

# Sampling Neural Radiance Fields for Refractive Objects

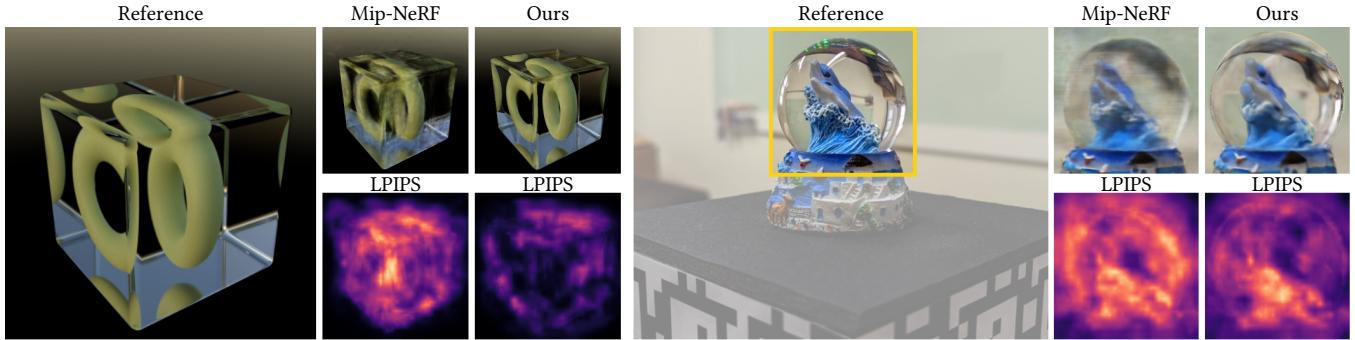
Jen-I Pan  
alexkeroro86@gmail.com  
National Tsing Hua University  
Hsinchu, Taiwan

Jheng-Wei Su  
jhengweisu@gapp.nthu.edu.tw  
National Tsing Hua University  
Hsinchu, Taiwan

Kai-Wen Hsiao  
kevin30112@gmail.com  
National Tsing Hua University  
Hsinchu, Taiwan

Ting-Yu Yen  
tingyus995@gmail.com  
National Tsing Hua University  
Hsinchu, Taiwan

Hung-Kuo Chu  
pigjohn@gmail.com  
National Tsing Hua University  
Hsinchu, Taiwan



**Figure 1:** Our framework takes multi-view images as inputs and renders novel views of both synthetic (left) and real (right) scenes containing refractive objects. With the benefit of considering refraction paths, our results on the surfaces (cube and sphere) and the interior objects (torus and dolphin) are more accurately rendered, as shown in the error maps using LPIPS index (brighter regions indicate higher errors).

## ABSTRACT

Recently, differentiable volume rendering in neural radiance fields (NeRF) has gained a lot of popularity, and its variants have attained many impressive results. However, existing methods usually assume the scene is a homogeneous volume so that a ray is cast along the straight path. In this work, the scene is instead a heterogeneous volume with a piecewise-constant refractive index, where the path will be curved if it intersects the different refractive indices. For novel view synthesis of refractive objects, our NeRF-based framework aims to optimize the radiance fields of bounded volume and boundary from multi-view posed images with refractive object silhouettes. To tackle this challenging problem, the refractive index of a scene is reconstructed from silhouettes. Given the refractive index, we extend the stratified and hierarchical sampling techniques in NeRF to allow drawing samples along a curved path tracked by the Eikonal equation. The results indicate that our framework outperforms the state-of-the-art method both quantitatively and qualitatively, demonstrating better performance on the perceptual similarity metric and an apparent improvement in the rendering quality on several synthetic and real scenes.

## CCS CONCEPTS

- Computing methodologies → Computer graphics; Machine learning.

## KEYWORDS

neural radiance fields, eikonal rendering

### ACM Reference Format:

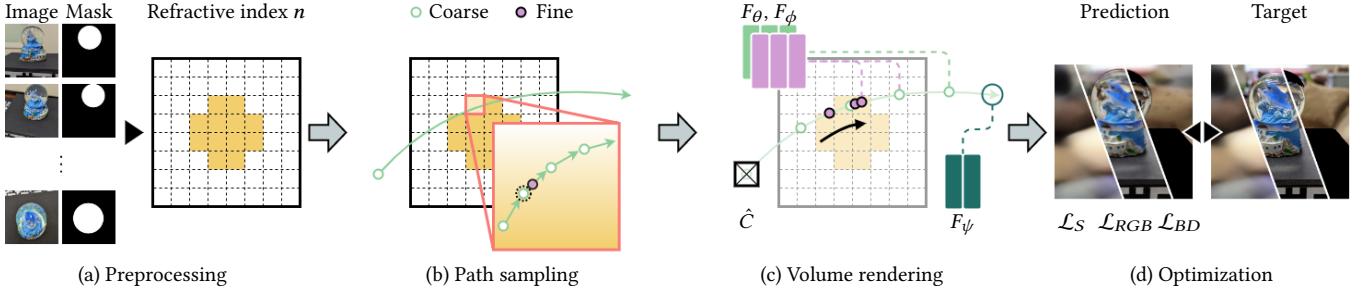
Jen-I Pan, Jheng-Wei Su, Kai-Wen Hsiao, Ting-Yu Yen, and Hung-Kuo Chu. 2022. Sampling Neural Radiance Fields for Refractive Objects. In . ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3550340.3564234>

## 1 INTRODUCTION AND RELATED WORK

Refraction is ubiquitous in everyday life. For example, distorted objects seen through the water, and a magnifying glass decreasing the field of view. Thus, accurate rendering of refraction is crucial to improve realism. Nevertheless, reconstructing the scene with a refractive object from multi-view images is an ill-posed problem due to the ambiguity among geometry, material and refractive index.

For the past two years, neural radiance field, or NeRF [Mildenhall et al. 2020], and its variants that treat a scene as a homogeneous volume have been widely explored. NeRF uses two multi-layer perceptrons (MLPs), one coarse  $F_\theta$  and one fine  $F_\phi$ , to represent a volumetric scene. The MLP takes the encoded position  $\mathbf{x}$  of a sample and view direction  $\mathbf{d}$  by positional encoding as inputs; it outputs the density  $\sigma$  and radiance  $\mathbf{c}$ . The pixel value  $\hat{C}$  is estimated by the differentiable volume rendering equation in Eqn. 1 with all

This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in *SA '22 Technical Communications*, December 6–9, 2022, Daegu, Republic of Korea © 2022 Association for Computing Machinery.  
<https://doi.org/10.1145/3550340.3564234>



**Figure 2: Framework overview.** Given the multi-view posed images and refractive object silhouettes, we first reconstruct the refractive index  $n$  of a scene from silhouettes and store it in a voxel grid (a). Next, we track the ray of a pixel  $C$  and draw the samples along the traversed path (b). Then, we query the density and radiance of each sample from the networks, namely  $F_\theta$  and  $F_\phi$ , and combine them with the boundary radiance evaluated by the boundary network  $F_\psi$  to estimate the resulting color (c). Finally, we optimize the three networks with respect to the re-rendering error ( $\mathcal{L}_{RGB}$ ) and regularizers ( $\mathcal{L}_S$  and  $\mathcal{L}_{BD}$ ) (d).

the samples  $\mathbf{r}$  along a ray cast from the camera origin  $\mathbf{o}$  to the pixel.

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) \mathbf{c}_i, \quad (1)$$

where  $T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$  is the accumulated transmittance, and  $\delta$  is the distance between adjacent samples. Although NeRF can replace the background with white via the mask of an opaque object, the background seen through a refractive object cannot be easily removed. Thus, the original equation in NeRF (Eqn. 1) is insufficient for our problem due to the lack of a boundary term. The samples used to estimate the pixel value consist of two sets, one coarse  $\mathbf{r}_c$  and one fine  $\mathbf{r}_f$ , for the coarse and fine networks. Specifically, the fine network takes the union of both sets after sorting. The coarse samples  $\mathbf{r}_c = \{(\mathbf{x}_i, \mathbf{d})\}_{i=0}^{N_c}$  are drawn uniformly from the evenly-spaced bins between the near  $t_N$  and far  $t_F$  planes with a stratified sampling of  $t_i$  in Eqn. 2 for a continuous representation; the fine samples  $\mathbf{r}_f = \{(\mathbf{x}_i, \mathbf{d})\}_{i=1}^{N_c}$  are then allocated to visible regions that most likely contribute to the pixel value based on the coarse network with a hierarchical sampling of  $t_i$  in Eqn. 3, respectively by the ray equation  $\mathbf{x}_i = \mathbf{o} + t_i \mathbf{d}$ .

$$\mathbf{t}_c = \{t_i \sim \mathcal{U}[(i-1)/N_c, i/N_c] \cdot (t_F - t_N) + t_N\}_{i=1}^{N_c} \quad (2)$$

$$\mathbf{t}_f = \{t_i \sim \text{InverseTransformSampling}(\hat{w}_i)\}_{i=1}^{N_f} \quad (3)$$

where  $\hat{w}_i = w_i / \sum_{j=1}^{N_c} w_j$ , and  $w_i = T_i (1 - \exp(-\sigma_i \delta_i))$ . However, these two sampling techniques in NeRF cannot be used with only the distance  $t$  and view direction  $\mathbf{d}$  if a path is curved due to the refraction. To this end, we combine light transport simulation based on the *Eikonal equation* with NeRF for the refraction, and we extend the original sampling techniques in NeRF to curved paths.

In terms of the rendering quality and refraction, mip-NeRF [Barren et al. 2021] proposes an integrated positional encoding and achieves the best quality for both single- and multi-scale contents. Ref-NeRF [Verbin et al. 2021] uses a concept of environment mapping to enable a sharp view-dependent effect. CompLum [Zhu et al. 2021] brings a surface light field to avoid the costly evaluation of light paths inside the complex refractive geometry. Still, none of them takes refraction paths into account. Furthermore, Matusik et

al. [2002] introduce an image-based rendering method with the 3D scanner for refractive objects.

Moreover, instead of a bounded scene, a boundary can be treated as different representations [Hao et al. 2021; Zhang et al. 2020]. To handle the leftover transmittance or density, GANcraft [Hao et al. 2021] proposes a regularizer. Overall, we represent the boundary as a skybox and bring a proper regularizer to resolve the ambiguity of color blending between refractive object and boundary. Supplementary materials, codes, and datasets are released for academic usage at <https://github.com/alexkeroro86/SampleNeRFRO>.

In summary, we make the following contributions:

- A NeRF-based framework for generating high-quality novel view rendering of refractive objects presenting the refraction and total reflection effects.
- A tailor-made hierarchical path sampling technique for both straight and curved paths.

*Concurrent work.* Eikonal Fields [Bemana et al. 2022] aims at the same problem as ours. Compared to the concurrent work, both methods follow the ray equation of geometric optics derived from the Eikonal equation for a volumetric scene by Eikonal Rendering [Ihrke et al. 2007]. We use the piecewise-linear approximation in [Sun et al. 2008] (Eqn. 4) to construct refraction paths:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \frac{\Delta s}{n} v_i, \quad v_{i+1} = v_i + \Delta s \nabla \mathbf{n}, \quad (4)$$

where  $\Delta s$  is step size,  $v$  is defined by  $n \frac{d\mathbf{x}}{ds}$ ,  $n$  is refractive index and  $\nabla \mathbf{n}$  is gradient index. For the refractive index, Eikonal Fields tackles the challenging task of reconstructing the refractive index of a scene by the dedicated multi-stage training strategy. We assume the object's material is known, hence its corresponding refractive index (e.g., 1.52 for glass and 1.33 for water). In terms of scene complexity, instead of using a bounding box annotation, we provide more complex interior objects inside different refractive objects for both synthetic and real scenes to inspire the follow-up works. For the sampling, Eikonal Fields only draws samples uniformly between the bounds, but we further propose a hierarchical path sampling technique. Besides, NeRF [Wang et al. 2022] aims to recover the depth and normal of a flat fluid surface for one-time refraction of the last sample using the Snell's law.

*Assumption.* Since we consider the refractive index of a refractive object, samples behind the interior object should be occluded eventually. Therefore, a ray does not change its direction when crossing the interior object. In addition, we ignore the outer surface of refractive objects, such as the glass in a glass of water.

## 2 METHOD

Fig. 2 illustrates an overview of our NeRF-based framework. To be compatible with the Eikonal equation in Eqn. 4, we first reconstruct the proxy geometry of the refractive object by shape from silhouette and remove the noisy components manually. Then, we choose a voxel grid with tri-linear interpolation to represent the refractive index of a scene. For each vertex, its refractive index is calculated by  $A/(A+B) \cdot 1.0 + B/(A+B) \cdot n$ , where  $A$  and  $B$  is the number of samples within a voxel that are outside and inside the proxy geometry, respectively. In addition, to eliminate the stair-step artifacts in rendering, we smooth the voxel grid before compute the gradient index as Sun et al. [2008].

To construct a path according to the refractive index and gradient index, we leverage the Eikonal equation in Eqn. 4 to bend a piecewise-linear path at each step  $i$ . We choose the step size  $\Delta s = (t_f - t_N)/(N_c \times N_e)$  to track the path in  $N_c * N_e$  steps. Here all the steps are denoted as the *Eikonal samples*, and we also collect the distance of each step. However, if we use all the Eikonal samples to estimate the pixel value, it is costly to evaluate by using the coarse network. Therefore, we randomly draw one sample in every  $N_e$  samples to reduce the times of network evaluation, and these drawn samples are called the *coarse samples*  $\mathbf{r}_c = \{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=0}^{N_c}$ , where  $\mathbf{d}_i = v_i / \|v_i\|$ . For the *fine samples*  $\mathbf{r}_f = \{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=0}^{N_f}$ , we should not only allocate the fine samples to visible regions as NeRF but also make sure they are still along the piecewise-linear path by extending the hierarchical sampling. With Eqn. 3, we transform the set of fine distances  $t_f$  to the fine samples  $\mathbf{r}_f$  by assigning the direction of the nearest former Eikonal sample  $\mathbf{d}_{\lfloor t \rfloor}$  to  $\mathbf{d}_i$  according to the distance  $t$ . Then, we re-calculate the position  $\mathbf{x}_i$  based on the position  $\mathbf{x}_{\lfloor t \rfloor}$  and distance  $\lfloor t \rfloor$  of the nearest former Eikonal sample along the direction  $\mathbf{d}_i$  by  $\mathbf{x}_i = \mathbf{x}_{\lfloor t \rfloor} + \mathbf{d}_i(t - \lfloor t \rfloor)$ . We illustrate our sampling techniques and compare to NeRF in Fig. 3.

After the network evaluation of samples  $\mathbf{r}$ , we gather the corresponding density  $\sigma$  and radiance  $\mathbf{c}$  by the following volume rendering equation with boundary term:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - \exp(-\sigma_i \delta_i)) \mathbf{c}_i + T_{N+1} C'(\mathbf{d}_N), \quad (5)$$

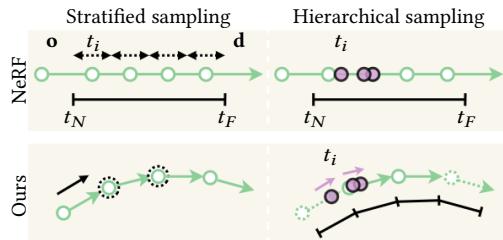


Figure 3: Comparison between NeRF's and our sampling techniques. A dotted double arrow is a bin in Eqn. 2.

where  $C'$  is a skybox represented as a small MLP  $F_\psi : \mathbf{d} \rightarrow \mathbf{c}$  whose architecture is based on the normal field in NeRFactor [Zhang et al. 2021], and  $\mathbf{d}_N$  is the leaving direction from the bounded volume.

Finally, we optimize the three MLPs, namely  $F_\theta$ ,  $F_\phi$  and  $F_\psi$ , with respect to the following objective function:

$$\mathcal{L} = \lambda_{RGB} \mathcal{L}_{RGB} + \lambda_{BD} \mathcal{L}_{BD} + \lambda_S \mathcal{L}_S, \quad (6)$$

where  $\lambda_{RGB}$ ,  $\lambda_{BD}$  and  $\lambda_S$  are the weighted hyper-parameters. We illustrate these terms in Fig. 2(d).

*Re-rendering error.* We use a L2 loss to compare the coarse  $\hat{C}_c$  and fine  $\hat{C}_f$  pixel values with the ground truth  $C(\mathbf{r})$  as NeRF:

$$\mathcal{L}_{RGB} = \|C(\mathbf{r}) - \hat{C}_c(\mathbf{r}_c)\|_2^2 + \|C(\mathbf{r}) - \hat{C}_f(\text{sort}(\mathbf{r}_c \cup \mathbf{r}_f))\|_2^2. \quad (7)$$

*Boundary regularizer.* It is calculated based on the re-rendering error but only updates the density  $\sigma$  evaluated by the fine network to preserve the visual quality and eliminate the blurry artifacts on the refractive surface:

$$\mathcal{L}_{BD} = \mathbb{1}(T_{f, N_c + N_f + 1}) \cdot \|C(\mathbf{r}) - T_{f, N_c + N_f + 1} C'(\mathbf{d}_{f, N_c + N_f})\|_1, \quad (8)$$

where  $\mathbb{1}(\cdot)$  is an indicator function to ignore the error of occluded region by the threshold 0.5 on the last accumulated transmittance.

*Smoothness regularizer.* We add an L2 gradient penalty to boundry as NeRFactor [Zhang et al. 2021] on a tile of directions  $\mathbf{d}'$ :

$$\mathcal{L}_S = \left( 0.5 \cdot \|[-1 \ 1] * C'(\mathbf{d}')\|_2^2 + 0.5 \cdot \left[ \begin{matrix} -1 \\ 1 \end{matrix} \right] * C'(\mathbf{d}')\|_2^2 \right). \quad (9)$$

## 3 RESULT

*Dataset.* We rendered four synthetic scenes, namely SHIP, TORUS, DEERGLOBE and STARLAMP, from viewpoints sampled on a full sphere with refraction and total reflection effects. The viewpoints are 100, 100, and 200 views of size  $800 \times 800$  for training, validation, and testing splits, respectively. We resize all the images by half for experiments. Moreover, we captured one real scene (DOLPHIN) from viewpoints sampled upon a hemisphere. The viewpoints are 100, 50, and 100 views of size  $2560 \times 1920$  for training, validation, and testing splits, respectively, and the camera poses are calibrated with AprilTag [Krogjus et al. 2019]. We resize all the images by half and crop the center for experiments. We also select three real scenes from Eikonal Fields [Bemana et al. 2022], namely BALL, GLASS and PEN, and compare to the provided video sequences.

*Experimental setting.* We choose PSNR and SSIM for low-level image similarity, and LPIPS for better mimicking human preference as our evaluation metrics. We set  $N_c = 64$ ,  $N_f = 128$  and 200k training iterations with batch rays 1024 for mip-NeRF [Barron et al. 2021] and ours. Moreover, during the first 2.5k warm-up iterations, only the re-rendering error in Eqn. 7 is used. Note that we crop the object region of an image for evaluating the real scenes.

*Ablation study.* We validate our design choices with two experiments on the synthetic scenes. *Ours w/o H* uses no hierarchical sampling but with  $N_c = 256$ . *Ours w/o BD* uses no additional boundary regularizer. The result in Fig. 4 with boundary regularizer shows the better refraction on the surface such as the green box of SHIP, and the hierarchical sampling further preserves the details such as the wave in the green box of DOLPHIN.

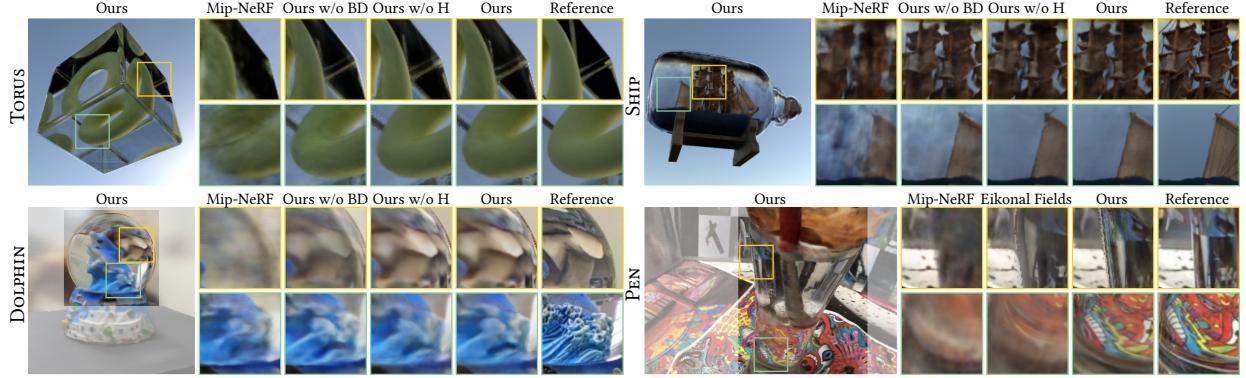


Figure 4: Qualitative comparison of the selected synthetic (top) and real (bottom) scenes.

*Competing method.* We compare our method with mip-NeRF [Barron et al. 2021] and Eikonal Fields [Bemana et al. 2022]. The results compared with mip-NeRF show that our method achieves a better performance of LPIPS across all the scenes (see Table 1). For the comparison against mip-NeRF in Fig. 4, our method preserves much more details (SHIP) and generates less blurry results (DOLPHIN). Then, we compare Eikonal Fields on the selected real scenes. Our method obtains a comparable LPIPS, and Eikonal Fields cannot reconstruct the refractive index of DOLPHIN scene (see Table 1). As shown in Fig. 4, our method could faithfully generate plausible results with better clearness than Eikonal Fields (PEN).

## 4 DISCUSSION AND FUTURE WORK

We present a NeRF-based framework that synthesizes the refraction in novel views and achieves better human perception performance in several scenes. The results show that explicitly tracking curved paths traversing through different refractive indices can produce more visually plausible refraction. Furthermore, with the help of sampling techniques and a boundary regularizer, our framework can further improve surface details and clarity. Our method still has limitations. The blurry geometric details in real scenes result from the imperfect camera poses compared to the synthetic data, and the foggy artifacts appear on refractive surfaces. In the future, we plan to tackle relighting via environment mapping to enable

**Table 1: Quantitative comparison of the selected synthetic and real scenes. The top three methods of each metric for a scene are marked by gold, silver and bronze.**

	TORUS			SHIP			DEERGLOBE			STARLAMP		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Mip-NeRF	22.32	0.759	0.268	23.93	0.828	0.151	26.79	0.881	0.134	22.04	0.885	0.092
Ours	25.46	0.853	0.130	24.76	0.840	0.122	27.43	0.896	0.109	22.08	0.878	0.086
Ours w/o H	25.57	0.852	0.136	24.77	0.838	0.127	27.58	0.894	0.108	22.08	0.876	0.091
Ours w/o BD	25.87	0.847	0.133	25.01	0.838	0.127	30.03	0.906	0.089	21.83	0.866	0.102
	DOLPHIN			BALL			GLASS			PEN		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Mip-NeRF	18.48	0.459	0.479	16.29	0.523	0.418	18.44	0.475	0.414	19.13	0.487	0.428
Eikonal Fields -	-	-	-	18.38	0.583	0.239	17.89	0.436	0.303	18.83	0.485	0.335
Ours	18.35	0.430	0.416	17.62	0.491	0.275	18.35	0.440	0.306	18.95	0.494	0.315

novel views under a new illumination and optimizing a voxel grid of refractive index to handle more complex refractive objects.

## REFERENCES

- Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.
- Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. 2022. Eikonal Fields for Refractive Novel-View Synthesis. *arXiv preprint arXiv:2202.00948* (2022).
- Zekun Hao, Arun Mallaya, Serge Belongie, and Ming-Yu Liu. 2021. Gancraft: Unsupervised 3d neural rendering of minecraft worlds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14072–14082.
- Ivo Ihrke, Gernot Riegler, Art Tevs, Christian Theobalt, Marcus Magnor, and Hans-Peter Seidel. 2007. Eikonal rendering: Efficient light transport in refractive objects. *ACM Transactions on Graphics (TOG)* 26, 3 (2007), 59–es.
- Maximilian Krogjus, Acshi Haggenmüller, and Edwin Olson. 2019. Flexible Layouts for Fiducial Tags. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard McMillan. 2002. Acquisition and Rendering of Transparent and Refractive Objects. In *Proceedings of the 13th Eurographics Workshop on Rendering (Pisa, Italy) (EGRW '02)*. Eurographics Association, Goslar, DEU, 267–278.
- 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 *European conference on computer vision*. Springer, 405–421.
- Xin Sun, Kun Zhou, Eric Stollnitz, Jiaoying Shi, and Baining Guo. 2008. Interactive relighting of dynamic refractive objects. In *ACM SIGGRAPH 2008 papers*. 1–9.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. 2021. Ref-Nerf: Structured View-Dependent Appearance for Neural Radiance Fields. *arXiv preprint arXiv:2112.03907* (2021).
- Ziyu Wang, Wei Yang, Junming Cao, Lan Xu, Junqin Yu, and Jingyu Yu. 2022. NeReF: Neural Refractive Field for Fluid Surface Reconstruction and Implicit Representation. *arXiv preprint arXiv:2203.04130* (2022).
- Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492* (2020).
- Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.
- Junqiu Zhu, Yaoyi Bai, Zilin Xu, Steve Bakó, Edgar Velázquez-Armendáriz, Lu Wang, Pradeep Sen, Miloš Hašan, and Ling-Qi Yan. 2021. Neural complex luminaires: representation and rendering. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.