Eikonal Fields for Refractive Novel-View Synthesis

Mojtaba Bemana
MPI Informatik
Germany
mbemana@mpi-inf.mpg.de

Karol Myszkowski
MPI Informatik
Germany
karol@mpi-inf.mpg.de

Jeppe Revall Frisvad
Technical University of Denmark
Denmark
jerf@dtu.dk

Hans-Peter Seidel
MPI Informatik
Germany
hpseidel@mpi-inf.mpg.de

Tobias Ritschel
University College London
UK
t.ritschel@ucl.ac.uk

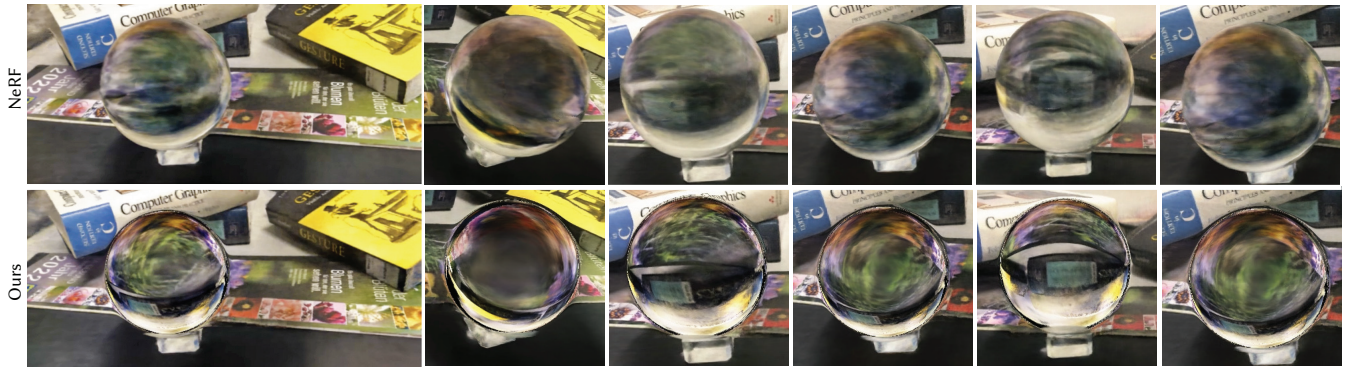

Figure 1: Novel-view synthesis using neural radiance fields (top) and our eikonal approach (bottom) for a real refractive scene.

ABSTRACT

We tackle the problem of generating novel-view images from collections of 2D images showing refractive and reflective objects. Current solutions assume opaque or transparent light transport along straight paths following the emission-absorption model. Instead, we optimize for a field of 3D-varying index of refraction (IoR) and trace light through it that bends toward the spatial gradients of said IoR according to the laws of *eikonal* light transport.

CCS CONCEPTS

• **Computing methodologies** → **Image-based rendering**.

KEYWORDS

refraction; deep learning; eikonal rendering

ACM Reference Format:
Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. 2022. Eikonal Fields for Refractive Novel-View Synthesis. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings (SIGGRAPH '22 Conference Proceedings), August 7–11, 2022, Vancouver, BC, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3528233.3530706

SIGGRAPH '22 Conference Proceedings, August 7–11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9337-9/22/08.
https://doi.org/10.1145/3528233.3530706

1 INTRODUCTION

Given images with different views of a refractive object, it is a challenging task to synthesize a novel view. The issue is that the refractive object takes its appearance from the surroundings by bending and internally reflecting the rays of light that travel through the object. If we fully digitize the object and its surroundings, we can synthesize novel views [Trifonov et al. 2006; Hullin et al. 2008; Ihrke et al. 2010; Stets et al. 2017], but this approach requires a lot more information than a simple set of images. We would for instance need a dedicated hardware setup to digitize a transparent object [Ihrke et al. 2010; Stets et al. 2017; Lyu et al. 2020]. Deep learning offers an alternative approach where we can instead render transparent objects and train a neural network to estimate the shape of such objects in more arbitrary surroundings [Stets et al. 2019; Sajjan et al. 2020; Li et al. 2020]. A deep learning technique based on a synthetic dataset, however, often returns a faulty estimate when presented with an image significantly different from those in the training data [Lyu et al. 2020].

We would like to avoid the difficulties in representing a wide enough range of transparent object appearances in one synthetic dataset. One way to do this is to learn the radiance field of a given object based on a set of images capturing its appearance as observed from different directions [Lombardi et al. 2019; Mildenhall et al. 2020]. This is useful for locating and estimating the distance to transparent objects [Ichnowski et al. 2021]. However, since neural radiance fields do not consider refraction, we can not use it out of the box for refractive novel-view synthesis (NVS).

To enable this, we devise a method that optimizes for the field of 3D spatially-varying index of refraction (IoR) given a set of 2D

images picturing a refractive object. Existing solutions to learn 3D fields capturing scene geometry are based on opaque or transparent light transport along straight paths. In the presence of transparent objects, however, light bends, i.e., it changes its direction. The precise way in which light paths are curved depends on a certain *eikonal equation* operating on spatial gradients of the IoR field, which we show can be solved – and differentiated over in learning – in practice with the appropriate formulation. The resulting method allows for novel-view synthesis (Fig. 1) in 3D scenes with complex objects involving strong refractive and internal reflection effects. Our code and training data are publicly available at https://eikonalfield.mpi-inf.mpg.de/.

2 RELATED WORK

As our focus is novel-view synthesis for refractive transparent objects, we discuss this problem with an emphasis on recent neural rendering solutions that can handle specular effects (Sec. 2.1). We overview also image-based modeling of transparent objects (Sec. 2.2), which is a more general setup than we require in this work, but still some similarities can be found. Finally, we discuss physics-based eikonal rendering that inspired our work (Sec. 2.3).

2.1 Novel-view synthesis

Recent learning-based approaches allow synthesizing images under new views without accurately reconstructing the physical parameters [Tewari et al. 2020, 2021].

Neural radiance fields (NeRF). Mildenhall et al. [2020] introduce a volumetric opacity representation that encodes both geometry and appearance using a multi-layered perceptron (MLP) and is trained on a large set of multiple-view RGB images and proves to be extremely successful in novel-view synthesis. More specifically, view-dependent RGB color and view-independent density are learned as sharp functions in space and smooth functions in angle. In the case of near-mirror or near-glass reflection/refraction, appearance cannot be described as a smooth function of angle anymore [Guo et al. 2021]. Consequently, reflected or refracted patterns appear notoriously ghosty and blurry [Ichnowski et al. 2021]. A number of solutions exist [Zhang et al. 2021b; Boss et al. 2021] that disentangle normal vectors and spatially-varying reflectance by manipulating the neural radiance fields (NeRF) density representation, but highly specular surfaces with clearly visible reflected environment cannot be reproduced. Our approach does not fully rely on the NeRF geometry and uses the diffuse scene only as a backdrop.

Inspired by traditional image-based rendering [Sinha et al. 2012; Xu et al. 2021] for scenes with planar reflections, Guo et al. [2021] introduce NeRFReN, where an additional NeRF structure is proposed that renders a reflected image and composes it additively with the traditional NeRF rendering. MirrorNeRF [Wang et al. 2021] employs a catadioptric imaging system based on an array of hemispherical mirrors enabling a single-shot portrait reconstruction and rendering. The position of a sample point is warped while its view direction is unchanged. In our approach, the view directions bend.

Other volumetric representations. Lombardi et al. [2019] and others [Yu et al. 2021a,b] propose a voxel grid with interpolation optimized using a CNN (or a gradient method directly) that encodes both geometry (possibly dynamic) and appearance. Lombardi et al. [2019] perform learned warping to reduce memory requirements and to improve the resolvable details. We instead warp space according to physical laws that regularize the problem. Even sharper mirror reflection and transparency effects can be obtained using an extension of the multi-plane images (MPI) representation, where for every pixel in a stack of semi-transparent planes, directional information using learned basis functions is stored [Wizadwongsa et al. 2021]. Similarly, as for other MPI-based and neural light-field methods [Bemana et al. 2020; Attal et al. 2021], only narrow baselines are supported.

A signed distance field, possibly encoded into an MLP, can represent surface geometry and help recovering non-spatially-varying reflectance using spherical Gaussians that in turn enable good quality reflections [Zhang et al. 2021a]. In an alternative point-based representation [Kolos et al. 2020], where each point is associated with a learnable photometric, geometric, and transparency descriptor, relatively sharp depiction of semi-transparent objects is achieved, when the non-distorted background is also known. However, specular effects are explicitly excluded from the training data.

In all those solutions, an important limiting factor is a straight-path assumption in the rendering formulation that neglects light reflection and refraction effects. In our work, we additionally reconstruct a volumetric IoR field that along with simulating the laws of physics associated with refractive effects enables us to explain the input RGB images during learning, and consequently provides a meaningful synthesis of novel views at the test time.

2.2 Transparent surface reconstruction

For a survey on reconstruction of shape, illumination and materials of transparent objects, see Ihrke et al. [2010].

Kutulakos and Steger [2008] investigate two-interface refractive light interaction with a surface, and for every pixel recover multiple 3D points, so that a ray exiting the surface can be reconstructed.

Environment matting techniques solve for a background deformation by a transparent object, so that it can be composited onto different backgrounds [Zongker et al. 1999; Chuang et al. 2000; Peers and Dutré 2003; Matusik et al. 2002; Wexler et al. 2002]. Khan et al. [2006] and others [Yeung et al. 2011; Chen et al. 2019] demonstrate that even significant departs from physics can be tolerated by human perception to make such compositing look realistic.

Dedicated setups for transparent object reconstruction rely on light-field background displays [Wetzstein et al. 2011], X-ray computed tomography (CT) scanners [Stets et al. 2017], and transmission imaging [Kim et al. 2017]. In intrusive setups, which require immersing transparent objects into a liquid with matching IoR, straight light paths can be assumed greatly simplifying CT reconstruction [Trifonov et al. 2006] or range scanning when fluorescent liquid is employed [Hullin et al. 2008].

Inspired by environment matting, Wu et al. [2018] and Lyu et al. [2020] place a transparent object on a turntable in front of a coded background and capture its multiple views from a static camera position. Wu et al. [2018] derive the correspondence between the incident (camera) and exit rays that reach the background, which additionally requires rotating the background, and finally

consolidate the resulting point clouds into a clean geometric model. Lyu et al. [2020] perform coarse-to-fine mesh optimization, driven by differentiable tracing of refractive two-bounce light paths, so that distorted refractive patterns and object silhouettes match captured photographs. Differentiable rendering is also employed to optimize an IoR field in order to cast desired caustics [Nimier-David et al. 2019] or to design advanced optical systems that account for optical aberrations [Sun et al. 2021; Tseng et al. 2021].

Li et al. [2020] employ a cell phone to capture a small number of views that along with segmented transparent object masks and a known environment map are provided as the input for their method. Sajjan et al. [2020] show that by employing an RGB-D camera the segmentation task is further simplified. Similar goals can be achieved using even a single RGB image and a massively trained encoder-decoder network [Stets et al. 2019]. As pointed out in Lyu et al. [2020] the domain gap can still be expected, as these networks [Stets et al. 2019; Li et al. 2020; Sajjan et al. 2020] are trained on renderings.

2.3 Eikonal rendering

Light propagation in media with varying IoR has been modeled based on formulations derived from the eikonal and transport equations. Mirage rendering [Berger et al. 1990a,b; Musgrave 1990] is concerned with tracing of rays through discrete atmosphere layers, so that the IoR increases with elevation. Stam and Languénou [1996] extend this discrete formulation to media with continuously varying IoR by introducing the eikonal equation to rendering applications. Gutierrez et al. [2005] revisited mirages and other atmospheric effects rendering using such continuous formulation. Ihrke et al. [2007] derived a wavefront tracing technique from the eikonal equation to pre-compute the irradiance distribution in a volume that enables efficient rendering of media with non-homogeneous IoR. Our problem is rather inverse rendering, which has been applied to eikonal forward models in earth sciences [Smith et al. 2021] and interferometric tomography [Sweeney and Vest 1973; Liu and Yang 1989; Tian et al. 2011] to infer 2D or even 3D structure of physical parameters such as velocities, densities or temperatures.

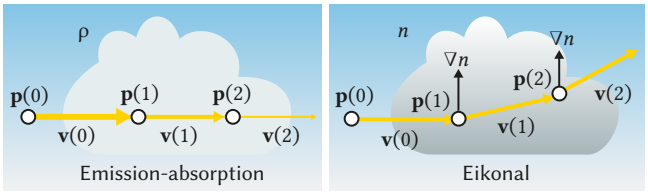

Figure 2: Emission-absorption (left) and eikonal light transport (right). Light is the yellow arrow, its thickness indicates strength. We show three discrete steps. In emission-absorption, direction remains unaltered. In the eikonal formuation, direction changes according to the gradient of the IoR, ∇n. In the eikonal case, strength remains unaffected.

3 LIGHT TRANSPORT ODE ZOO

We will here discuss three approaches to model interaction of light and matter as ordinary differential equations (ODEs): a complete model (Sec. 3.1), an emission-absorption-only model (Sec. 3.2) and an eikonal-only model (Sec. 3.3). The complete one handles refractive and non-refractive scenes, but was only applied to synthetic scenes in the literature. The emission-absorption one can be used for inverse rendering, but excludes refraction. Our eikonal one, in combination with the emission-absorption one, can handle refractive transparency in practical inverse rendering.

3.1 Complete model

When light travels through a scene, it changes its radiance Ldue to absorption and emission as described by the (refractive) radiative transfer equation (RTE) [Preisendorfer 1957]

$$n(s)^2 \frac{\mathrm{d}(L/n^2)}{\mathrm{d}s} = -\sigma(s)L(s) + q(s), \tag{1}$$

where n is the IoR and $s \in [0, \infty[$ is the distance along a (curved) light path, $\sigma$is the extinction coefficient, and $q/\sigma$is the source function (which includes in-scattering) [Chandrasekhar 1950]. The quantity L/n^2 is sometimes referred to as basic radiance. For a spatially varying n, light also changes its position $\mathbf{p}$and direction $\mathbf{v}$due to refraction according to the laws of eikonal light transport [Stam and Languénou 1996; Gutierrez et al. 2005; Ihrke et al. 2007], see Fig. 2. We can describe this using Hamilton's equations for ray tracing [Ihrke et al. 2007]:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \frac{\mathbf{v}(s)}{n(s)} \quad \text{and} \quad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}s} = \nabla n(s), \tag{2}$$

where $\mathbf{v}$ is not unit length but normalized by n. This model has been used in a virtual setting to render advanced visual phenomena including refraction, total internal reflection, and scattering [Gutierrez et al. 2005; Ihrke et al. 2007; Ament et al. 2014; Pediredla et al. 2020]. Unfortunately, this is an ideal model that has not been demonstrated to be tractably used for NVS directly. We will next show the typical simplifications made when ignoring refraction, and introduce a different, also simplified model, that will allow our NVS for refraction.

3.2 Emission-absorption-only model

In NeRF [Mildenhall et al. 2020], radiance remains subject to emission and absorption

$$\frac{\mathrm{d}L}{\mathrm{d}s} = -\sigma(s)L(s) + q(s), \tag{3}$$

but travels along a constant direction $\mathbf{v}$and the change of direction is assumed zero (Fig. 2-left):

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \mathbf{v} \quad \text{and} \quad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}s} = 0. \tag{4}$$

This is classic ray marching along straight rays [Max 1995].

3.3 Eikonal-only model

Complementary and finally, we consider a simplified light transport that does not emit or absorb,

$$\frac{\mathrm{d}L}{\mathrm{d}s} = 0, \tag{5}$$

but changes direction as per eikonal light transport (Fig. 2-right):

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}s} = \frac{\mathbf{v}(s)}{n} \quad \text{and} \quad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}s} = \nabla n(s). \tag{6}$$

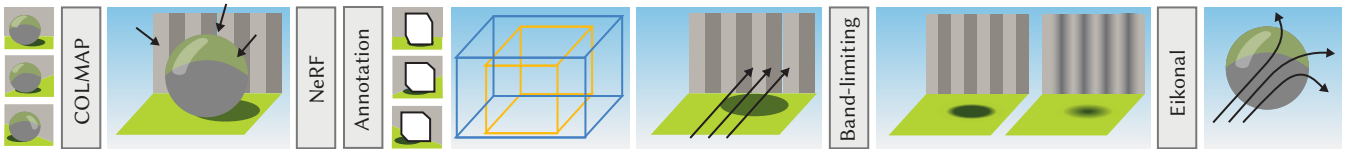

Figure 3: Overview of the pipeline enabling final eikonal training: We start by estimating camera poses [Schönberger and Frahm 2016]. We ask NeRF to explain the scene using emission-absorption and straight rays. In a semi-automated process, we identify a 3D box region not explained and consider this the refractive volume which we exclude from a second NeRF fit. We then grid the view-independent part of this fit to enable the final progressive training using eikonal equations and curved rays.

3.4 Solving

Concisely, all three variants can be formulated as position-motion-radiance state vector and its derivative:

$$\mathbf{z}(s) = (\mathbf{p}, \mathbf{v}, L) \qquad \text{and} \qquad \mathbf{z}'(s) = \frac{\mathrm{d}\mathbf{z}}{\mathrm{d}s}. \tag{7}$$

For all three approaches, coupled ODEs

$$\mathbf{z}(s_1) = \mathbf{z}(s_0) + \int_{s_0}^{s_1} \mathbf{z}'(s)\,\mathrm{d}s = \texttt{odeSolve}(s_0, s_1, \mathbf{z}, \mathbf{z}') \tag{8}$$

need to be solved to compute the final state given the initial state as well as the IoR, emission and absorption fields.

Typically, numerical integration such as Euler solvers are used to solve for the state [Hairer and Wanner 1996]. Working backwards, to compute gradients of the emission or absorption is done by automatic differentiation of forward Euler solvers [Mildenhall et al. 2020; Henzler et al. 2019]. Unfortunately, this requires memory in the order of the number of steps a solver takes. When also accounting for IoR with many small steps, this can quickly become prohibitive. Instead, we use the adjoint [Pontryagin et al. 1962] formulation from Neural ODE [Chen et al. 2018; Stam 2020] that uses constant memory also in backward mode to perform `odeSolve`.

4 OUR APPROACH

Our approach has two main steps: First (Sec. 4.1), reconstructing the opaque scene using a non-eikonal emission-absorption model with straight rays (Sec. 3.2) and, second (Sec. 4.2), modeling the remaining refractive part using an eikonal formulation (Sec. 3.3).

The result of the first step is an input to the second step, i.e., we first train a non-refractive 3D explanation of the world which is input to a second training that 3D-bends rays inside a fixed non-refractive world so that 2D input images can be explained (Fig. 3).

4.1 Non-eikonal step

In this step, we train a NeRF model of emission ($\bar{q}$) and absorption ($\bar{\sigma}$), assuming straight rays. This is used to represent the background and to find the 3D region not explained by the model. We also learn a multi-scale version of this model to be used in the next step.

Registration. In a first step, we compute matrices to transform the camera space of each input image into one reference view using COLMAP [Schönberger and Frahm 2016]. Hence, we also know the 3D ray for every 2D pixel.

Diffuse-opaque init. Given this information an off-the-shelf NeRF is learned that describes emission and absorption as two MLPs that fit continuous functions $\bar{q}(\mathbf{p}, \omega) \in \mathbb{R}^3 \times \Omega \mapsto \mathbb{R}^3$ and $\bar{\sigma}(\mathbf{p}) \in \mathbb{R}^3 \times \Omega \mapsto \mathbb{R}$ mapping position and direction to RGB color or scalar opacity. Let $\theta$and ϕdenote the MLP parameters of the emission and absorption models resulting from this optimization.

Masking. The model above of $\bar{q}$and $\bar{\sigma}$will not be reliable for refractive objects. Hence, we would like to eliminate these parts of 3D space, and explain them by our eikonal approach. The parts that are non-refractive, will be input to this step. We assume the refractive part of the scene can be bounded by a 3D box $\Pi \in \mathbb{R}^{3\times2}$ that exclusively contains refractive objects. This results in a *masked* emission model q, respectively a masked σ:

$$q(\mathbf{p}, \omega) \text{ resp. } \sigma(\mathbf{p}) = \begin{cases} 0 & \text{for} \mathbf{p} \in \Pi \\ \bar{q}(\mathbf{p}, \omega) \text{ resp. } \bar{\sigma}(\mathbf{p}) & \text{otherwise.} \end{cases} \tag{9}$$

We find the box $\Pi$by providing a user with 10 percent of the training images uniformly distributed around the refractive object. The user selects a few points on the horizontal and vertical extent of the refractive object in the image. Once we have collected these 2D points from the images, we use the depth map computed from the NeRF model to find their corresponding 3D locations. We then take 0.02 and 0.98 percentiles of all points along each spatial dimension and multiply them by a constant value of 1.2 to make sure the box encompasses the entire object. The parameters of $\Pi$are given by the minimum and maximum coordinate values of the points.

Progressive grids. Our experiments have shown that solving for the eikonal directly given $\sigma$and $q$is challenging. The problem is that when rays bend a lot it becomes harder to find correspondences between input images and background. Moreover, the bending depends on the spatial gradient of the IoR rather than the IoR directly, which is an operation known to be numerically demanding to optimize over. Addressing this challenge, we will instead learn eikonal transport using different progressively finer versions of the emission and absorption models. This is inspired by progressive spatial encodings [Park et al. 2021], but instead of blurring the periodic spatial functions, we blur the radiance function itself.

It is not obvious how to make a coarser version of $q$or σwhich are MLPs. In particular, our preliminary experiments using slower-varying or fewer spatial encodings did not result in the desired band-limiting. Instead, we recur to relying on regular grids. These are typically struggling to resolve fine details, or to work in 5D, but fortunately, this is not required in our case. Hence, we sample the

masked emission and absorption solutions to a 3D grid as $Q$and P, collapsing Q over the angular domain:

$$Q_i(\mathbf{p}) = \mathbb{E}y[\mathbb{E}_\omega[q(y,\omega)\kappa_i(|\mathbf{p}-y|)]] \quad (10)$$
$$P_i(\mathbf{p}) = \mathbb{E}y[\sigma(y)\kappa_i(|\mathbf{p}-y|)], \quad (11)$$

where κ_i is a Gaussian kernel of increasing frequency bandwidth for increasing levels i. In our experiments, we use a grid size of 128^3 and the values inside the grid are interpolated with a trilinear interpolation scheme.

4.2 Eikonal step

At this step, we have access to a hierarchy of grids describing the emission and absorption in the scene for all locations $\mathbf{p} \notin \Pi$ outside the refractive box. We now find an IoR field defined on $\mathbf{p} \in \Pi$ to explain both the non-refractive 3D grids and the 2D images. A key concept is to *enter and exit* the refractive box in a masked traversal, as well as training with masked rays and progressively.

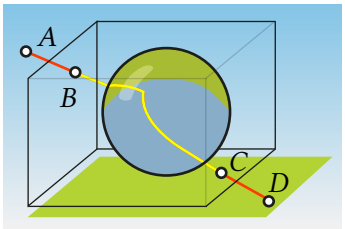

Figure 4: Enter and exit.

Masked traversal. Fig. ?? shows a red ray starting from point A and traversing the world outside Π, which is hit at point B. We use the emission and absorption models $\bar{q}$and $\bar{\sigma}$to trace the straight ray from A to B (Sec. 3.2). Starting at B, eikonal ray-marching curves out the yellow path (Sec. 3.3) according to an IoR model that maps spatial position to IoR: $n(\mathbf{p}) \in \mathbb{R}^3 \mapsto \mathbb{R}$. When this ray leaves Πat C, we let it continue with emission and absorption on a straight path, eventually receiving a contribution at D or other points.

Training of n – that is also an MLP whose parameters are denoted as ψ – proceeds similar to NeRF, but instead of marching geometrically, we solve (and back-propagate through) an ODE in position-motion-radiance space.

Recall that we let $\mathbf{z}$denote a position-motion-radiance vector. We use dot notation to pick an element in the vector so that $\mathbf{z}.\mathbf{p}$denotes the position and $\mathbf{z}.L$the radiance, for example. In the mixed refractive/non-refractive case, the state ODE is

$$\mathbf{z}'_\psi(s) = \begin{cases} \text{Eqs. 5 and 6 s.t. } n_\psi & \text{if } \mathbf{z}_\psi(s).\mathbf{p} \in \Pi \\ \text{Eqs. 3 and 4 s.t. } q_\theta \text{ and } \sigma_\phi & \text{otherwise,} \end{cases} \quad (12)$$

so the state change is non-eikonal outside the box and eikonal inside. It is made to depend on ψ, but not on $\theta$and ϕ, as these are fixed both in the forward and backward pass of this step.

Let $\mathbf{z}_i$ denote the state of a ray through pixel i. We then find

$$\psi^\star = \arg\min_\psi \mathbb{E}_i[|\texttt{odeSolve}(s_0, s_1, \mathbf{z}, \mathbf{z}', \psi).L - \mathbf{z}_i.L|], \quad (13)$$

where $\psi$is an extra argument for `odeSolve` with parameters that condition $\mathbf{z}'$.

As a ray cannot change direction outside Π, the condition in Eq. 12 can be handled by loop splitting in practice: First the ray is traced straight, then traced eikonal, and then it is traced straight once more, eliminating the conditional statement in Eq. 12.

Masked rays. Since our MLP for estimating the IoR is only evaluated inside the bounding box, we start the eikonal training by making sure a batch contains only the rays that are hitting the box.

Progression. We start by finding an IoR field that explains a coarse version of the emission-absorption grid. When the change of error falls below a threshold, we switch one level up to a finer grid. The number of parameters in the MLP to represent the IoR is the same at all levels. We render final images using the full NeRF model instead of a grid.

Interior radiance field. Non-transparent objects might be present in the interior of the transparent object that we located in Π. To explain these, we train another NeRF for the radiance in Π. The IoR field in $\Pi$is available from the eikonal step (Sec. 4.2), and we can now keep it fixed together with the opaque NeRF (Sec. 4.1) and trace paths that bend according to the eikonal when encountering the transparent object in Π. Conclusively, our solution consists of the opaque NeRF, the MLP for the IoR field, and a NeRF for the interior of the transparent object. Together, these have been trained sequentially to explain the input images.

4.3 Implementation details

Our NeRF implementation follows Mildenhall et al. [2020], and our second MLP to represent the IoR field is a 6-layer MLP with 64 hidden dimensions with a skip connection that concatenates the input to the third layer's activation. Similar to NeRF, we also apply positional encoding with five frequencies to the input. For stable training, as suggested by Chen et al. [2018], we use softplus activation with $\beta = 5$ for all layers instead of a non-smooth function like ReLU and all layers are initialized with the Xavier uniform. In the non-eikonal step (training NeRF), we use the same training setting as described by Mildenhall et al. [2020], and let the optimization run for 150k iterations. This takes around 12 hours to converge on a single NVIDIA 1080Ti with 12 GB RAM. For the eikonal step, we use a batch size of 1024 rays and traverse the space with 128 ODE steps and the training takes around 5 hours for 5k iterations. We use the Neural ODE PyTorch implementation [Chen et al. 2018; Stam 2020] to backpropagate through the ODE with the adjoint method. In the progression part, we smooth the grid with a Gaussian kernel with a normalized frequency bandwidth of 0.08 cycles per sample and double this for every 1k iterations. For the last step, we use a single MLP similar to the NeRF fine network [Mildenhall et al. 2020], but with 128 hidden dimensions to represent the interior radiance field. As we do not adopt any hierarchical volume sampling, we consider 512 steps along the ray to properly sample both interior and exterior radiance fields, and it takes around 12 hours to optimize over 10k iterations. With our complete model, it takes around 85 seconds to render a frame of 672×504 resolution.

5 RESULTS

Our aim is NVS with plausible coherence in scenes with transparent objects. We look at a range of scenes where we apply different methods and – as no metric exists to quantify our main aim – we refer to standard peak signal-to-noise ratio (PSNR), structural

Table 1: Quantitative comparison of different methods (rows) using different scenes and metrics (columns). The numbers in the User column say how often in our user study the method was considered closer to the reference than ours. As these numbers are significantly ($p < 0.01$) smaller than the chance level 50%, our method was for all scenes considered closest to the reference in the majority of the comparisons shown to the users.

	Ball				Glass				Pen				WineGlass			
	PSNR	SSIM	LPIPS	User	PSNR	SSIM	LPIPS	User	PSNR	SSIM	LPIPS	User	PSNR	SSIM	LPIPS	User
NeRF	**27.384**	0.945	0.042	0.27%	**27.146**	**0.924**	0.066	3.83%	27.749	0.933	0.059	9.58%	**29.011**	**0.947**	0.045	24.93%
Trivial	24.373	0.933	0.034	9.31%	25.930	0.914	0.059	7.39%	23.070	0.912	0.060	37.26%	26.739	0.935	0.052	1.33%
Direct	27.247	0.942	0.054		26.031	0.914	0.081		26.624	0.928	0.073		27.379	0.938	0.060	
Ours	26.720	**0.951**	**0.023**		26.525	0.922	**0.050**		**27.803**	**0.935**	**0.047**		27.789	0.940	**0.042**	

similarity (SSIM), and learned perceptual image patch similarity (LPIPS) metrics, and we perform a user study too.

Scenes. We selected four real scenes including refractive objects with unknown geometry: Ball, Glass, Pen and WineGlass. We used an Iphone 8 camera to capture 96, 97, 105, and 102 views, respectively for each scene, and we hold out 1/10 of all views for the test set. All images are of resolution 672×504 pixels.

Methods. We compared Ours with NeRF and two other methods named Direct and Trivial. In Direct, we jointly optimize for the emission-absorbtion and IoR models. In this setup, similar to progressive grid, we provide a coarse-to-fine optimization scheme by applying progressive positional encoding [Park et al. 2021] for the emission-absorbtion MLP. In Trivial, we try to reconstruct the IoR field using the density field of refractive objects recovered by NeRF. We do this by first executing the NeRF model for a discrete set of samples along the rays coming from the input camera poses and crossing the bounding box Π, and we set the density to zero for the samples outside the box. Then, for each ray, we estimate both the front and back surface position of the refractive object by forward and backward ray marching until an opacity threshold is reached (similar to how the depth maps are computed in NeRF). For the samples that fall between the intersections, we assign a constant IoR value (1.5 for Glass, 1.33 WineGlass), and we choose 1.0 for the regions outside. We then try to fit an MLP to map each 3D point inside the box Πto its calculated IoR.

Qualitative comparisons. Fig. 5 facilitates a visual comparison of Ours with NeRF and Trivial. Please refer to the supplemental material for our visual comparison with the Direct method. The insets show novel view reconstructions of different view points for all methods. Please refer to the supplemental video for an animated version of these results. NeRF tends to "fake" refraction by considering a diffuse content on the surface of the transparent object and assigning view-dependent color for each point on the surface. Under the condition of extreme view changes, as can be seen in all scenes, NeRF fails to properly reproduce the color and it tends to average all observations leading to a blurry result. NeRF also seems to struggle with reconstruction of an occluder inside the transparent objects although multi-view consistency holds for the object inside. In the Pen scene, NeRF failed to assign a transparent content on the surface of the glass in order to properly reconstruct the pen inside. Trivial assumes a constant IoR field inside the entire refractive object and in case of spatially varying IoR, the refraction tends to be wrong for some regions (e.g., towards the top and the bottom of the glass in the Glass and the Pen scenes). Trivial performs better on the Ball scene as the crystal ball has a constant IoR inside. However, due to the mere fact that the NeRF density field for the refractive object is not always valid, the estimated IoR of Trivial might not be very accurate and the refracted background becomes misplaced in some regions. In contrast, Ours reproduces sharper details and aligns better with the reference. Moreover, in order to assess the temporal consistency of each method, in the right block, we also show the corresponding pseudo-epipolar image that is created by stacking a selected scanline for 30 subsequent video frames using a continuous camera trajectory. A good optical flow continuity can be observed between the stacked scanlines for all methods, but clearly the flow fidelity with respect to the reference is best for Ours. NeRF and Trivial feature significant blur that is visible also in the insets in the middle column.

User study. Unfortunately, no method exists to quantify the main aim of this work, plausible refractive and reflective flow. To quantify the coherency, we performed a small user study. A reference photograph and two images produced by Ours and either NeRF or Trivial (selected randomly) were shown to 73 participants, 10 image triplets for each scene. The participants then had to indicate which one is visually closer to the reference in a two-alternative forced choice (2AFC) experiment. All three images were presented simultaneously without any time limit; the position of the reference was fixed, while it was randomized for the other two. We selected five different views for each of the four scenes and aggregated the participant selection over those views. For each scene, we report how often a competitor was selected, hence less is better for us, while the chance level is 50%. All outcomes are significant at the $p < 0.01$-level for a binomial test at $N = 73$.

Quantitative comparisons. Tab. 1 presents quantitative results of our user study (in the "User" columns) and for the different metrics averaged over our test set. We see NeRF has consistently the highest PSNR, which is a metric relatively insensitive to blur or structure preservation. When it comes to SSIM, already a metric more aware of the structures we want to preserve, it comes to a draw. At the most advanced metric, LPIPS – which is based on human image artifact perception and better tolerates small spatial misalignments with respect to the reference – Ours always wins. Direct and Trivial are sometimes better than other methods but never win. The participants of our user study almost consistently indicate that Ours leads to less perceived differences with respect to the reference views. As for a relatively high score of Trivial for

Figure 5: The left block shows the cross section of recovered IoR by our method for a scanline between the white dots shown in the reconstructed test view in the second block. The third block shows insets taken from novel views produced by three different methods (rows) for different view points (columns). The right block shows a pseudo-epipolar view using a continuous camera trajectory, again for all methods.

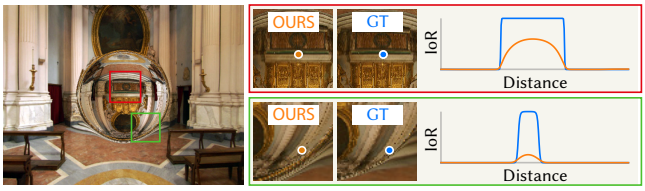

Figure 6: IoR cross section of our method (orange) and ground truth (blue) for a scanline along the pixel marked with dot in each inset.

Pen, we hypothesize that the background sharpness and its color saturation could have appeal to some participants, who neglected strong background distortions, and the pen's absence, visible in Fig. 5. The relatively high score of NeRF for WineGlass can be attributed to views where the background contained less high-frequency details, so that blur became perceivable.

Reference comparison. While our method makes use of an IoR, it is not forced to use actual physical values. For a scene with a known IoR (Fig. 6), we see that it can reconstruct the image faithfully, while a cross-section shows the IoR is indeed quite different from the

reference IoR. We would hence like to iterate that our method is suitable for NVS, not for the reconstruction of 3D structure.

6 CONCLUSION

Given a set of 2D images containing refractive materials, we explored the problem of optimizing for the field of 3D-spatially varying IoR with the purpose of NVS. Existing solutions that learn 3D fields for NVS are based on opaque or transparent light transport along straight paths. As opposed to this, we model the bending of light according to the eikonal equation from geometric optics. This enables us to do perceptually better NVS in 3D scenes with complex objects exhibiting strong refractive effects.

Our work is subject to several assumptions. The eikonal equation deals with refraction and total internal reflection but not separation into partial reflection and refraction. Partial reflection and refraction in continuously varying media is difficult even in forward simulation, and left for future work. We also first learn the diffuse world, followed by the transparent objects in a second pass, where we rely on a user marking the bounding box of the specular object to aid the task. Ideally, this would be done jointly and in a fully automated way. As a consequence, we have to assume that we sufficiently observe the diffuse world directly, making us unable to reconstruct parts exclusively revealed in the refraction.

Despite our simplifying assumptions, we have by means of eikonal light transport for the first time included refraction and total internal reflection in a model that learns 3D fields from images of transparent objects to accomplish synthesis of novel views. We compared our results with other methods and conducted a user study which strongly indicates that we achieve results that are perceptually closer to reference images.

REFERENCES

A. Pediredla, Y. K. Chalmiani, M. G. Scopelliti, M. Chamanzar, S. Narasimhan, and I. Gkioulekas. 2020. Path tracing estimators for refractive radiative transfer. *ACM Trans. Graph.* 39, 6, Article 241 (2020), 15 pages.

Marco Ament, Christoph Bergmann, and Daniel Weiskopf. 2014. Refractive radiative transfer equation. *ACM Trans. Graph.* 33, 2, Article 17 (2014), 22 pages.

B. Attal, J.-B. Huang, M. Zollhoefer, J. Kopf, and C. Kim. 2021. Learning Neural Light Fields with Ray-Space Embedding Networks. arXiv:cs.CV/2112.01523

Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. 2020. X-Fields: Implicit Neural View-, Light- and Time-Image Interpolation. *ACM Trans. on Graph.* 39, 6, Article 257 (2020), 15 pages.

Marc Berger, Nancy Levit, and Terry Trout. 1990a. Rendering mirages and other atmospheric phenomena. In *Eurographics*. Eurographics Association, 459–468.

M. Berger, T. Trout, and N. Levit. 1990b. Ray Tracing Mirages. *IEEE Computer Graphics and Applications* 10, 3 (1990), 36–41.

Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P. A. Lensch. 2021. NeRD: Neural Reflectance Decomposition from Image Collections. In *ICCV*. IEEE, 12684–12694.

S. Chandrasekhar. 1950. *Radiative Transfer.* Oxford University Press, Oxford.

Guanying Chen, Kai Han, and Kwan-Yee K. Wong. 2019. Learning transparent object matting. *Int J Computer Vision* 127, 10 (2019), 1527–1544.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. 2018. Neural ordinary differential equations. In *NeurIPS*. 6572–6583.

Yung-Yu Chuang, Douglas E. Zongker, Joel Hindorff, Brian Curless, David H. Salesin, and Richard Szeliski. 2000. Environment matting extensions: Towards higher accuracy and real-time capture. In *SIGGRAPH*. ACM, 121–130.

Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. 2021. NeRFReN: Neural Radiance Fields with Reflections. arXiv:cs.CV/2111.15234

Diego Gutierrez, Adolfo Munoz, Oscar Anson, and Francisco J. Seron. 2005. Non-linear Volume Photon Mapping. In *EGSR*. Eurographics Association, 291–300.

Ernst Hairer and Gerhard Wanner. 1996. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems* (second revised ed.). Springer.

Philipp Henzler, Niloy J. Mitra, and Tobias Ritschel. 2019. Escaping Plato's cave: 3D shape from adversarial rendering. In *ICCV*. IEEE, 9984–9993.

M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch. 2008. Fluorescent Immersion Range Scanning. *ACM Trans. Graph.* 27, 3, Article 87 (2008), 10 pages.

Jeffrey Ichnowski, Yahav Avigal, Justin Kerr, and Ken Goldberg. 2021. Dex-NeRF: Using a Neural Radiance Field to Grasp Transparent Objects. arXiv:cs.RO/2110.14217

I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. 2010. Transparent and specular object reconstruction. *Comp. Graph. Forum* 29, 8 (2010), 2400–2426.

Ivo Ihrke, Gernot Ziegler, Art Tevs, Christian Theobalt, Marcus Magnor, and Hans-Peter Seidel. 2007. Eikonal rendering: Efficient light transport in refractive objects. *ACM Trans. Graph.* 26, 3, Article 59 (2007), 9 pages.

Erum Arif Khan, Erik Reinhard, Roland W. Fleming, and Heinrich H. Bülthoff. 2006. Image-based material editing. *ACM Trans. Graph.* 25, 3 (2006), 654–663.

J. Kim, I. Reshetouski, and A. Ghosh. 2017. Acquiring Axially-Symmetric Transparent Objects Using Single-View Transmission Imaging. In *CVPR*. IEEE, 3559–3567.

M. Kolos, A. Sevastopolsky, and V. Lempitsky. 2020. TRANSPR: Transparency Ray-Accumulating Neural 3D Scene Point Renderer. In *3DV*. IEEE, 1167–1175.

Kiriakos Kutulakos and Eron Steger. 2008. A Theory of Refractive and Specular 3D Shape by Light-Path Triangulation. *Int J Computer Vision* 76, 1 (2008), 13—-29.

Zhengqin Li, Yu-Ying Yeh, and Manmohan Chandraker. 2020. Through the looking glass: neural 3D reconstruction of transparent shapes. In *CVPR*. IEEE, 1262–1271.

K. Liu and J. Y. Yang. 1989. Reconstruction of 3-D Refractive Index Fields from Multi-Frame Interferometric Data. In *New Methods in Microscopy and Low Light Imaging (Proc. SPIE)*, Vol. 1161. SPIE, 42–46.

Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural volumes: learning dynamic renderable volumes from images. *ACM Trans. Graph.* 38, 4, Article 65 (2019), 14 pages.

Jiahui Lyu, Bojian Wu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2020. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Trans. Graph.* 39, 6, Article 195 (2020), 13 pages.

Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard Mcmillan. 2002. Acquisition and Rendering of Transparent and Refractive Objects. In *EGWR*. Eurographics Association, 267–278.

Nelson Max. 1995. Optical models for direct volume rendering. *IEEE Trans Vis Comput Graph* 1, 2 (1995), 99–108.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*. Springer, 405–421.

F. Kenton Musgrave. 1990. A Note on Ray Tracing Mirages. *IEEE Computer Graphics and Applications* 10, 6 (1990), 10–12.

M. Nimier-David, D. Vicini, T. Zeltner, and W. Jakob. 2019. Mitsuba 2: A Retargetable Forward and Inverse Renderer. *ACM Trans. Graph.* 38, 6, Article 203 (2019), 17 pages.

Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. 2021. Nerfies: Deformable Neural Radiance Fields. In *ICCV*. IEEE, 5865–5874.

Pieter Peers and Philip Dutré. 2003. Wavelet Environment Matting. In *EGSR*. Eurographics Association, 157–166.

L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. 1962. *The Mathematical Theory of Optimal Processes.* Interscience Publishers, New York.

Rudolph W. Preisendorfer. 1957. A mathematical foundation for radiative transfer theory. *Journal of Mathematics and Mechanics* 6, 6 (1957), 685–730.

Shreeyak Sajjan, Matthew Moore, Mike Pan, Ganesh Nagaraja, Johnny Lee, Andy Zeng, and Shuran Song. 2020. ClearGrasp: 3D shape estimation of transparent objects for manipulation. In *ICRA*. IEEE, 3634–3642.

Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *CVPR*. IEEE, 4104–4113.

Sudipta N. Sinha, Johannes Kopf, Michael Goesele, Daniel Scharstein, and Richard Szeliski. 2012. Image-Based Rendering for Scenes with Reflections. *ACM Trans. Graph.* 31, 4, Article 100 (2012), 10 pages.

Jonathan D. Smith, Kamyar Azizzadenesheli, and Zachary E. Ross. 2021. EikoNet: Solving the Eikonal Equation With Deep Neural Networks. *IEEE Trans Geoscience and Remote Sensing* 59, 12 (2021), 10685–10696.

Jos Stam. 2020. Computing Light Transport Gradients using the Adjoint Method. arXiv:cs.GR/2006.15059

Jos Stam and Eric Languénou. 1996. Ray tracing in non-constant media. In *Rendering Techniques '96 (EGWR)*. Springer, 225–234.

J. D. Stets, A. Dal Corso, J. B. Nielsen, R. A. Lyngby, S. H. N. Jensen, J. Wilm, M. B. Doest, C. Gundlach, E. R. Eiriksson, K. Conradsen, A. B. Dahl, J. A. Bærentzen, J. R. Frisvad, and H. Aanæs. 2017. Scene reassembly after multimodal digitization and pipeline evaluation using photorealistic rendering. *Applied Optics* 56, 27 (2017), 7679–7690.

J. D. Stets, Z. Li, J. R. Frisvad, and M. Chandraker. 2019. Single-shot analysis of refractive shape using convolutional neural networks. In *WACV*. IEEE, 995–1003.

Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. 2021. End-to-End Complex Lens Design with Differentiate Ray Tracing. *ACM Trans. Graph.* 40, 4, Article 71 (2021), 13 pages.

D. W. Sweeney and C. M. Vest. 1973. Reconstruction of Three-Dimensional Refractive Index Fields from Multidirectional Interferometric Data. *Applied Optics* 12, 11 (1973), 2649–2664.

A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B Goldman, and M. Zollhöfer. 2020. State of the art on neural rendering. *Comp. Graph. Forum* 39, 2 (2020), 701–727.

A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, Y. Wang, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Niessner, J. T. Barron, G. Wetzstein, M. Zollhoefer, and V. Golyanik. 2021. Advances in Neural Rendering. arXiv:cs.GR/2111.05849

Chao Tian, Yongying Yang, Yongmo Zhuo, Tao Wei, and Tong Ling. 2011. Tomographic reconstruction of three-dimensional refractive index fields by use of a regularized phase-tracking technique and a polynomial approximation method. *Applied Optics* 50 (2011), 6495–6504.

Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. 2006. Tomographic Reconstruction of Transparent Objects. In *EGSR*. Eurographics Association, 51–60.

E. Tseng, A. Mosleh, F. Mannan, K. St-Arnaud, A. Sharma, Y. Peng, A. Braun, D. Nowrouzezahrai, J.-F. Lalonde, and F. Heide. 2021. Differentiable Compound Optics and Processing Pipeline Optimization for End-to-end Camera Design. *ACM Trans. on Graph.* 40, 2, Article 18 (2021), 19 pages.

Ziyu Wang, Liao Wang, Fuqiang Zhao, Minye Wu, Lan Xu, and Jingyi Yu. 2021. MirrorNeRF: One-Shot Neural Portrait Radiance Field from Multi-Mirror Catadioptric Imaging. arXiv:cs.CV/2104.02607

Gordon Wetzstein, David Roodnick, Wolfgang Heidrich, and Ramesh Raskar. 2011. Refractive shape from light field distortion. In *ICCV*. IEEE, 1180–1186.

Yonatan Wexler, Andrew W. Fitzgibbon, and Andrew Zisserman. 2002. Image-Based Environment Matting. In *EGSR*. Eurographics Association, 279–290.

Suttisak Wizadwongsa, Pakkapon Phongthawee, Jiraphon Yenphraphai, and Supasorn Suwajanakorn. 2021. NeX: Real-time View Synthesis with Neural Basis Expansion. In *CVPR*. IEEE, 8534–8543.

B. Wu, Y. Zhou, Y. Qian, M. Cong, and H. Huang. 2018. Full 3D reconstruction of transparent objects. *ACM Trans. Graph.* 37, 4, Article 103 (2018), 11 pages.

Jiamin Xu, Xiuchao Wu, Zihan Zhu, Qixing Huang, Yin Yang, Hujun Bao, and Weiwei Xu. 2021. Scalable Image-Based Indoor Scene Rendering with Reflections. *ACM Trans. Graph.* 40, 4, Article 60 (2021), 14 pages.

Sai-Kit Yeung, Chi-Keung Tang, Michael S. Brown, and Sing Bing Kang. 2011. Matting and Compositing of Transparent and Refractive Objects. *ACM Trans. Graph.* 30, 1, Article 2 (2011), 13 pages.

A. Yu, S. Fridovich-Keil, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa. 2021a. Plenoxels: Radiance Fields without Neural Networks. arXiv:cs.CV/2112.05131

A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa. 2021b. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *ICCV*. IEEE, 5752–5761.

Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021a. PhySG: Inverse Rendering with Spherical Gaussians for Physics-based Material Editing and Relighting. In *CVPR*. IEEE, 5453–5462.

X. Zhang, P. P. Srinivasan, B. Deng, P. Debevec, W. T. Freeman, and J. T. Barron. 2021b. NeRFactor: Neural Factorization of Shape and Reflectance Under an Unknown Illumination. *ACM Trans. Graph.* 40, 6, Article 237 (2021), 18 pages.

D. E. Zongker, D. M. Werner, B. Curless, and D. H. Salesin. 1999. Environment Matting and Compositing. In *SIGGRAPH*. ACM/Addison-Wesley, 205–214.