freeman, and Jonathan T. Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics*, 2021. 3

[53] Xiang Zhang, Kaori Ikematsu, Kunihiro Kato, and Yuta Sugiura. Reflectouch: Detecting grasp posture of smartphone using corneal reflection images. In *CHI Conference on Human Factors in Computing Systems*, 2022. 1

[54] Xuaner Cecilia Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *CVPR*, 2018. 3