# Domain Reduction Strategy
# for Non-Line-of-Sight Imaging

Hyunbo Shim*, In Cho*, Daekyu Kwon, and Seon Joo Kim

Yonsei University

**Abstract.** This paper presents a novel optimization-based method for non-line-of-sight (NLOS) imaging that aims to reconstruct hidden scenes under general setups with significantly reduced reconstruction time. In NLOS imaging, the visible surfaces of the target objects are notably sparse. To mitigate unnecessary computations arising from empty regions, we design our method to render the transients through partial propagations from a continuously sampled set of points from the hidden space. Our method is capable of accurately and efficiently modeling the view-dependent reflectance using surface normals, which enables us to obtain surface geometry as well as albedo. In this pipeline, we propose a novel domain reduction strategy to eliminate superfluous computations in empty regions. During the optimization process, our domain reduction procedure periodically prunes the empty regions from our sampling domain in a coarse-to-fine manner, leading to substantial improvement in efficiency. We demonstrate the effectiveness of our method in various NLOS scenarios with sparse scanning patterns. Experiments conducted on both synthetic and real-world data support the efficacy in general NLOS scenarios, and the improved efficiency of our method compared to the previous optimization-based solutions. Our code is available at https://github.com/hyunbo9/domain-reduction-strategy.

**Keywords:** Computational imaging · Non-line-of-sight imaging

## 1 Introduction

The ability to see objects placed over the direct line-of-sight has the potential to advance various applications, *e.g.*, autonomous driving, medical care, and rescue operations. Non-line-of-sight (NLOS) imaging aims to reconstruct hidden scenes from measurements of indirect reflections. These transient measurements are commonly obtained via "looking around the corner", where a light source and a time-of-flight sensor illuminate and scan the relay wall (see Fig. 1 (a)).

Recent advancements in NLOS imaging have been achieved by FFT-based inverse solutions, including 3D convolution-based methods [1, 32, 50], and wave-based methods [22, 25]. Despite promising results, one or more assumptions are made to model the exact solutions, *e.g.*, planar relay walls or certain scanning
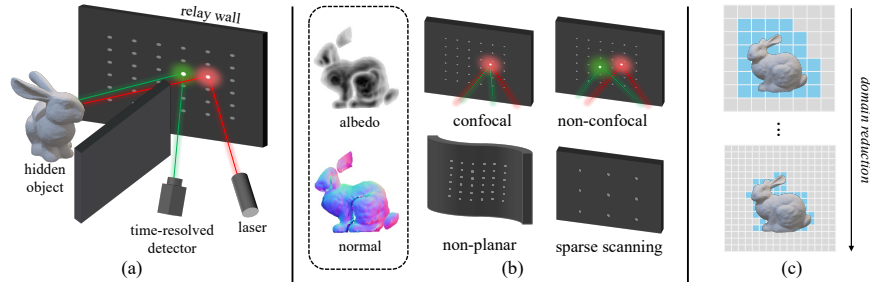
---

* Equal contribution.

**Fig. 1:** **(a)** A common NLOS scanning system. A laser and a time-resolved sensor illuminate and scan the relay wall. **(b)** Our method reconstructs both albedo and surface normal of the hidden objects in general scenarios, including non-confocal, non-planar relay walls and sparse sampling. **(c)** Our domain reduction gradually prunes empty regions in a coarse-to-fine manner, achieving significant efficiency improvement.

systems. These methods also necessitate a dense scanning procedure to reconstruct high-resolution outputs, which is too time-consuming for practical uses.

Considering the practical applications of NLOS imaging, there are numerous scenarios where the exact solutions are not sufficient to meet the demands. In such scenarios, an ideal relay wall may not be available, and dense scanning of the relay wall becomes impractical. There is also a need to reconstruct precise surface geometry. These scenarios give rise to the need for solutions capable of solving NLOS imaging under broader settings. Several works have attempted to meet this demand with optimization-based NLOS methods [13,33,39,43]. Albeit versatile, they require substantial computational resources per iteration, with a significant portion dedicated to the unoccupied regions of the hidden scenes.

In this paper, we present a novel optimization-based method that inherits the versatility of previous optimization-based solutions while simultaneously enhancing efficiency. In NLOS imaging scenarios, the visible surfaces of the target objects are notably sparse, occupying less than 5% regions of the entire hidden space. Since the empty regions do not actually contribute to both transients and the target volumes, we aim to exclude these unnecessary computations.

Pruning the empty regions requires processing varying size inputs. Therefore, we tailor our method to render the transients by partially propagating the light returning from a set of an arbitrary number of points. We begin by modeling the transients as a superposition of point-wise light propagation functions. Given a set of points sampled from the hidden space, light propagation of each point is computed using its albedo and surface normal. The predicted transients are obtained by a linear combination of the computed propagations, optimized to reconstruct the target volumes through inverse rendering. Our pipeline is accurately and efficiently implemented through a matrix-free CUDA kernel. This facilitates our method to accurately model both surface geometry and albedo, without being constrained to specific types of relay walls and scanning systems.

Once we model the transients as a superposition of point-wise propagations, superfluous computations for the empty regions can be easily removed. Here, we

propose a novel domain reduction strategy to eliminate such unnecessary computations, resulting in a notable improvement in the efficiency of the optimization process. During the optimization process, our domain reduction strategy periodically identifies regions of which albedo becomes lower than a certain threshold and prunes these empty regions from our sampling domain. Sampling a set of points only from the active regions effectively boosts the efficiency of our method, achieving about 20× acceleration of the reconstruction time. This domain reduction procedure is conducted in a coarse-to-fine manner, enabling our method to reconstruct high-resolution output volumes with a single commercial GPU.

We demonstrate the validity of our method in various NLOS scenarios, including a non-planar relay wall, sparse scanning patterns, confocal and non-confocal, and surface geometry reconstruction (Fig. 1 (b)). Our experimental results from both synthetic and real-world inputs confirm that our method can efficiently reconstruct hidden objects with fine details across various scenarios. Notably, our method is capable of producing 128×128 outputs in under a minute.

## 2   Related Work

**NLOS imaging.** NLOS imaging was first introduced by Kirmani *et al.* [19], and experimentally validated in [44] by using a femtosecond laser and a streak camera. Various methods have been developed for NLOS imaging, which can be broadly divided into two parts: active [8, 15, 17, 22, 32, 42, 45, 47, 52] and passive methods [4–6, 36, 37, 41, 49]. Passive methods rely on indirect light under various setups, *e.g.*, using videos of a blank wall [38], a large transmissive window [40], or recovering videos projected in a hidden scene [2]. On the other hand, active NLOS methods utilize controllable a light source and a time-of-flight detector, usually producing higher-quality results with the use of more abundant information. In the active NLOS setup, the pulsed laser emits photons to the visible wall and measures the amount of returning photons to reveal the shape of the hidden object. The active NLOS imaging methods has been rapidly enhanced with various input sources, such as acoustic waves [21], and diverse types outputs, *e.g.*, albedo [1, 32], normal [11, 50], surface [43], and depth [9].

**Inverse NLOS methods.** Significant development has been recently made in NLOS imaging through various inverse solutions. By assuming isotropic light scatters and a planar relay wall, LCT [32] demonstrates that the NLOS imaging problem can be interpreted as a 3D deconvolution problem under the confocal setup. Their main concept is further extended by DLCT [50], which uses a vector deconvolution to obtain both the surface normal and albedo. Lindell *et al.* [22] present a wave-based solution named f-k migration. Phasor field methods [25, 26] formulate the NLOS imaging problem as a diffractive wave propagation and utilize the Rayleigh-Sommerfeld integral. Despite impressive results, they require several assumptions, *e.g.* planar relay wall, specific (*e.g.*, confocal) scanning system and exhaustive raster scanning for high-resolution outputs. In addition, methods except DLCT [50] can only reconstruct the albedo of hidden scenes.

Although such constraints allow faster computations, general applicability and accurate surface modeling are sacrificed in most inverse NLOS methods.

**NLOS methods for general setups.** A number of studies have explored NLOS imaging algorithms that can be applied in various environments, *e.g.* diverse acquisition patterns and geometry [12, 16, 27, 35, 48], and imaging over long ranges [46]. Several works utilize additional hardware for various purposes, such as SPAD arrays for faster scanning [30,31,33], which are often expensive for real-world applications, and new hardware designs for dynamic relay walls [20]. Back-projection (BP) solvers [1, 3, 44] are one of the commonly used techniques for general NLOS imaging. Iterative optimization-based methods [14,34,43] could be effective solutions for NLOS imaging in challenging real-world scenarios. Ahn *et al.* [1] propose the iterative approach that can operate without being bound to a certain scan system and the lateral resolution of the transients. Heide *et al.* [13] propose a volumetric-based optimization method that can handle partial occlusion and surface normal. The optimization-based method using a point spread function [33] was also proposed. Tsai *et al.* [43] propose the surface optimization method that can reconstruct the continuous surfaces of the hidden scenes, of which results are sensitive to the initial state. NeTF [39] proposes neural representations similar to NeRF [29] for NLOS imaging. Importantly, most of them suffer from huge computational costs of the optimization, usually $O(N^5)$ or $O(N^6)$ per iteration. This paper aims to inherit the general applicability of the optimization-based methods and to address expensive computation overheads of the optimization by identifying and pruning unnecessary computations.

## 3   Method

### 3.1   NLOS Measurement Model

We begin by formulating the transient imaging model for an arbitrary measurement environment. In the active NLOS methods, a time-resolved sensor measures returning photons after light pulse is emitted onto a relay wall. Let $\mathbf{p}$ be a point in the hidden volumes which follows Lambert's cosine law. We assume no interactions between the surfaces, including inter-reflections and self-occlusion. The measured transients can be formulated as

$$\tau(t, \mathbf{l}, \mathbf{s}) = \int_{\mathbf{p} \in \Omega} \rho(\mathbf{p}) \cdot \Phi_{\mathbf{p}}(\mathbf{l}, \boldsymbol{n}(\mathbf{p})) \cdot \Upsilon_{\mathbf{p}}(\mathbf{l}, \mathbf{s}) \cdot \delta(d_{\mathbf{l}} + d_{\mathbf{s}} - tc) \, \mathbf{dp},$$

$$\Phi_{\mathbf{p}}(\mathbf{l}, \boldsymbol{n}(\mathbf{p})) = \langle \frac{\mathbf{l} - \mathbf{p}}{||\mathbf{l} - \mathbf{p}||}, \boldsymbol{n}(\mathbf{p}) \rangle, \ \Upsilon_{\mathbf{p}}(\mathbf{l}, \mathbf{s}) = \frac{1}{d_{\mathbf{l}}^2} \cdot \frac{1}{d_{\mathbf{s}}^2}, \tag{1}$$

where $\mathbf{l}$ and $\mathbf{s}$ denote the laser and scan point at the relay wall, and $\rho(\mathbf{p})$ and $\boldsymbol{n}(\mathbf{p})$ are the albedo and the surface normal of point $\mathbf{p}$ in the hidden space $\Omega$. $\Phi_{\mathbf{p}}(\mathbf{l}, \boldsymbol{n})$ models the view-dependent cosine term according to Lambert's cosine law. $\Upsilon_{\mathbf{p}}(\mathbf{l}, \mathbf{s})$ models the distance fall-off, where $d_{\mathbf{l}}$, $d_{\mathbf{s}}$ are the distances from the hidden scene to the laser and the scan point respectively. The Dirac delta $\delta(\cdot)$ relates time to the light travel distance, $c$ is the speed of light and $t$ is
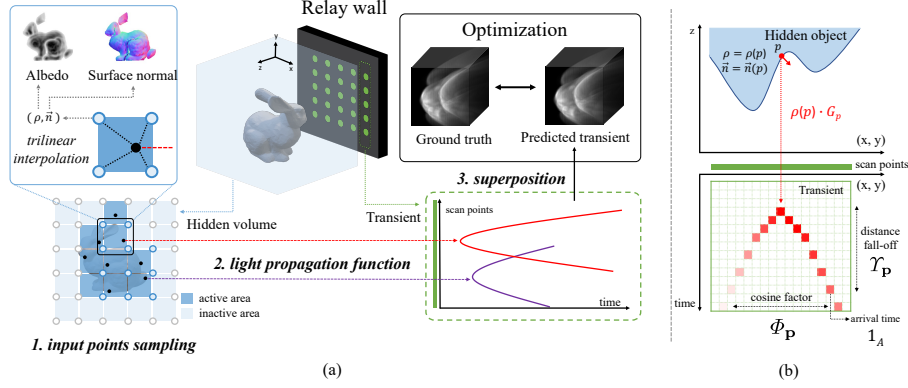
**Fig. 2: (a)** The overview of our reconstruction scheme. We divide hidden space to a grid shape and assign albedo $\rho$ and surface normal $\boldsymbol{n}$ for each vertex. Input points are randomly sampled from the hidden space. Then the light propagation from each point $G_{\mathbf{p}}$ is computed, of which superposition is used as predicted transients. The variables are optimized by minimizing the L2 distance. Our domain is gradually reduced by pruning the empty regions during the optimization. **(b)** The computation of light propagation function from each point $\mathbf{p}$. The arrival time at scan point $\mathbf{s}$ is first identified, and then the fall-off terms $\Phi_{\mathbf{p}}$ and $\Upsilon_{\mathbf{p}}$ are computed using $\mathbf{p}$, $\mathbf{l}$, $\mathbf{s}$ and $\boldsymbol{n}(\mathbf{p})$.

the arrival time of photons. Note that our formation model is defined without assumptions on relay walls and scanning systems. By additionally assuming a confocal scanning system and a planar relay wall, our formation model becomes equivalent to the formation model of DLCT [50].

Since each point $p$ in $\Omega$ independently contributes to the transients in our formation model, we can represent the transients as a superposition of functions that models the light propagation from each $p$. To decompose the transients into a set of point-wise propagation functions, we first model the light propagation function for given $\mathbf{p}$:

$$g_{\mathbf{p}}(t, \mathbf{l}, \mathbf{s}) = \Phi_{\mathbf{p}}(\mathbf{l}, \boldsymbol{n}(\mathbf{p})) \cdot \Upsilon_{\mathbf{p}}(\mathbf{l}, \mathbf{s}) \cdot \delta(d_{\mathbf{l}} + d_{\mathbf{s}} - tc). \tag{2}$$

The propagation function $g_{\mathbf{p}}$ determines the arrival time, distance fall-off and cosine factor according to $\mathbf{l}$ and $\mathbf{s}$. It can be interpreted as a tailored point spread function that depends on $\mathbf{p}$, $\mathbf{l}$, $\mathbf{s}$ and the surface normal $\boldsymbol{n}(\mathbf{p})$ at $\mathbf{p}$. With the above equation, the entire transients can be expressed as

$$\tau = \int_{\mathbf{p} \in \Omega} \rho(\mathbf{p}) \cdot g_{\mathbf{p}} \, \mathbf{dp}. \tag{3}$$

Each $\rho(\mathbf{p}) \cdot g_{\mathbf{p}}$ represents 3D sub-signals containing all measured photons returning from point $\mathbf{p}$. Our goal becomes revealing $\rho$ and $\boldsymbol{n}$ to $p$ that correspond to each decomposed function, without posing constraints on $\mathbf{l}$, $\mathbf{s}$ and without losing accurate modeling of light propagation.

### 3.2   Reconstruction Scheme

By ignoring $\Phi$ and $\Upsilon$, the above equation can be addressed using an elliptical Radon transform [24]. It would enable easier approaches to find the inverse, but the accurate modeling would also be sacrificed. Therefore, we aim to reconstruct $\rho$ and $\boldsymbol{n}$ of the hidden volumes through the optimization. Instead of using a discretized matrix with a fixed size, we leverage a set of an arbitrary number of points as inputs, which are continuously sampled from the hidden space.

Fig. 2 (a) illustrates the overview of our proposed reconstruction scheme. We first divide the hidden space into a grid and assign 4-dimensional variables, albedo $\rho$ and normal $\boldsymbol{n}$ to each vertex of the grid. In each optimization step, input points are sampled from the hidden space. Then the light propagation of each point is computed, which are composed to obtain the transient predictions. The 4-dimensional variables are optimized using gradient descent by minimizing the difference between the rendered and the ground truth measurements. We illustrate details of our optimization scheme in the following paragraphs.

**Input points sampling.** Instead of using fixed points as inputs, we randomly sample a set of continuous points, from the hidden space at every step. We utilize a grid-based random sampling technique to obtain a spatially balanced points set. We sample one point from each cell of the 3D voxel grid, resulting in the set of points $H$. The albedo and normal at each point are determined by the trilinear interpolation of the surrounding 8 vertices. The normals are converted to unit vectors after the interpolation.

**Light propagation and superposition.** To synthesize the transients with a sampled points set $H$, we discretize the transients in Equation (3) to a finite linear combination of $\{G_{\mathbf{p}}\}$, where $G_{\mathbf{p}}$ can be expressed as

$$
\begin{aligned}
G_{\mathbf{p}} &= \Phi_{\mathbf{p}} \cdot \Upsilon_{\mathbf{p}} \cdot 1_A, \\
A &= \{(t, \mathbf{l}, \mathbf{s}) \mid d_{\mathbf{l}} + d_{\mathbf{s}} = tc\}.
\end{aligned}
\tag{4}
$$

To compute $G_{\mathbf{p}}$, we first identify the non-zero histograms of each scan point $\mathbf{s}$ with given $\mathbf{l}$ and $\mathbf{s}$, by computing $t$ that satisfies the indicate function $1_A$. The contribution at the corresponding location is then determined by $\Phi_{\mathbf{p}}$ and $\Upsilon_{\mathbf{p}}$, which are calculated as in Fig. 2 (b). Computations of light propagation are done in parallel for all points in $H$, and thus can be accelerated by GPU. Finally, we compute the predicted transients with a linear summation:

$$
T = \sum_{\mathbf{p} \in H} \rho(\mathbf{p}) \cdot G_{\mathbf{p}} \cdot \Delta V,
\tag{5}
$$

where $\Delta V$ is volume of a grid bin. We update $\rho$ and $\boldsymbol{n}$ using the gradient descent, by minimizing the L2 distance between rendered and ground truth transients.

**Noise regularization.** The sensor noise present in NLOS imaging could adversely affect to our optimization process. To mitigate this issue, we utilize L1 regularization to the albedo to encourage sparsity, and jointly optimize additional noise parameters $d(t, \mathbf{l}, \mathbf{s})$ to cover the effects of the background noise, including both ambient light and dark counts. Specifically, these learnable parameters

are defined and assigned for each transient histogram $\tau(t, \mathbf{l}, \mathbf{s})$. To constrain the range of these parameters, and thereby prevent them from covering too much intensities of the rendered transients, we model the noise parameters $d(t, \mathbf{l}, \mathbf{s})$ as

$$d(t, \mathbf{l}, \mathbf{s}) = b + \lambda \sigma(z(t, \mathbf{l}, \mathbf{s})), \tag{6}$$

where $b$ is the base noise level controlling the minimum values of the noise parameters, $\sigma$ denotes the sigmoid activation function, and $z$ is the learnable parameters that are jointly optimized throughout the optimization process. $\lambda$ is the hyperparameter controlling the maximum values of $d(t, \mathbf{l}, \mathbf{s})$. These simple techniques greatly improves robustness of our method to the noise.

### 3.3 Domain Reduction Strategy

One crucial observation is that most of the hidden volumes to reconstruct is empty in the NLOS imaging scenarios, since photons are only reflected from the visible surfaces of the objects. These empty regions do not actually affect either the measured transients or the reconstruction volumes. By taking the advantages of this sparsity nature, most of the computations in synthesizing transients can be eliminated, as our reconstruction pipeline is capable of computing transients with an arbitrary number of points. The synthesized transients in Equation (5) can be divided as

$$T = \sum_{\mathbf{p} \in \Omega_1} \rho(\mathbf{p}) \cdot G_{\mathbf{p}} \cdot \Delta V + \sum_{\mathbf{p} \in \Omega_2} \rho(\mathbf{p}) \cdot G_{\mathbf{p}} \cdot \Delta V,$$
$$\Omega_1 = \{ \, \mathbf{p} \in \Omega' | \, \rho(\mathbf{p}) > \epsilon \, \}, \Omega_2 = \Omega' \setminus \Omega_1, \tag{7}$$

where $\Omega' \subset \Omega$ is a set of points that corresponds to the vertices in the grids. Since the values in the summation are always positive, we can ignore the operations of the second term if $\epsilon$ is a sufficiently small value ($||\rho(\mathbf{p}) \cdot G_{\mathbf{p}}||_\infty \to 0$ as $\epsilon \to 0$ $\forall \mathbf{p} \in \Omega_2$). Hence, we can exclude $\Omega_2$ from our domain and eliminate unnecessary computations on $\Omega_2$. This domain reduction procedure is periodically conducted during the optimization, significantly accelerating the efficiency of our method.
**Soft domain reduction.** We observed that the albedo $\rho$ often takes on incorrect values during the optimization, which can result in the accidental removal of non-empty regions from the domain. To this end, we propose a soft domain reduction strategy that gently reduces the domain with a low-pass filter. Specifically, we apply a Gaussian kernel to the albedo volumes and then we exclude the grid bins with values smaller than a certain threshold from the domain. This soft domain reduction expands the domain to the surrounding areas and provides additional chances to be optimized for the accidentally removed regions.
**Coarse to fine strategy.** Predicting transients with a large number of sampled points is computationally exhaustive, while the coarse shapes of the hidden objects can be sufficiently discovered with coarsely sampled input points. Therefore, we initialize the hidden space with a coarse grid and gradually increase the resolution to progressively reconstruct fine details. After sufficiently reducing the domain, we expand the grid to a higher resolution. The grid bins within

---

**Algorithm 1** Reconstruction Algorithm.

---

1:  $\Omega' \leftarrow$ Set of points of grid vertices of hidden scene
2:  **for** $step \leftarrow 1, 2, ..., N$ **do**
3:      $S \leftarrow$ Set of variables $\rho$ and $\boldsymbol{n}$ on $\Omega'$
4:      $H \leftarrow$ Set of points, one for each bin on $\Omega'$
5:      $T \leftarrow$ Linear combination of $\{G_p\}_{p \in H}$                    $\triangleright$ Equation (5)
6:      $T \leftarrow T+$ Synthesized noise                               $\triangleright$ Equation (6)
7:      Update $\rho, \boldsymbol{n} \in S$ by minimizing $\|T - \tau_{gt}\|_2$
8:      **if** $step \in reduction\_step$ **then**
9:          $\rho_{\texttt{low}} \leftarrow$ Low-pass filter$(\rho)$
10:         $\Omega' \leftarrow$ Domain reduction$(\Omega', \rho_{\texttt{low}})$            $\triangleright$ Equation (7)
11:     **end if**
12:     **if** $step \in expansion\_step$ **then**
13:         $\Omega' \leftarrow$ Dividing $\Omega'$ into a fine grid
14:         $S \leftarrow$ Trilinear interpolation$(S)$
15:     **end if**
16: **end for**
17: **return** $\rho, \boldsymbol{n}$

---

the active regions are subdivided into smaller bins, and variables of the divided bins are assigned values obtained with trilinear interpolation of previous coarse grids. Reconstructing in a coarse-to-fine fashion, which shares similar philosophy with various studies in 3D computer vision [23, 28, 39, 51], enables our method to efficiently reconstruct high-resolution outputs with fine details.

**Overall algorithm.** Equipped with the above equations and techniques, our method can be described as in Algorithm. 1. The optimization process fully exploits the parallel nature of our formulations with GPU implementations.

## 4    Experiments

We demonstrate the effectiveness of our method through the extensive experiments. We compare the results on transients having $32 \times 32$ scannings, with both confocal and non-confocal scanning systems, and with a non-planar relay wall.

**Baselines.** We compare our method with several state-of-the-art baselines: FK [22], DLCT [50], Phasor field [26] with a BP solver (without FFT), Gram [1], and NeTF [39]. We apply maximum intensity projection through the $z$ axis for all methods. For the results of DLCT, we follow the authors and report the $z$-directional albedo. We provide results of more baselines in Supplement.

**Dataset.** We validate the results on the ZNLOS [10] dataset and the Stanford real-world measurements [22]. ZNLOS consists of a number of synthetic transients, simulated with a 1 m $\times$ 1 m relay wall and hidden objects placed approximately 0.5 m from the relay wall. The temporal resolution of bins corresponds to the light travel time of 0.3 cm. The Stanford real-world measurements [22] are obtained with a 2 m $\times$ 2 m relay wall (1m $\times$ 1 m for a non-planar case) and a bin resolution 32 ps. The raw $512 \times 512$ measurements are first spatially

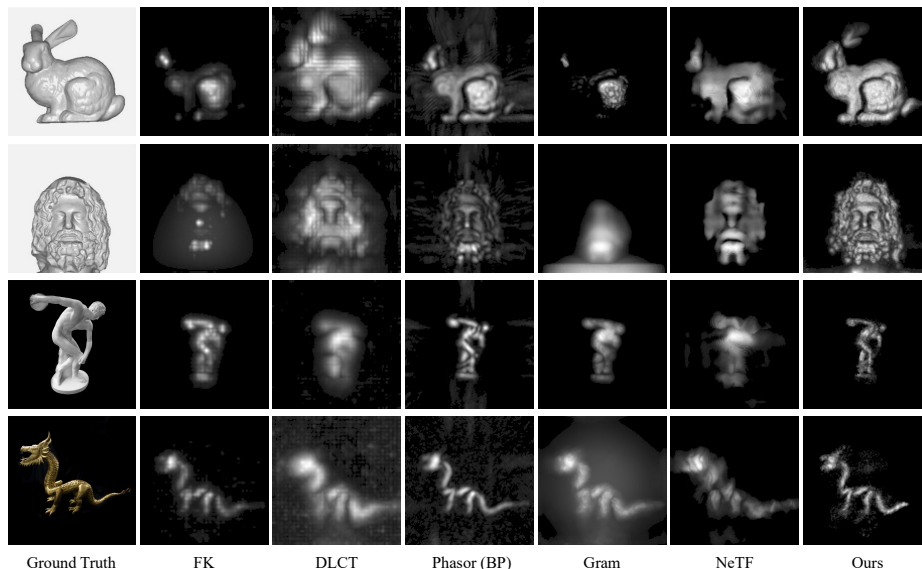| Ground Truth | FK | DLCT | Phasor (BP) | Gram | NeTF | Ours |

**Fig. 3:** Reconstructed albedo on the transients with a confocal $32 \times 32$ scan system and a planar relay wall. Our method achieves the highest reconstruction quality.

downsampled by 2, which results in $256 \times 256$ transients. We adopt Bunny and Serapis instances from ZNLOS dataset, and Statue, Dragon, NT instances from the Stanford dataset. The NT instance is used for the non-planar evaluation.

**Implementation detail.** Our method is implemented using PyTorch, and the light propagation function is implemented using the PyTorch CUDA extension. We optimize the variables using the Adam optimizer [18] with learning rate 1 and total 1000 steps. We set the reduction threshold as 5% (3% for non-confocal) of the maximum albedo value. The domain reduction step is performed at every 50 steps. We refer to Supplement for more implementation details.

**Sparse scanning.** It is worth mentioning that most previous studies [22,26,32, 50] have reported results on high-resolution transients obtained by dense scanning. Such exhaustive procedures require much longer scanning time, which is often difficult in several real-world scenarios. To demonstrate the applicability of our method in such scenarios, we compare the results on $32 \times 32$ sparsely sampled transients. These sparsely sampled measurements are obtained by taking the subset of the measurements with appropriate spatial strides.

### 4.1 Confocal NLOS Imaging Result

We first evaluate the performance of our method on sparsely sampled confocal measurements. As shown in Fig. 3, results of DLCT contain pale artifacts around the hidden objects, making some objects indistinguishable from these artifacts. FK [22] often fail to reveal fine details of the objects, *e.g.* the ear of Bunny and the head of Dragon. Phasor [26] with a BP solver yields some artifacts and some

**Table 1:** Quantitative results of Depth map and Normal map on the ZNLOS Bunny under confocal setting with threshold values of 5% and 10% of maximum albedo.

| Method | Depth (5%) | | Normal (5%) | | Depth (10%) | | Normal (10%) | |
|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| LCT [32] | 0.1977 | 0.2875 | - | - | 0.1048 | 0.2069 | - | - |
| FK [22] | 0.0719 | 0.1871 | - | - | 0.1023 | 0.2173 | - | - |
| Phasor (BP) [26] | 0.1348 | 0.2049 | - | - | 0.1283 | 0.2068 | - | - |
| Phasor (FFT) [25] | 0.1300 | 0.2808 | - | - | 0.0818 | 0.2095 | - | - |
| Gram [1] | 0.0913 | 0.1628 | - | - | 0.0751 | 0.1827 | - | - |
| NeTF [39] | 0.0679 | 0.1748 | - | - | 0.0681 | 0.1754 | - | - |
| DLCT [50] | 0.3189 | 0.4220 | 0.3796 | 0.4856 | 0.1449 | 0.2729 | 0.2021 | 0.3438 |
| Ours | **0.0477** | **0.1523** | **0.1147** | **0.2394** | **0.0493** | **0.1552** | **0.1115** | **0.2341** |
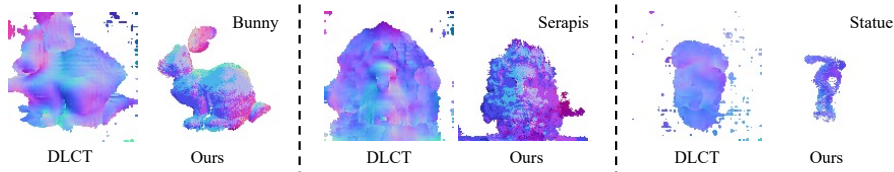


**Fig. 4:** Comparison of the reconstructed normal map with DLCT [50] on the transients with a confocal $32 \times 32$ scan system and a planar relay wall. Our method accurately reconstructs surface geometry, whereas DLCT only produces coarse structures.

details are missing. Gram [1] delivers blurry results or fails to capture several details. NeTF [39] tends to reconstruct the objects with blurry and distorted albedo. On the other hand, our method successfully reconstructs clean and sharp results with fine details on both synthetic and real-world transients.

**Normal maps.** To evaluate the capability of accurately reconstructing surface geometry, we deliver the reconstructed normal maps in Fig. 4. DLCT is unable to produce accurate surface geometry on $32 \times 32$ transients, indicating the necessity of raster-scanning for the inverse methods. On the other hand, our method delivers normal maps in high-quality, reconstructing details of the objects, and achieves significantly reduced scanning time compared to DLCT.

**Quantitative results.** We also quantitatively evaluate depth maps and normal maps on $32 \times 32$ transients of Bunny. We upsample the results of all methods to the resolution $256 \times 256$, matching that of the ground truth, with trilinear interpolation. We also apply 10% threshold to acquire depth and normal maps. We measure the mean absolute error (MAE) and root mean squared error (RMSE) for both depth and surface normals. As reported in Table 1, our method outperforms all other baselines in both depths and surface geometry, demonstrating that our method reconstructs accurate geometry as well as visual appearance.

## 4.2   Various NLOS Imaging Result

**Non-planar relay wall result.** We conduct the experiments on the real-world transients captured with a non-planar wall [22], to show that our method can be applied to arbitrary types of the sampling geometry. Following previous works
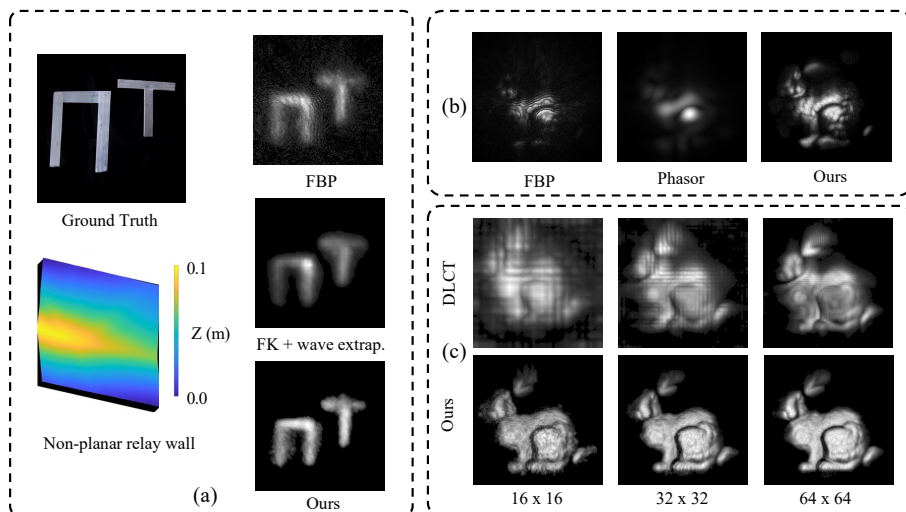
**Fig. 5:** Results on various setting. **(a)** Results on the real-world $32 \times 32$ transients of the NT instance [22], measured with a non-planar relay wall. **(b)** Results on non-confocal ZNLOS Bunny. **(c)** Results on Bunny with various sampling resolutions. Our method consistently delivers high-quality results in various scanning scenarios.

[32, 50], we slightly adjust our light propagation model in this experiment to properly deal with the retro-reflective targets. We compare the results with FBP [44] and FK [22]. The wave extrapolation technique is applied to report the results of FK as in their paper. Fig. 5 (a) demonstrates the results of non-planar relay wall setup. FBP is effective at reconstructing the silhouette, but its results tend to be noisy and show some distortions. FK [22] recovers the blurry shapes of the objects with the incorrect albedo values at some regions such as the shape of N. Our method reconstructs the clean shapes of the objects with more correct albedo values, showing the extensibility of our method beyond planar relay walls.

**Non-confocal NLOS imaging result.** We evaluate the performance of our method with a non-confocal scanning system. We present our results on the non-confocal $32 \times 32$ transients of Bunny. Methods designed to address the non-confocal measurements, Phasor [26] and FBP [44] are chosen and compared as baselines. Fig. 5 (b) reports the qualitative results on the non-confocal Bunny. FBP reconstructs only some parts of Bunny, in which streak artifacts can be observed. Phasor field fails to recover some details and yields the coarse and blurry output. Unlike these baselines, our method delivers much cleaner results with more detailed shapes, exhibiting general applicability of our method.

**Results with various sampling resolutions.** We report the results on the transients by gradually increasing the sampling resolutions by 2, from $16 \times 16$ to $64 \times 64$. We compare our results with DLCT [50], which shares most similar image formation model with our method and FFT-based method. As in Fig. 5 (c), our approach delivers plausible results with $16 \times 16$ samplings and
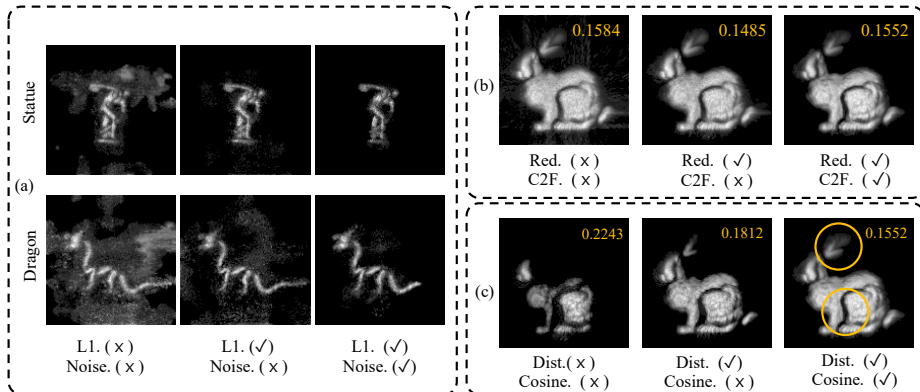
**Fig. 6:** Qualitative and qualitative ablation results. **(a)** Ablation results on the noise regularization techniques. 'L1' denotes the L1 regularization and 'Noise.' denotes optimized noise parameters. **(b)** Ablation results on the reduction. 'Red.' denotes the domain reduction and 'C2F' denotes the coarse-to-fine strategy. Both 'Red.' and 'C2F' do not sacrifice the reconstruction quality. **(c)** Ablation results on the accurate light modeling. 'Dist.' denotes the distance fall-off term and 'Cosine.' denotes the cosine factor. The upper-right values denote RMSE of the depth maps.

already achieves high-quality outputs with $32 \times 32$ samplings. On the other hand, DLCT fails to produce comparable outputs with our method until the sampling resolution reaches $64 \times 64$. Interestingly, no meaningful enhancement is observed by increasing the sampling resolution higher than $32 \times 32$, indicating that sparsely sampled $32 \times 32$ scan points are sufficient for our method to achieve high-resolution results, without requiring exhaustive raster-scanning.

### 4.3  Analysis and Ablation Study

In this section, we provide deeper insights and analysis of our proposed concepts. All analysis and ablations are conducted on the confocal $32 \times 32$ transients of Bunny, except for the ablation the noise regularization techniques.

**Noise regularization.** We conduct the ablation on our noise regularization techniques to assess the robustness of our method in the presence of sensor noise. Fig. 6 (a) delivers the results on the real-world measurements of Statue and Dragon. Results of the model without any regularization are severely affected by the presence of noise (No reg.). Imposing L1 regularization alleviates the effects of the noise, but we can still observe some noisy artifacts (L1 only). With the jointly optimized noise parameters, our method achieves clean results with much reduced artifacts, showing the robustness of our method to the noise.

**Domain reduction and runtime analysis.** We conduct the ablation on our domain reduction strategy to discover the efficiency improvement of our method. We compare three models, which are without domain reduction, domain reduction without coarse-to-fine strategy, and the proposed method. We report total runtime and the ratio of remaining active regions at 100, 500, 1000 steps,

**Table 2:** Ablation results on the domain reduction. 'Red.' denotes the domain reduction and 'C2F' denotes the coarse-to-fine strategy. We report elapsed time (seconds) of the optimization and the ratio of remaining domain in the parentheses at each step.

| Red. | C2F | 100 iter | 500 iter | 1000 iter |
|------|-----|----------|----------|-----------|
|      |     | 109 s | 539 s | 1087 s |
| ✓    |     | 67 s (11%) | 103 s (4%) | 134 s (4%) |
| ✓    | ✓   | 4 s (15%) | 27 s (3%) | 54 s (3%) |

**Table 3:** Comparisons of GPU memory usage (MB) with FFT-based methods, where $N$ is the hidden volume resolution. Our domain reduction (DR) significantly reduces the consumption especially when the target resolution is high. We use GPU versions [7] of LCT and FK.

| Resolution | LCT | FK | Ours w/o DR | Ours w/ DR |
|------------|-----|-----|-------------|------------|
| $N = 32$ | 1473 | 1445 | 1187 | 1211 |
| $N = 128$ | 7101 | 6813 | 4065 | 1659 |

with the final reconstructed albedo. As shown in Fig. 6 (b) and Table 2, our model with coarse-to-fine domain reduction achieves substantial efficiency improvement, without sacrificing the quality of the outputs. Compared with the model without reduction strategy, reconstruction time of our method is significantly reduced. The model without coarse-to-fine strategy suffers from heavy computations at the early stage, where the empty regions are not sufficiently pruned yet 100 steps. Thanks to the coarse-to-fine domain reduction strategy, our method efficiently eliminates most of the unnecessary computations with a low-resolution grid, achieving more than $20\times$ efficiency improvement as a result.

**Accurate light modeling.** We discover the effects of accurate light modeling through the ablation on the cosine factor and the distance fall-off term. We compare the results of three variants of our method: optimized without both cosine factor and distance fall-off, model assuming isotropic scatter (no cosine factor), and model with both terms. As reported in Fig. 6 (c), The results of both experiments with less accurate modeling are degraded. The model without both terms shows the worst result, only recovering the silhouette of the bunny. The model only without cosine term reconstructs almost all details, but still fails to recover some details (circled parts). These results support that the accurate modeling of light propagation leads to the improvement.

## 5   Discussion

**Comparison of general applicability.** We illustrate the general applicability of the NLOS imaging algorithms in Table 4. As can be seen, none of the methods exhibit general applicability as our method except NeTF [39]. The achievable resolution of all FFT-based inverse methods are bounded to the sampling resolution, which inevitably requires time-consuming scanning procedures. Although NeTF could operate without being limited to all types of constraints, the training time of NeTF typically exceeds 2 days, limiting the applicability of this method. Our method, on the other hand, is capable of reconstruct hidden scenes without being limited to scanning resolution, a planar relay wall, and can accurately reconstruct surface normals as well as albedo, typically in a minute.

**Table 4:** Comparison of the general applicability of various NLOS imaging algorithms.

| Method | Scan system | Sparse scanning | Non-planar | Surface recon. |
|--------|-------------|-----------------|------------|----------------|
| LCT [32] | Confocal | | | |
| FK [22] | Confocal | | △ | |
| DLCT [50] | Confocal | | | ✓ |
| Phasor [26] | - | | | |
| FBP [44] | - | ✓ | ✓ | |
| Gram [1] | - | ✓ | | |
| NeTF [39] | - | ✓ | ✓ | |
| Ours | - | ✓ | ✓ | ✓ |

**Comparison with FFT-based methods.** The FFT-based methods exhibit $O(N^3 log N)$ complexity and faster reconstruction time than our method, where $N$ is the spatial and temporal resolution. However, exhaustive raster-scanning procedures required for these methods serve as a major bottleneck, which takes about 46 minutes to achieve $128 \times 128$ outputs on the Stanford dataset. While FFT-based solutions have faster runtime, our optimization-based pipeline can reconstruct high-quality outputs only using $32 \times 32$ samplings, requiring scanning time less than 3 minutes. Considering both scanning and reconstruction time, our method can serve as an efficient solution for real-world applications. We provide more results and discussion on sparse samplings in Supplement.

**Reconstruction time and memory usage.** The time complexity of our method $O(cM^2N^3)$ per iteration, where $M$ is the sampling resolution and $c$ is the active ratio. Since the ratio of active regions is mostly less than 4%, our domain reduction reduces overall computations about $20\times$, achieving reconstruction time of about a minute. Moreover, as shown in Table 3, our domain reduction effectively reduces memory consumption, particularly when the target resolution is high. This enables our method to efficiently reconstruct hidden scenes using a single commercial GPU, typically requiring less than 2GB memory.

**Limitation and future works.** While our method presented the robustness to noise, we expect that such noise can be more effectively addressed. Reconstructing the objects with self-occlusion is also a worth exploring subject. We also expect interesting future research for extending the domain reduction to more complex scenes *e.g.*, scenes with multiple objects, and analyze its behavior.

## 6   Conclusion

This paper presented the new optimization-based method that can address various NLOS imaging scenarios. Thanks to the domain reduction strategy, conducted in a coarse-to-fine manner, our method can efficiently reconstruct hidden scenes in high-quality, less than a minute, significantly reducing acquisition time. We expect our method to facilitate broader range of real-world applications.

## Acknowledgements

## References

1. Ahn, B., Dave, A., Veeraraghavan, A., Gkioulekas, I., Sankaranarayanan, A.C.: Convolutional approximations to the general non-line-of-sight imaging operator. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7889–7899 (2019)
2. Aittala, M., Sharma, P., Murmann, L., Yedidia, A., Wornell, G., Freeman, B., Durand, F.: Computational mirrors: Blind inverse light transport by deep matrix factorization. Advances in Neural Information Processing Systems (NeurIPS) **32** (2019)
3. Arellano, V., Gutierrez, D., Jarabo, A.: Fast back-projection for non-line of sight reconstruction. Optics express **25**(10), 11574–11583 (2017)
4. Batarseh, M., Sukhov, S., Shen, Z., Gemar, H., Rezvani, R., Dogariu, A.: Passive sensing around the corner using spatial coherence. Nature communications **9**(1), 1–6 (2018)
5. Boger-Lombard, J., Katz, O.: Passive optical time-of-flight for non line-of-sight localization. Nature communications **10**(1), 1–9 (2019)
6. Bouman, K.L., Ye, V., Yedidia, A.B., Durand, F., Wornell, G.W., Torralba, A., Freeman, W.T.: Turning corners into cameras: Principles and methods. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2270–2278 (2017)
7. Chen, W., Wei, F., Kutulakos, K.N., Rusinkiewicz, S., Heide, F.: Learned feature embeddings for non-line-of-sight imaging and recognition. ACM Transactions on Graphics (TOG) **39**(6), 1–18 (2020)
8. Choi, K., Kim, I., Choi, D., Marco, J., Gutierrez, D., Kim, M.H.: Self-calibrating, fully differentiable nlos inverse rendering. In: SIGGRAPH Asia 2023 Conference Papers. pp. 1–11 (2023)
9. Chopite, J.G., Hullin, M.B., Wand, M., Iseringhausen, J.: Deep non-line-of-sight reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 960–969 (2020)
10. Galindo, M., Marco, J., O'Toole, M., Wetzstein, G., Gutierrez, D., Jarabo, A.: A dataset for benchmarking time-resolved non-line-of-sight imaging (2019), `https://graphics.unizar.es/nlos`
11. Grau, J., Plack, M., Haehn, P., Weinmann, M., Hullin, M.: Occlusion fields: An implicit representation for non-line-of-sight surface reconstruction. arXiv preprint arXiv:2203.08657 (2022)
12. Gu, C., Sultan, T., Masumnia-Bisheh, K., Waller, L., Velten, A.: Fast non-line-of-sight imaging with non-planar relay surfaces. In: 2023 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2023)

13. Heide, F., O'Toole, M., Zang, K., Lindell, D.B., Diamond, S., Wetzstein, G.: Non-line-of-sight imaging with partial occluders and surface normals. ACM Transactions on Graphics (TOG) **38**(3), 1–10 (2019)
14. Iseringhausen, J., Hullin, M.B.: Non-line-of-sight reconstruction using efficient transient rendering. ACM Transactions on Graphics (ToG) **39**(1), 1–14 (2020)
15. Isogawa, M., Chan, D., Yuan, Y., Kitani, K., O'Toole, M.: Efficient non-line-of-sight imaging from transient sinograms. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 193–208. Springer (2020)
16. Isogawa, M., Chan, D., Yuan, Y., Kitani, K., O'Toole, M.: Efficient non-line-of-sight imaging from transient sinograms. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 193–208. Springer (2020)
17. Isogawa, M., Yuan, Y., O'Toole, M., Kitani, K.M.: Optical non-line-of-sight physics-based 3d human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7013–7022 (2020)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Kirmani, A., Hutchison, T., Davis, J., Raskar, R.: Looking around the corner using transient imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 159–166 (2009)
20. La Manna, M., Nam, J.H., Reza, S.A., Velten, A.: Non-line-of-sight-imaging using dynamic relay surfaces. Optics express **28**(4), 5331–5339 (2020)
21. Lindell, D.B., Wetzstein, G., Koltun, V.: Acoustic non-line-of-sight imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6780–6789 (2019)
22. Lindell, D.B., Wetzstein, G., O'Toole, M.: Wave-based non-line-of-sight imaging using fast fk migration. ACM Transactions on Graphics (TOG) **38**(4), 1–13 (2019)
23. Liu, L., Gu, J., Lin, K.Z., Chua, T.S., Theobalt, C.: Neural sparse voxel fields. NeurIPS (2020)
24. Liu, X., Bauer, S., Velten, A.: Analysis of feature visibility in non-line-of-sight measurements. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10140–10148 (2019)
25. Liu, X., Bauer, S., Velten, A.: Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. Nature communications **11**(1), 1–13 (2020)
26. Liu, X., Guillén, I., La Manna, M., Nam, J.H., Reza, S.A., Huu Le, T., Jarabo, A., Gutierrez, D., Velten, A.: Non-line-of-sight imaging using phasor-field virtual wave optics. Nature **572**(7771), 620–623 (2019)
27. Liu, X., Wang, J., Xiao, L., Fu, X., Qiu, L., Shi, Z.: Few-shot non-line-of-sight imaging with signal-surface collaborative regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13303–13312 (2023)
28. Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D.: Nerf in the wild: Neural radiance fields for unconstrained photo collections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7210–7219 (2021)
29. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 405–421. Springer (2020)

30. Mu, F., Mo, S., Peng, J., Liu, X., Nam, J.H., Raghavan, S., Velten, A., Li, Y.: Physics to the rescue: Deep non-line-of-sight reconstruction for high-speed imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2022)
31. Nam, J.H., Brandt, E., Bauer, S., Liu, X., Renna, M., Tosi, A., Sifakis, E., Velten, A.: Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second. Nature communications **12**(1), 6526 (2021)
32. O'Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. Nature **555**(7696), 338–341 (2018)
33. Pei, C., Zhang, A., Deng, Y., Xu, F., Wu, J., David, U., Li, L., Qiao, H., Fang, L., Dai, Q.: Dynamic non-line-of-sight imaging system based on the optimization of point spread functions. Optics Express **29**(20), 32349–32364 (2021)
34. Plack, M., Callenberg, C., Schneider, M., Hullin, M.B.: Fast differentiable transient rendering for non-line-of-sight reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3067–3076 (2023)
35. Rapp, J., Saunders, C., Tachella, J., Murray-Bruce, J., Altmann, Y., Tourneret, J.Y., McLaughlin, S., Dawson, R.M., Wong, F.N., Goyal, V.K.: Seeing around corners with edge-resolved transient imaging. Nature communications **11**(1), 5929 (2020)
36. Saunders, C., Murray-Bruce, J., Goyal, V.K.: Computational periscopy with an ordinary digital camera. Nature **565**(7740), 472–475 (2019)
37. Seidel, S.W., Murray-Bruce, J., Ma, Y., Yu, C., Freeman, W.T., Goyal, V.K.: Two-dimensional non-line-of-sight scene estimation from a single edge occluder. IEEE Transactions on Computational Imaging **7**, 58–72 (2020)
38. Sharma, P., Aittala, M., Schechner, Y.Y., Torralba, A., Wornell, G.W., Freeman, W.T., Durand, F.: What you can learn by staring at a blank wall. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2330–2339 (2021)
39. Shen, S., Wang, Z., Liu, P., Pan, Z., Li, R., Gao, T., Li, S., Yu, J.: Non-line-of-sight imaging via neural transient fields. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **43**(7), 2257–2268 (2021)
40. Shi, X., Tang, M., Zhang, S., Qiao, K., Gao, X., Jin, C.: Passive localization and reconstruction of multiple non-line-of-sight objects in a scene with a large visible transmissive window. Optics Express **32**(6), 10104–10118 (2024)
41. Tanaka, K., Mukaigawa, Y., Kadambi, A.: Polarized non-line-of-sight imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2136–2145 (2020)
42. Tsai, C.Y., Kutulakos, K.N., Narasimhan, S.G., Sankaranarayanan, A.C.: The geometry of first-returning photons for non-line-of-sight imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7216–7224 (2017)
43. Tsai, C.Y., Sankaranarayanan, A.C., Gkioulekas, I.: Beyond volumetric albedo–a surface optimization framework for non-line-of-sight imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1545–1555 (2019)
44. Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M.G., Raskar, R.: Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. Nature communications **3**(1), 1–8 (2012)
45. Wang, J., Liu, X., Xiao, L., Shi, Z., Qiu, L., Fu, X.: Non-line-of-sight imaging with signal superresolution network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17420–17429 (2023)

46. Wu, C., Liu, J., Huang, X., Li, Z.P., Yu, C., Ye, J.T., Zhang, J., Zhang, Q., Dou, X., Goyal, V.K., et al.: Non–line-of-sight imaging over 1.43 km. Proceedings of the National Academy of Sciences **118**(10), e2024468118 (2021)
47. Xin, S., Nousias, S., Kutulakos, K.N., Sankaranarayanan, A.C., Narasimhan, S.G., Gkioulekas, I.: A theory of fermat paths for non-line-of-sight shape reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6800–6809 (2019)
48. Ye, J.T., Huang, X., Li, Z.P., Xu, F.: Compressed sensing for active non-line-of-sight imaging. Optics Express **29**(2), 1749–1763 (2021)
49. Yedidia, A.B., Baradad, M., Thrampoulidis, C., Freeman, W.T., Wornell, G.W.: Using unknown occluders to recover hidden scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12231–12239 (2019)
50. Young, S.I., Lindell, D.B., Girod, B., Taubman, D., Wetzstein, G.: Non-line-of-sight surface reconstruction using the directional light-cone transform. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1407–1416 (2020)
51. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: Plenoctrees for real-time rendering of neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5752–5761 (2021)
52. Zhu, D., Cai, W.: Fast non-line-of-sight imaging with two-step deep remapping. ACS Photonics **9**(6), 2046–2055 (2022)