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

Reference
Mip-NeRF
Ours
Reference
Mip-NeRF
Ours

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 SIGGRAPH Asia 2022 Technical Communications (SA ’22 Technical Communications), December 6–9, 2022, Daegu, Republic of Korea. 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 radiation 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 $F_0$ and one fine $F_\phi$, to represent a volumetric scene. The MLP takes the encoded position $x$ of a sample and view direction $d$ by positional encoding as inputs; it outputs the density $\sigma$ and radiance $c$. The pixel value $\hat{C}$ is estimated by the differentiable volume rendering equation in Eqn. 1 with all the samples $r$ along a ray cast from the camera origin $o$ to the pixel.

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

(1)

where $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$ 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 $r_c$ and one fine $r_f$, for the coarse and fine networks. Specifically, the fine network takes the union of both sets after sorting. The coarse samples $r_c = \{(x_i, d_i)\}_{i=0}^{N_c}$ 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 $r_f = \{(x_i, d_i)\}_{i=0}^{N_f}$ 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 $x_i = o + t_i d$.

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

(2)

$$t_f = \{t_i \sim \text{InverseTransformSampling} (\hat{w}_i)\}_{i=1}^{N_f}$$  

(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 $d$ 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 [Bar-ron et al. 2021] proposes an integrated positional encoding and achieves the best quality for both single- and multi-scale contents. Ref-NeRF [Verbini 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, Matsuk 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:

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

**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:

$$x_{i+1} = x_i + \frac{\Delta}{n} n_i, \quad v_{i+1} = v_i + \Delta \nabla n,$$  

(4)

where $\Delta$ is step size, $v$ is defined by $\frac{\Delta}{n}$, $n$ is refractive index and $\nabla n$ 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.
Within a voxel that are outside and inside the proxy geometry, voxel grid with tri-linear interpolation to represent the refractive density. 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 of the first derivative of the refractive surface: $C'$ 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 $d_N$ is the leaving direction from the bounded volume.

Finally, we optimize the three MLPs, namely $F_0$, $F_\phi$ and $F_{\psi}'$, with respect to the following objective function:

$$
\mathcal{L} = \lambda_{\text{RGB}} \mathcal{L}_{\text{RGB}} + \lambda_{\text{BD}} \mathcal{L}_{\text{BD}} + \lambda_{\text{S}} \mathcal{L}_{\text{S}},
$$

where $\lambda_{\text{RGB}}$, $\lambda_{\text{BD}}$ and $\lambda_{\text{S}}$ 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 $C(r)$ as NeRF:

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

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:

$$
\mathcal{L}_{\text{BD}} = \mathbb{1}(T_f, N_c + N_f + 1) \cdot ||C(r) - T_f, N_c + N_f + 1, C'(d_f, N_c + N_f)||_1.
$$

Smoothness regularizer. We add an L2 gradient penalty to boundary as NeRFactor [Zhang et al. 2021] on a tile of directions $d'$:

$$
\mathcal{L}_S = 0.5 \cdot || -1 \cdot C'(d') ||_2^2 + 0.5 \cdot || -1 \cdot C'(d') ||_2^2.
$$

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 [Krogius 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.
We present a NeRF-based framework that synthesizes the refraction of refractive objects. As shown in Fig. 4, our method could faithfully generate plausible results in novel views and achieves better human perception performance compared with mip-NeRF (see Table 1). For the comparisons on the synthetic scenes, 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 novel views under a new illumination and optimizing a voxel grid of refractive index to handle more complex refractive objects.

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.

|      | PSNR | SSIM | LPIPS | PSNR | SSIM | LPIPS | PSNR | SSIM | LPIPS |
|------|------|------|-------|------|------|-------|------|------|-------|
|      | DOLPHIN | Torus | SHIP | DeerGlobe | STARLAMP |
| Mip-NeRF | 22.32 | 0.759 | 0.268 | 24.70 | 0.861 | 0.209 |
| Ours | 25.46 | 0.833 | 0.110 | 27.17 | 0.916 | 0.197 |
| Ours w/o BD | 25.57 | 0.852 | 0.136 | 27.78 | 0.894 | 0.108 |
| Ours w/o H | 25.87 | 0.842 | 0.133 | 28.06 | 0.890 | 0.089 |

REFERENCES

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Bruxalla, 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.

Zekun Hao, Arun Mallya, 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 Ihle, Gernot Ziegler, 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 Krogus, Aeshi Haggenmüller, and Edwin Olson. 2019. Flexible Layouts for Facial Tags. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

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, Junqing Yu, and Jingyi Yu. 2022. NeReF: Neural Refractive Field for Fluid Surface Reconstruction and Implicit Representation. arXiv preprint arXiv:2203.01430 (2022).

Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020. NeRF+: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010.07492 (2020).

Xuxing Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nefactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (TOG) 40, 6 (2021), 1–18.

Jen-I Pan, Jheng-Wei Su, Kai-Wen Hsiao, Ting-Yu Yen, and Hung-Kuo Chu...