

Through the Curved Cover: Synthesizing Cover Aberrated Scenes with Refractive Field

Liuyue Xie
 Carnegie Mellon University
 liuyue@andrew.cmu.edu

Jiancong Guo
 Google
 jiancong@google.com

László A. Jeni
 Carnegie Mellon University
 laszlojeni@cmu.edu

Zhiheng Jia
 Google
 lukezjia@google.com

Mingyang Li
 Google
 mingyangli@google.com

Yunwen Zhou
 Google
 verse@google.com

Chao Guo
 Google
 chaoguo@google.com

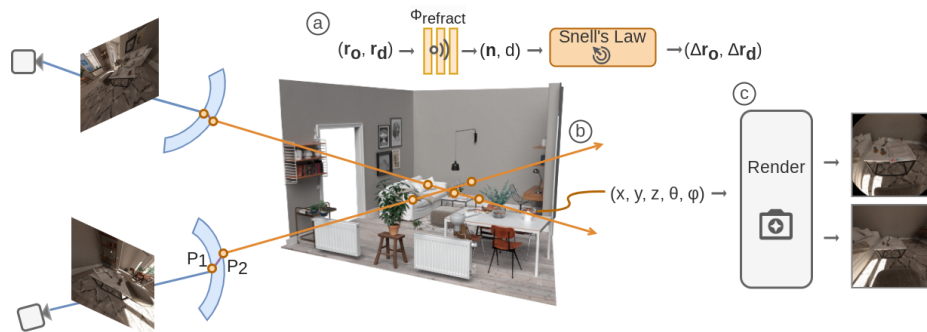


Figure 1. This figure illustrates the four modules of the proposed framework: (a) Refractive Field: Estimates surface normals and ray-to-cover distances, using Snell’s Law to calculate ray offsets. (b) Ray Sampling: Follows refracted paths, calculating positional offsets for rays. (c) Radiance Field Rendering: Renders the scene using the sampled points.

Abstract

Recent extended reality headsets and field robots have adopted covers to protect the front-facing cameras from environmental hazards and falls. The surface irregularities on the cover can lead to optical aberrations like blurring and non-parametric distortions. Novel view synthesis methods like NeRF and 3D Gaussian Splatting are ill-equipped to synthesize from sequences with optical aberrations. To address this challenge, we introduce SynthCover to enable novel view synthesis through protective covers for downstream extended reality applications. SynthCover employs a Refractive Field that estimates the cover’s geometry, enabling precise analytical calculation of refracted rays. Experiments on synthetic and real-world scenes demonstrate our method’s ability to accurately model scenes viewed through protective covers, achieving a significant improvement in rendering quality compared to prior methods. We also show that the model can adjust well to various cover

geometries with synthetic sequences captured with covers of different surface curvatures. To motivate further studies on this problem, we provide the benchmarked dataset containing real and synthetic walkable scenes captured with protective cover optical aberrations.

1. Introduction

The deployment of extended reality (XR) devices in commercial settings depends on their ability to perceive their surroundings accurately and reliably. Recently, XR headsets have been equipped with translucent covers over their front-facing cameras to enhance durability and safety. However, these protective covers present challenges for novel view synthesis due to surface irregularities that can cause image aberrations, leading to degraded synthesis quality. We propose SynthCover, a novel framework for neural view synthesis through refractive surfaces, which addresses this challenge by providing a geometric estima-

tion of surface irregularities that is jointly optimized with the synthesis process, enabling high-quality rendering even from cover-aberrated captures.

Image aberration has been a persistent issue in photography and imaging, particularly for 3D reconstruction. Common optical aberrations, such as blurring and distortions, occur even with monochromatic light sources. To mitigate these effects, specialized blind deblurring algorithms are often used to preprocess and restore image quality, while distortion correction typically relies on physical camera model approximations [1, 18, 19, 23, 31, 36], using parametric models to realign distorted pixels. Photogrammetry and novel view synthesis frameworks usually derive these parameters using Structure-from-Motion (SfM). However, these methods often struggle to generalize across the non-symmetrical distortions introduced by curved covers [29, 33, 47], a design feature increasingly common in modern XR headsets and dome cameras used in field robotics.

Surface irregularities from the polishing process of these covers can deviate from the intended geometry, causing irregular distortions and image blurring. While high-precision interferometers can measure these irregularities in a lab environment before deployment, scaling this measurement process for mass production and after deployment remains a major hurdle. Synthesizing novel viewpoints from aberrated image sequences is therefore non-trivial, as the surface irregularities introduce non-radially symmetric distortions and optical artifacts that conventional calibration techniques struggle to handle [12, 21, 29, 35, 42, 43]. These distorted features cannot be reliably matched to their warped counterparts, especially in complex scenes [21, 23, 27, 31, 46]. Recent advancements have explored embedding camera parameter tuning within reconstruction objectives, allowing for synthesis with noisy calibration parameters [24, 42] and reducing dependency on precise initializations. However, these methods struggle when camera rays deviate significantly from the assumed camera model, which is often the case with curved covers.

To address these limitations, we introduce SynthCover, an innovative framework that explicitly models the cover's geometry and adjusts the camera rays accordingly to account for optical aberrations. SynthCover effectively handles the complex distortions introduced by curved covers, surpassing the limitations of conventional camera model-based methods. By learning the cover geometry, our framework adds an additional degree of freedom for ray propagation, enabling an accurate representation of cover-induced distortions while adhering to physical refractive laws.

Our framework demonstrates superior performance compared to state-of-the-art methods, including camera-calibrating novel view synthesis frameworks, in rendering cover-aberrated sequences from real-world and simulated captures. We demonstrate that our cover geometry model-

ing approach effectively handles ray distortions in complex scenes, producing results superior to existing novel view synthesis frameworks [24, 25, 44] for modeling camera-aberrated sequences. Our contributions can be summarized as follows:

- End-to-end novel view synthesis for cover-protected cameras.
- Estimating protective-cover's surface figure geometries through ray tracing.
- Protective cover aberrated walkable scene captures for indoor and outdoor environments.

We note several assumptions in our work for estimating the surface geometries. Surface geometry estimation in optics usually involves considering spherical aberration, coma, astigmatism, curvature of field, and distortion. Since images, once captured, contain compressed light information and thus incomplete ray characteristics, we here simplify the modeling to handle two general categories of aberration: image blurring and distortions. We assume a monochromatic incoming light compatible with the ray tracing representation in the novel view synthesis methods. The material dispersion model is assumed to be constant with wavelength.

2. Related Works

Camera distortion modeling and optical aberrations.

The protective covers result in geometric optic aberration in the form of distortion, and visual aberration perceived as image blurring. In the past, camera distortion has been addressed in computer vision through the use of parameterized models tailored to specific camera types, as evidenced by a range of studies [9, 18, 22, 26, 27, 30, 34, 35, 39, 47]. These models have shown good generalization across various capturing devices. However, they fall short in accurately representing per-pixel distortions, especially with image blurring, leading to the exploration of more flexible calibration methods that can accommodate these nuances for novel view synthesis, which relies on accurate ray-to-pixel associations [20, 24, 42].

Recent advances in camera modeling have focused on estimating per-pixel distortions, addressing the limitations of traditional camera models, and accommodating sensors that deviate from conventional models [7, 31, 38, 43]. Subsequent research has extended these concepts, proposing advanced optimization strategies for imaging models that transform 3D rays into pixels and exploring the impact of different perspective distortion effects on the image plane [8, 15, 20, 39]. While these per-pixel calibration methods offer promising flexibility and generality, they introduce a significant number of parameters, often relying on sensor-specific knowledge [10, 11, 22, 26, 35, 38], which complicates the optimization process and makes them less suitable

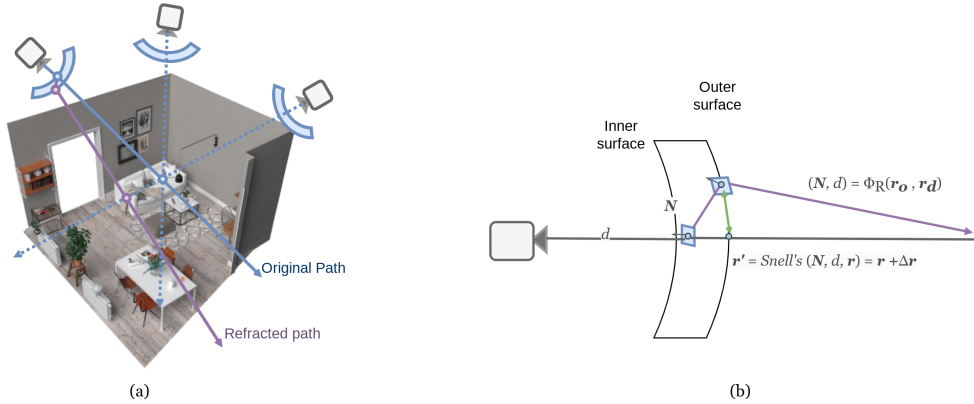


Figure 2. (a) Refraction from curved cover distorts light paths and causes rendering artifacts. (b) Our method addresses this by modeling the cover geometry and analytically bending the rays conforming to Snell’s physical refraction law.

for integration into novel view synthesis pipelines. In contrast to these approaches, our method departs from hand-crafted parameterization of camera models, focusing instead on per-ray estimation across generic captures.

Self-calibrating novel-view synthesis. Neural Radiance Fields (NeRF) have revolutionized the synthesis of novel views by leveraging a coordinate network to implicitly learn a volumetric scene function [29]. Subsequent enhancements to NeRF have been proposed, focusing on various aspects of volumetric rendering and scene reconstruction for more accurate surface and scene geometry representations [3, 4, 44, 45]. Gaussian Splatting was subsequently proposed to provide an explicit alternative to modeling the scenes with an improved rendering speed and training efficiency.

While initially focused on perspective image sequences, recent works have extended NeRF to distorted image captures [24, 42]. These approaches introduce an optimizable camera model that learns parameter offsets to the initialized parameters, enabling the rendering when SfM techniques like COLMAP [37] fail to accurately initialize camera parameters. However, these models do not specifically account for refractive mediums, which limits their effectiveness in rendering scenes looking through covers. 3DGS, on another hand, does not propagate gradients to camera intrinsic parameters as it relies on perspective rasterization. In the context of extended reality, device captures, we propose to embed cover-aberrated refraction modeling to achieve end-to-end learning on real-world captures, enabling both NeRF and 3DGS to account for the complex distortions introduced by the curved cover, leading to more flexible synthesis for challenging mobile captures.

3. Preliminary on Novel View Synthesis

Neural Radiance Field (NeRF) [29] and 3D Gaussian Splatting (3D-GS) [25] are distinct 3D scene synthesis tech-

niques from 2D images. NeRF’s implicit volumetric rendering Gaussian Splatting’s explicit α -blending operate on a similar principle. For NeRF, with color C , density σ , transmittance T , and point sampling interval δ , the color is produced with volumetric rendering along a ray as:

$$C = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) c_i, \text{ with } T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j). \quad (1)$$

As an explicit approach, Gaussian Splatting instead computes the color blending from N ordered points overlapping each pixel with their respective α values as:

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j). \quad (2)$$

For both approaches, the obtained color is contrasted against the corresponding color in the target image to refine the model’s parameters. Both techniques are adept at synthesizing three-dimensional scenes, offering comparable degrees of signal supervision to fine-tune ray deformations [25, 29]. Pose matrices $M = [R, T]$ and projection matrix P governing the outward ray directions are required to project a learned 3D canonical volume to the 2D image plane for rendering the colors. SfM methods have previously been used to provide camera estimates, but they fall short with aberrated captures. We address this by learning the protective cover geometries and accordingly correcting the outward rays. This ensures that the outward camera rays adhere to the aberrated renders, mitigating the issues of inaccurate canonical space to camera plane projections.

4. Modeling Optical Aberrations from Cover Geometry

We study the problem of novel view synthesis from a monocular cover-aberrated video sequence. The surface figure of the translucent protective cover induces image blur-



Figure 3. Sample images from the CurvedCover dataset and the capturing rig.

ring and distortions, we model these optical irregularities explicitly with a Refractive Field such that the rays coming out from the camera assembly reflect the optical surface geometry. Using the Structure-from-Motion method COLMAP, we obtain an initial estimation of camera poses, intrinsic parameters, and distortion coefficients. The camera intrinsics establish the basic geometric properties of the camera, determining the overall projection behavior of the rays, while the distortion parameters introduce non-linear modifications to the rays, bending them in a way that deviates from ideal pinhole projection. Lifting the plane coordinates \mathbf{x} to the corresponding ray \mathbf{r} , the rays originating from each camera’s viewpoint can be expressed as:

$$\mathbf{r} = f^{-1}(\mathbf{x}) = \mathbf{r}(d) = \mathbf{r}_o + d \cdot \mathbf{r}_d, \quad (3)$$

where $f(\mathbf{r})$ derived from the camera intrinsic and distortion parameters denotes the mapping of a 3D ray \mathbf{r} to a 2D image plane coordinate \mathbf{x} . $\mathbf{r}_o \in \mathbb{R}^n$ and $\mathbf{r}_d \in \mathbb{R}^n$ respectively denote the ray’s origin and direction, and d is the ray’s travel distance. Assuming that the ray’s origin is provided by SfM initialization as the camera’s world coordinate at a given frame, we seek the optimal outward ray directions \mathbf{r}_d such that the rendered image $\hat{\mathcal{I}}(\mathbf{r})$ matches the input image $\mathcal{I}(\mathbf{r})$.

Recognizing the potential inaccuracy of SfM-derived camera intrinsic and distortion parameters, we employ a differential fine-tuning strategy. This process optimizes the parametric camera model until the remaining optical aberrations are solely attributed to the curved cover.

The curved cover optical aberrations are usually caused by a thick cover’s refractions and surface shape irregularities. When the cover surfaces depart from flatness, a collimated beam would expand or focus and lead to deviations from rectilinear projections. While intuitively, the non-parametric per-ray aberrations can be modeled by learning ray offsets with a ray Refractive Field, directly predicting the distortions with each individual ray is inherently ill-posed since it essentially demands the network to implicitly understand both the geometry of the medium and the physical behavior of Snell’s refractive law. In our proposed SynthCover model, we instead decouple the learning objective of the Refractive Field by modeling the physical geometry of the refractive medium with a network and then analyti-

cally bending the lights according to the learned geometry through Snell’s Law.

Decoupling the Refractive Field objective alleviates the network’s learning demand and adds transparency to the optimization. The mathematical formulations of the differentiable parametric camera model, Refractive Field network, ray sampling, and regularizations are detailed in the following sections. The implementation details are provided in Appendix Sec. A.2.

4.1. Differentiable Parametric Camera Model

We adopt a differentiable Brown-Conrady model to describe the camera without protective cover aberrations. We note that when the optical power induced by the protective cover is zero, the capturing system would resort to Brown’s camera model. The camera model consists of intrinsic parameters describing the optical center and focal lengths, as well as distortion parameters describing the deviations from an ideal pinhole camera. Directly learning the camera intrinsics \mathbf{K} leads to a highly non-convex optimization problem [24]. We instead decompose the camera intrinsics into an initialized \mathbf{K}_o matrix and learnable offset parameter matrix $\Delta\mathbf{K}$ such that the initialized intrinsics reduce the local minima that can compromise the learning process. The revised intrinsic matrix can thus be expressed as $\mathbf{K} \in \mathbb{R}^{3 \times 3} = \mathbf{K}_o + \Delta\mathbf{K}$ where the norm of $\Delta\mathbf{K}$ would be bounded. Similarly, the distortion model \mathcal{D} consists of the initial parameter estimates and offsets to be refined through optimization.

The camera model describes the mapping between points on the camera sensor $\mathbf{x} \in \mathbb{R}^{2 \times 2}$ to their 3D locations $\mathbf{X} \in \mathbb{R}^{3 \times 3}$. To express a camera with a ray bundle, we define the ray origins as the camera’s location in the world, and we lift the points from the camera sensor to their locations in the world coordinate, then subtract the camera centers to get the ray directions. Deriving the lifting from pixels to rays, the rays in the camera coordinate system are expressed as $\mathbf{r}_d(\mathbf{p}) = \mathcal{D}(\mathbf{K}^{-1}\mathbf{p})$ with $\mathbf{r}_o = 0$ as rays originating from the camera’s position. The extrinsic \mathbf{R} and \mathbf{T} can be used to transform the rays from the local camera coordinate to the world coordinate as: $\mathbf{r}_d = \mathbf{R} \cdot \mathcal{D}(\mathbf{K}^{-1}\mathbf{p}) + \mathbf{T}$; $\mathbf{r}_o = \mathbf{T}$.

Detailed formulation of the differentiable camera model can be found in Appendix Sec. A.5.

4.2. Refractive Medium Estimation

We design a network to model the refractive medium $\Phi_R(\mathbf{r}_o, \mathbf{r}_d) = (d, \mathbf{n})$ to estimate the cover surface geometries. The Refractive Field network takes ray direction vectors, \mathbf{r}_o and \mathbf{r}_d , as inputs and then predicts the distance to the two incident surfaces as well as the incident normal. The predicted distances describe the travel distance of each ray to both the cover surfaces. This is estimated to track the incident locations of the rays when the placement of the protective cover and thickness of the cover are both unknown. The outputs from the distortion network are used to analytically compute ray offsets, which are subsequently added to the corresponding rays before they are projected to the world space for rendering. These offsets implicitly account for the distortions introduced by the refractive medium. We leverage an invertible neural network (INN) architecture for this network since it can effectively compute both the forward transformation (distortion) and its inverse (undistortion). This invertibility allows the network to model the complex, two-way relationship between the undistorted and distorted rays.

Since a lens generally has two refractive surfaces, we assume that the rays refract exactly twice, with the first refraction passing the rays from the air to the covered inner surface, and the second refraction passing the rays through the outer surface back to the air. We neglect considerations of self-occlusions on the refractive path to constrain the problem. The estimation of outward rays is detailed as follows.

For each time step in the sequence, the previously derived rays from the camera are sent to the Refractive Field network Φ_R . The network outputs the travel distance $d \in \mathbb{R}^{n \times 2}$ of each ray to two surfaces \mathcal{S}_1 and \mathcal{S}_2 and the corresponding incident normal vectors $\mathbf{n} \in \mathbb{R}^{n \times 2}$. With the estimated geometry, a ray \mathbf{r} refracts through the first surface at \mathbf{X}_{s1} with the estimated incident normal using Snell’s Law which governs the angle of light bending when passing across two mediums with different indexes of refraction (z_1, z_2) .

We use a vector form of Snell’s Law for the refraction computation. In the vector form notation, $\mathbf{r} \in \mathbb{R}^3$ denotes the incident ray, $\mathbf{n} \in \mathbb{R}^3$ denotes the normal vector, and $\mathbf{r}' \in \mathbb{R}^3$ is the transmitted ray. The Snell’s Law can then be expressed as:

$$\mathbf{r}' = \eta(\mathbf{r} + c_1 \mathbf{n}) - c_2 \mathbf{n}, \quad (4)$$

where $c_1 = \langle \mathbf{n}, \mathbf{r} \rangle$, $c_2 = \sqrt{1 - \eta^2(1 - c_1^2)}$ are two compressed terms, and $\eta = \frac{z_1}{z_2}$ denotes the ratio between two medium’s refractive indices. The refracted ray directions through the first stage can thus be computed accordingly, with the rays originating from the cameras defining the in-

cident rays. Similarly, the ray passing through the second surface \mathcal{S}_2 of the curved cover to exit the refractive medium can be computed again using Snell’s Law. The resultant rays outward from the second surface are then used for subsequent rendering.

Directly estimating the distances and normal vectors from an unconstrained network is ill-posed, since the predicted cover geometry through distance estimation may deviate from the surface geometry described by the normal predictions. To ensure that the network embeds the correct geometry of the cover, we regularize the distance and normal estimations with a consistency loss detailed in Section 4.4.

4.3. Sampling and Rendering Through Refractive Cover

The captured scene is distributed along the direction outward from the cover’s outer surface. With the surface approximated with the estimated intersection \mathbf{X}_{s2} , the ray $\mathbf{r}'(d)$ beyond the outer refractive surface can be expressed as:

$$\mathbf{r}'(t) = \mathbf{X}_{s2} + d \cdot \mathbf{r}'_d. \quad (5)$$

For each ray in NeRF synthesis, we sample points \mathbf{x} along $\mathbf{r}'(d)$ then feed the sampled 3D points to the Radiance Field for further prediction of the volumetric scene. With the rays accounting for the refracted offsets, the modified NeRF can now be optimized directly through the aberrated captures.

Synthesizing 3D Gaussian Splatting (3DGS) scenes requires slight modifications to the optimization as a rasterization method. 3DGS is initialized with a sparse point cloud from SfM, and it necessitates rectifying the training views to align with the perspective projection in the rasterization process. To adapt our approach to 3DGS’s formulation, we rectify the training views, modify the training objective to correct residual optical aberrations, and include parameter estimates in addition to the ray offsets. The Gaussians are represented by parameters $(\mu, \mathbf{R}, \mathbf{S}, \mathbf{sh})$, which denote the Gaussian’s mean, rotation, scale, and ray-dependent spherical harmonics, respectively. Since the Gaussian parameters, excluding the Spherical Harmonics terms, are not directly associated with the camera ray directions, we introduce two additional layers to capture Gaussian deformations caused by distorted rays. The parameter offsets $(\Delta\mu, \Delta\mathbf{R}, \Delta\mathbf{S})$ are estimated by the deformation network, while $\Delta\mathbf{sh}$ is derived from the ray offsets. These offsets are applied to the canonical Gaussians before rendering, and the rendered results are compared with the rectified image.

4.4. Regularization

We optimize the differentiable camera parameters and distortion field with photometric loss and normal consistency loss. The photometric loss compares the 2D image

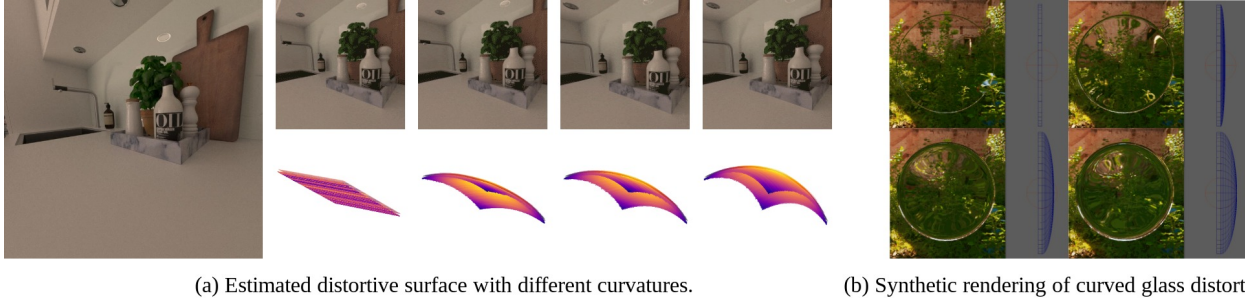


Figure 4. (a) Surface reconstruction of the curved cover. (b) Simulation setup for recreating the distortion induced by the curved cover.

Table 1. **Quantitative comparison on the FisheyeNeRF dataset and EyefulTower dataset.** See Sec 5.1 for an analysis of the performance.

method	globe			cube			table			sofa area		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRF	19.3	0.472	0.731	18.5	0.534	0.672	21.5	0.788	0.333	21.7	0.672	0.364
SCNeRF	23.8	0.598	0.633	23.2	0.605	0.616	22.9	0.728	0.275	25.1	0.769	0.348
NeuroLens	<u>24.5</u>	<u>0.724</u>	<u>0.148</u>	<u>25.3</u>	<u>0.716</u>	<u>0.163</u>	<u>23.4</u>	<u>0.817</u>	<u>0.260</u>	<u>27.8</u>	<u>0.837</u>	<u>0.189</u>
Ours	28.4	0.730	0.124	27.3	0.728	0.147	27.2	0.833	0.189	29.2	0.859	0.136

Table 2. **Quantitative comparison on the CurvedCover dataset.** See Sec 5.1 for an analysis of the performance.

method	robot arm			scotty			statue			coffee table		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRF	16.1	0.711	0.8176	16.9	<u>0.7346</u>	0.8682	22.1	0.6082	0.6177	23.4	0.7678	0.5977
SCNeRF	25.9	<u>0.8761</u>	<u>0.2664</u>	25.4	0.6065	0.5890	<u>26.8</u>	0.7187	0.3197	29.5	0.8473	0.4103
NeuroLens	<u>26.5</u>	0.7799	0.3138	<u>25.6</u>	0.6463	<u>0.4455</u>	26.6	<u>0.7897</u>	<u>0.2483</u>	<u>32.6</u>	<u>0.8926</u>	<u>0.3280</u>
Ours	28.8	0.9287	0.1564	26.8	0.8972	0.2649	27.3	0.8137	0.2121	32.9	0.9048	0.2873

observations and the rendered results. It builds on the intuition that if the camera forms the images correctly, the color of the rendered pixels from each ray \mathbf{r} modeled by the learnable terms should match its projected 3D counterpart in the reference image as:

$$\mathcal{L}_{\text{photometric}} = \sum_{\mathbf{r} \in \mathcal{N}} \|\hat{\mathcal{I}}(\mathbf{r}) - \mathcal{I}(\mathbf{r})\|_2^2 \quad (6)$$

Besides the photometric differences between the rendered image and the captured image, we also minimize the difference between the predicted normals \mathbf{n}_i from the Refractive Field and the fitted local normals from the predicted distances \mathbf{n}_{d_i} . The normal consistency loss is defined as follows:

$$\mathcal{L}_{\text{normals}} = \sum_{\mathbf{r} \in \mathcal{N}} \|\mathbf{n}_{d_r} - \mathbf{n}_r\|_2^2, \quad (7)$$

where \mathcal{N} denotes the collection of ray paths that cross the refractive medium. We fit the local normals in (3×3) neighborhoods from the predicted distances to surfaces d as \mathbf{n}_d then regularize the network’s direct normal estimations \mathbf{n} with the fitted normals.

5. Experiments

In this section, we validate our proposed method on four static scene datasets captured with different levels of cam-

era distortions. We demonstrate the quantitative and qualitative results achieved on the LLFF dataset [28], EyefulTower dataset [44], FisheyeNeRF dataset [24], and our own CurvedCover dataset. Notably, our method achieves state-of-the-art performance across both cameras without covers and those equipped with curved covers.

5.1. Evaluation

This section presents a quantitative and qualitative analysis of our method’s performance. Our aim is to investigate whether explicit cover modeling helps disambiguate ray bending paths and enhance the reconstruction of distorted scenes. We also explore whether the proposed camera model potentially hinders the reconstruction of scenes captured without a cover. For implementation details, please see Appendix Sec. A.2.

Benchmarked datasets. We evaluate our method using both real-world and synthetic monocular datasets with varying camera distortions. The real-world datasets include LLFF [28] with minimal distortion, EyefulTower [44] captured with fish-eye camera in office scenes, FisheyeNeRF [24] with wide-angle outdoor scenes, and our CurvedCover dataset captured by dual fish-eye cameras with a curved

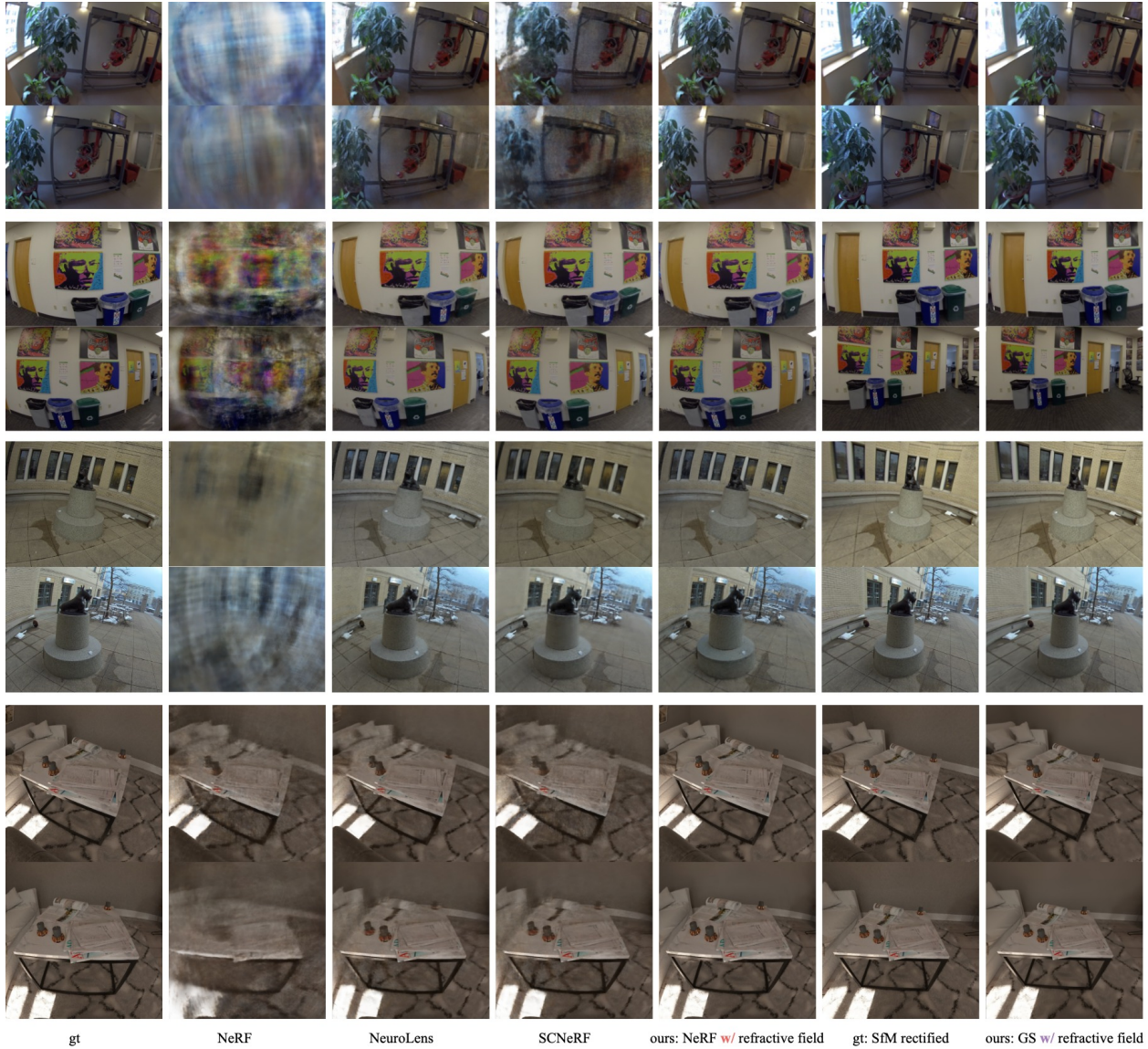


Figure 5. This figure compares NeRF variants, including ours. Ground truth images are shown in the first column. All methods use RealityCapture for initial calibration. Our approach shows consistent accuracy across datasets.

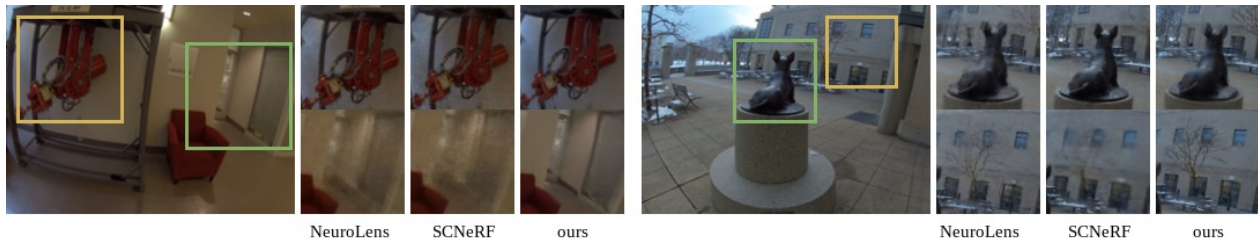


Figure 6. This figure compares NeRF variants, including ours. Ground truth images are shown in the first column. All methods use RealityCapture for initial calibration. Our approach shows consistent accuracy across datasets.

Apple Vision Pro front-facing cover as shown in Fig. 3. We also include synthetic scenes with varying cover curvatures to study the extent of our proposed cover modeling

approach. Detailed explanations of the dataset can be found in Appendix Sec. A.1.

Quantitative baseline comparisons. We evaluate

our method’s novel-view synthesis capabilities on LLFF, FisheyeNeRF, EyefulTower, and CurvedCover datasets, ordered by ascending camera distortion levels. These datasets present increasing challenges, with the CurvedCover dataset posing the greatest difficulties due to its external cover and scene complexities. We assess visual quality using LPIPS [48], SSIM [41], and PSNR [32], comparing our method to NeRF, SCNeRF, NeuroLens, as shown in Table 1 and 2.

Table 3. **Quantitative comparison on the LLFF dataset.** See Sec 5.1 for an analysis of the performance.

Scene	Model	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Flower	NeRF	32.2	0.937	0.067
	Ours	33.1	0.951	0.062
Trex	NeRF	31.4	0.955	0.099
	Ours	32.3	0.963	0.094

Our method consistently outperforms these baselines on scenes with greater distortions (FisheyeNeRF, EyefulTower, and CurvedCover). These comparisons highlight the effectiveness of our holistic camera distortion modeling compared to other camera-calibrating synthesis frameworks. To further assess and ensure that our method does not hinder novel-view synthesis quality with sufficiently calibrated datasets, we compare its results with the baselines on the LLFF dataset, as shown in Table 3. Notably, our method does not hinder the innate novel view synthesis capability as the cover geometry essentially resorts back to a flat surface when the camera does not have a curved cover and is sufficiently calibrated. Finally, we support the aforementioned decomposed geometric Refractive Field learning objective as in Table 4. For more details on the implementation and experiments on the 3DGS model, see Appendix Sec. A.3 and Sec. A.4.

Qualitative evaluations on distorted captures. We compared our method with NeRF, SCNeRF and NeuroLens on both synthetic and real curved cover datasets as shown in Fig. 5. The vanilla NeRF was unable to synthesize distorted captures, as consistent with previous findings in the literature [24]. Our approach achieves superior rendering quality due to its ability to explicitly model the curved cover geometry. This is particularly evident in scenes with significant refractive distortions, where the baseline methods struggle to accurately capture the light behavior, as shown in Fig. 6. With the modeled Refractive Field, the cover geometry can be recovered as visualized in Fig. 4.

Refractive Field representations. We conduct ablation studies on the synthetic “Coffee Table” sequence in the CurvedCover dataset to dissect the impact of the warm-up phase and Refractive Field, as detailed in Table 4. End-to-end training solely on distortion parameters highlighted the need for a warm-up phase of 6000 iterations, to avoid nu-

Table 4. Ablation studies on the warmup process, Refractive Field, deformation field, and generalizability to the 3D-GS model.

Model	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
ours-NeRF	32.9	0.904	0.287
ours-GS	33.4	0.896	0.262
– warm up	25.3	0.786	0.442
– Refractive Field	29.4	0.861	0.307

merical instabilities. To probe the necessity for explicit geometric estimation of the Refractive Field, we directly output the ray offsets by modifying the Refractive Field’s output layer to produce Δr . Our findings confirm that the geometric Refractive Field is crucial for accurate light modeling. It alleviates the network’s burden of simultaneously learning both the cover geometry and the physics of refraction, leading to more efficient and effective scene reconstruction.

6. Limitations and Future Work

Our method is primarily limited in its ability to handle specular effects from the protective cover since the Refraction Field only accounts for the primary refracted path for each ray. The specular reflections are present when ambient lights are reflected on the covers. Scenes with strong specular effects are generally not well-captured by NeRF frameworks without explicitly extending its ray tracing to capture higher-order light paths from the ambient environment, as discussed in Eikonal Field [6]. For enabling specular reflections in splatting methods, additional attributes to handle specular surfaces could be amended to the 3D Gaussian color parameterization.

7. Conclusion

In this study, we introduce SynthCover to enable novel view synthesis with sequences aberrated by protective covers, a common design choice adopted by extended reality headsets and field robots for durability. Our approach fits into existing novel view synthesis pipelines, enabling them to model sequences with cover-induced optical aberrations. Alongside this, we present a new dataset featuring both synthetic and real captures with aberrative covers to facilitate further research on this topic. Our evaluations, utilizing our dataset and existing novel view synthesis datasets, validate our model’s broad applicability across various cover geometries. This underscores its potential as a distortion rendering tool for future robotic novel view synthesis applications. SynthCover is currently limited to handle constant material dispersion with monochromatic lights, without consideration of specular effects from the cover. For future work, we plan on extending our work to handle other optic effects caused by the cover such as specular reflections and scattering for more general applications.

References

- [1] Marina Alterman, Yoav Y. Schechner, and Yohay Swirski. Triangulation in random refractive distortions, 2017. **2**
- [2] Arashi Vision Inc. Insta360Pro 2, 2024. **11**
- [3] Benjamin Attal, Jia Bin Huang, Michael Zollhofer, Johannes Kopf, and Changil Kim. Learning Neural Light Fields with Ray-Space Embedding. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, 2022. **3**
- [4] 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 *Proceedings of the IEEE International Conference on Computer Vision*, 2021. **3**
- [5] Jens Behrmann, Will Grathwohl, Ricky T.Q. Chen, David Duvenaud, and Jörn Henrik Jacobsen. Invertible residual networks. In *36th International Conference on Machine Learning, ICML 2019*, volume 2019-June, 2019. **11**
- [6] Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. Eikonal fields for refractive novel-view synthesis. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*, pages 1–9, 2022. **8**
- [7] Paul Bergmann, Rui Wang, and Daniel Cremers. Online Photometric Calibration of Auto Exposure Video for Realtime Visual Odometry and SLAM. *IEEE Robotics and Automation Letters*, 3(2):627–634, 4 2018. **2**
- [8] alexandre Bousaid, Theodoros Theodoridis, Samia Nefti-Meziani, and Steve Davis. Perspective Distortion Modeling for Image Measurements. *IEEE Access*, 8, 2020. **2**
- [9] BROWN DC. Close-range camera calibration. *Photogramm Eng*, 37(8), 1971. **2, 12**
- [10] Federico Camposeco, Torsten Sattler, and Marc Pollefeys. Non-parametric Structure-Based Calibration of Radially Symmetric Cameras. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2192–2200. IEEE, 12 2015. **2**
- [11] David Claus and Andrew W. Fitzgibbon. A rational function lens distortion model for general cameras. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, volume I, 2005. **2**
- [12] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997*, 1997. **2**
- [13] Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear independent components estimation. In *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*, 2015. **11**
- [14] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017. **11**
- [15] Aubrey K. Dunne, John Mallon, and Paul F. Whelan. Efficient generic calibration method for general cameras with single centre of projection. *Computer Vision and Image Understanding*, 114(2), 2010. **2**
- [16] Epic Games Inc. RealityCapture, 3 2024. **11**
- [17] Christian Etmann, Rihuan Ke, and Carola Bibiane Schonlieb. IUNets: Learnable invertible up-and downsampling for large-scale inverse problems. In *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, volume 2020-September, 2020. **11**
- [18] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 588 LNCS, 1992. **2**
- [19] Ricardo Ferreira, João P. Costeira, and João A. Santos. Stereo reconstruction of a submerged scene. In *Lecture Notes in Computer Science*, volume 3522, 2005. **2**
- [20] Andrew W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, 2001. **2**
- [21] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer. Generation of fiducial marker dictionaries using Mixed Integer Linear Programming. *Pattern Recognition*, 51, 2016. **2**
- [22] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, 2001. **2**
- [23] Richard I. Hartley. Kruppa’s equations derived from the fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), 1997. **2**
- [24] Yoonwoo Jeong, Seokjun Ahn, Christopher Choy, Animesh Anandkumar, Minsu Cho, and Jaesik Park. Self-Calibrating Neural Radiance Fields. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021. **2, 3, 4, 6, 8, 12**
- [25] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4), 2023. **2, 3, 11**
- [26] Einari Kilpelä. Compensation of systematic errors of image and model coordinates. *Photogrammetria*, 37(1), 1981. **2**
- [27] Seon Joo Kim and Marc Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4), 2008. **2**
- [28] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 2019. **6**
- [29] 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 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12346 LNCS, 2020. **3, 11**

- [30] David Nistér, Henrik Stewénus, and Etienne Grossmann. Non-parametric self-calibration. In *Proceedings of the IEEE International Conference on Computer Vision*, volume I, 2005. [2](#)
- [31] Marc Pollefeys and Luc Van Gool. Stratified self-calibration with the modulus constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8), 1999. [2](#)
- [32] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, and Marco Carli. Modified image visual quality metrics for contrast change and mean shift accounting. In *2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, pages 305–311, 2011. [8](#)
- [33] Srikumar Ramalingam and Peter Sturm. A Unifying Model for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7), 2017. [2](#)
- [34] Srikumar Ramalingam, Peter Sturm, and Suresh K. Lodha. Towards complete generic camera calibration. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, volume I, 2005. [2](#)
- [35] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, ICVS'06*, volume 2006, 2006. [2](#)
- [36] Johannes L. Schonberger and Jan Michael Frahm. Structure-from-Motion Revisited. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, 2016. [2](#)
- [37] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [3](#)
- [38] Thomas Schops, Viktor Larsson, Marc Pollefeys, and Torsten Sattler. Why Having 10,000 Parameters in Your Camera Model Is Better Than Twelve. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. [2](#)
- [39] P Sturm and S Ramalingam. A Generic Calibration Concept: Theory and Algorithms. *Technical Report*, 2003. [2](#)
- [40] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David Mcallister, Justin Kerr, and Angjoo Kanazawa. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *Proceedings - SIGGRAPH 2023 Conference Papers*, 2023. [11](#)
- [41] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. [8](#)
- [42] Wenqi Xian, Aljaž Božič, Noah Snavely, and Christoph Lassner. Neural Lens Modeling. 2023. [2](#), [3](#)
- [43] Jinhui Xiong and Wolfgang Heidrich. In-the-Wild Single Camera 3D Reconstruction Through Moving Water Surfaces. In *Proceedings of the IEEE International Conference on Computer Vision*, 2021. [2](#)
- [44] Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder, Aljaž Božič, Dahua Lin, Michael Zollhöfer, and Christian Richardt. VR-NeRF: High-Fidelity Virtualized Walkable Spaces. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–12, New York, NY, USA, 12 2023. ACM. [2](#), [3](#), [6](#)
- [45] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-NeRF: Point-based Neural Radiance Fields. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, 2022. [3](#)
- [46] Mustafa B. Yaldiz, Andreas Meuleman, Hyeonjoong Jang, Hyunho Ha, and Min H. Kim. DeepFormableTag: End-to-end generation and recognition of deformable fiducial markers. *ACM Transactions on Graphics*, 40(4), 2021.
- [47] Cyril Zeller and Olivier Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. *Rapports de recherche-INRIA*, 1996. [2](#)
- [48] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. [8](#)