Efficient Non-Line-of-Sight Imaging from Transient Sinograms

Supplementary Materials

Mariko Isogawa, Dorian Chan, Ye Yuan, Kris Kitani, and Matthew O'Toole

Contents

1. Overview of the Supplementary Materials

This supplementary document contains additional details on our methodology for C^2NLOS scanning. Please also refer to the supplementary video for additional information and results. We highlight reference numbers associated with the main paper in blue, and those associated with this supplementary document in red.

2. Prototype Confocal NLOS Imaging System

The design of our prototype confocal NLOS imaging system (shown in Figure 1) is based on the system proposed in O'Toole *et al*. [4]. Our laser is a low-power picosecond pulse diode from ALPHALAS with a wavelength of $520\,\mathrm{nm}$, a full width at half maximum (FWHM) of $60\,\mathrm{ps}$, and a peak power of $280\,\mathrm{mW}$. The laser emits pulses of light at a rate of $10\,\mathrm{MHz}$. A fast-gated single photon avalanche diode (SPAD) from Micro Photon Devices (MPD) measures the response; the gate feature of the SPAD is turned off in our experiments. Our time-correlated single photon counting system (TCSPC) is a PicoHarp 300 from PicoQuant, and its role is to convert the SPAD's output into a stream of photon events. A MATLAB script then bins these photon events into a transient sinogram. The laser and SPAD are aligned with a beamsplitter (Thorlabs PBS251), and a Nikon lens focuses the light from the scene onto the SPAD.

Another MATLAB script interfaces with a National Instruments Data Acquisition Device (NI-DAQ USB-6343) to control a pair of large beam galvo mirrors (Thorlabs GVS012). The mirrors control the point on a wall illuminated with laser light, and measured by the SPAD. After calibrating for the position of the wall relative to our setup, the galvo mirrors continuously scan a circle on the wall of a user-specified radius at a rate of $130\,\mathrm{Hz}$.

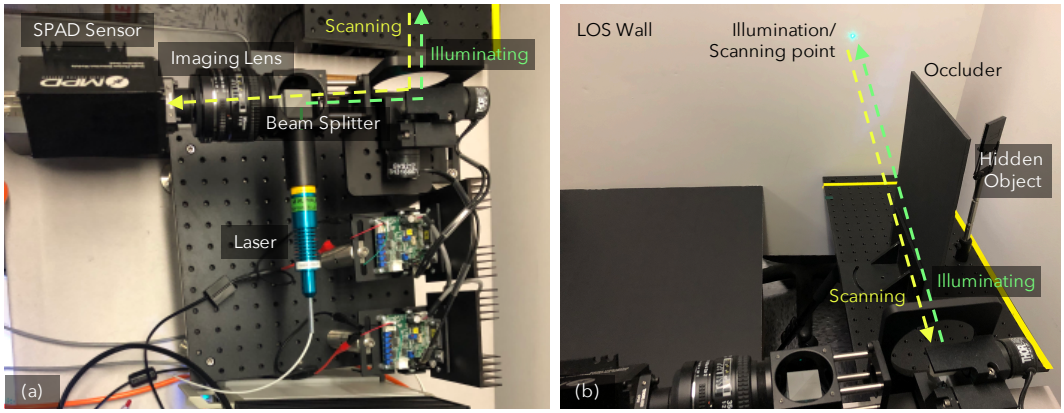


Figure 1. **(a)** Our hardware prototype. **(b)** A hidden scene with a single NLOS object (a retroreflector) used for our object localization experiments.

3. Additional Analysis and Results

3.1. 1D Reconstruction: Additional Object Localization Results with Real Captured Data

Figure 2 expands upon the object localization results shown in Figure 6 of the main paper. Here, we demonstrate the ability to estimate the positions of one, two, or three scatterers hidden from direct line-of-sight with our prototype NLOS system. Even in the case of three scatterers, our Hough voting approach accurately estimates the parameters (amplitude α, phase β, and offset γ) of the three corresponding sinusoids from the transient sinogram. We then convert the recovered parameters to each object's 3D position.

3.2. 2D Reconstruction

3.2.1 The Radon Transform and 2D NLOS Imaging

Let's consider the case of a single scatterer, located at some distance r_{gt} away from the center of the scanning circle (see Figure 3(a)). As shown in Figure 3(b), this scatterer will contribute a sinusoid to the transient measurements with the following form:

$$v(\phi') = r_{\text{gt}}^2 + r'^2 - 2r_{\text{gt}}r' \sin(\theta)\cos(\phi - \phi') = \gamma - \alpha\cos(\beta - \phi'), \tag{1}$$

where

$$\alpha = 2r_{\text{gt}}r' \sin(\theta), \quad \beta = \phi, \quad \gamma = r_{\text{gt}}^2 + r'^2. \tag{2}$$

Consider the case where we assume the correct distance is r_{est}, which may or may not equal r_{gt}. Following the methodology outlined in the paper, we recenter our transient sinogram at offset $r_{\text{est}}^2 + r'^2$. For simplicity, we ignore the effect of cropping for the following theoretical analysis. This formulation has a number of implications, which we describe below.

Now consider the case where $r_{\text{est}} = r_{\text{gt}}$. Then, the sinusoid of the scatterer is perfectly centered as input to the inverse Radon reconstruction. Because a single spatial point maps exactly to a perfectly centered sinusoid under the Radon Transform [2], our output 2D image captures a perfect scaled orthographic projection of the scatterer, located at coordinate $2r'[x, y]$ as given in Equation (13) in the main paper.

Now, suppose $r_{\text{est}} \neq r_{\text{gt}}$. The sinusoid of the scatterer will be instead shifted to be centered at $\Delta = r_{\text{gt}}^2 - r_{\text{est}}^2$ (see Figure 3(c)). Let $G_\beta(\omega)$ be the 1D Fourier Transform of the resampled transient measurement for a spatial circular sample β. Because every transient measurement is shifted by the same amount Δ, $G_\beta(\omega) = e^{-i\omega\Delta}F_\beta(\omega)$, where $F_\beta(\omega)$ is the 1D Fourier Transform of a correctly shifted measurement and $e^{-i\omega\Delta}$ is the same complex exponential for every β. By the projection-slice theorem, $G_\beta(u)$ exactly gives the 1D slice with angle β through the origin of the 2D Fourier Transform of the output image [2]. Therefore, the 2D Fourier Transform of this incorrect scene $G(u, v)$ will equal the 2D Fourier Transform of a correctly-shifted scene $F(u, v)$, modulated by the sinusoid given by $e^{-i\omega\Delta}$ rotated about the **0**-frequency origin. In the spatial domain, this circular sinusoidal pattern corresponds to a convolution with a circular kernel with radius Δ. In effect,

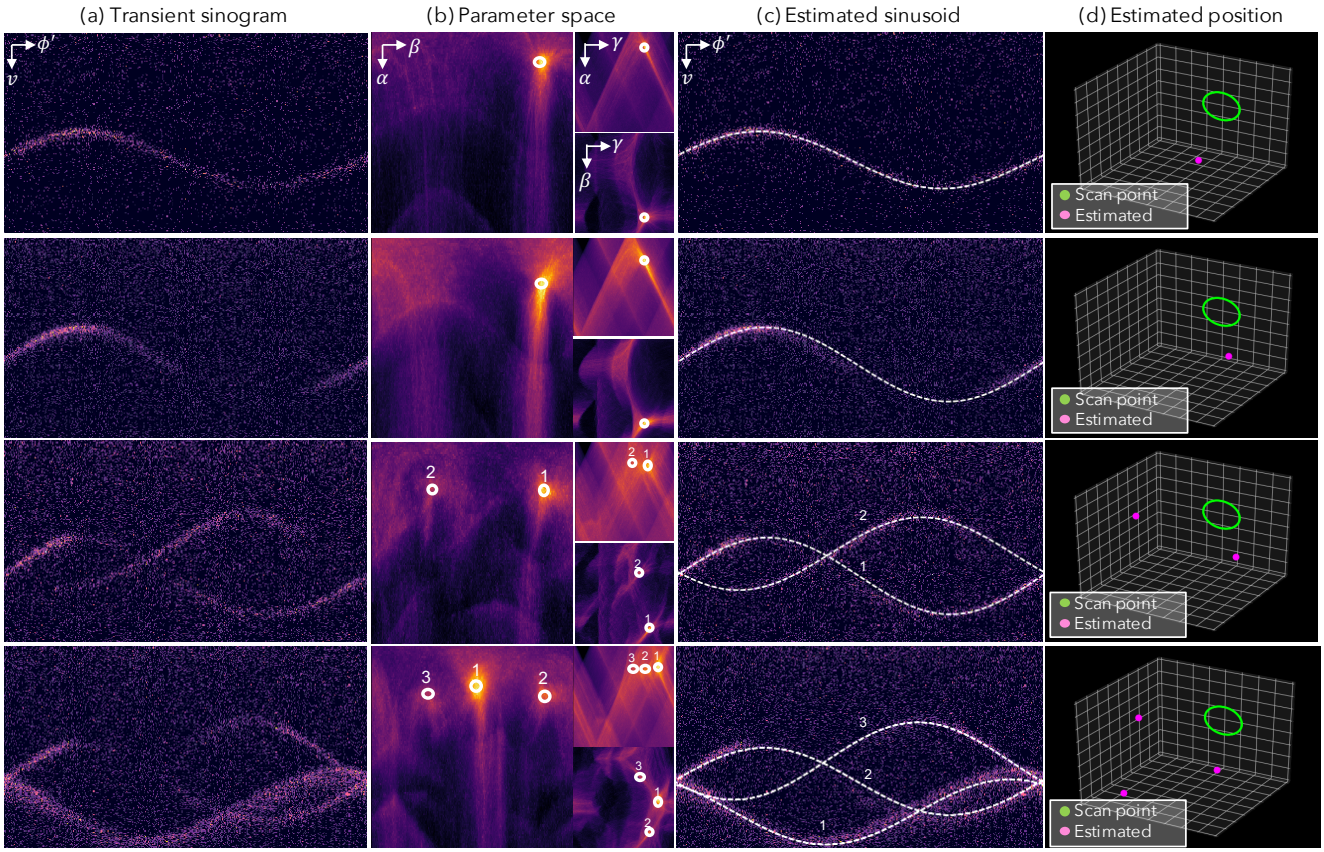


Figure 2. Additional experimental results for estimating the position of scatterers. **(a)** Transient sinograms of different numbers of scatterers. **(b)** From every transient sinogram, we estimate the amplitude, phase, and offset of the sinusoids using a Hough transform procedure. **(c)** By identifying peaks in the parameter space shown in (b), our approach recovers the sinusoid parameters that best fit the measured transient sinograms. **(d)** The 3D position of the object is finally reconstructed with the estimated parameters.

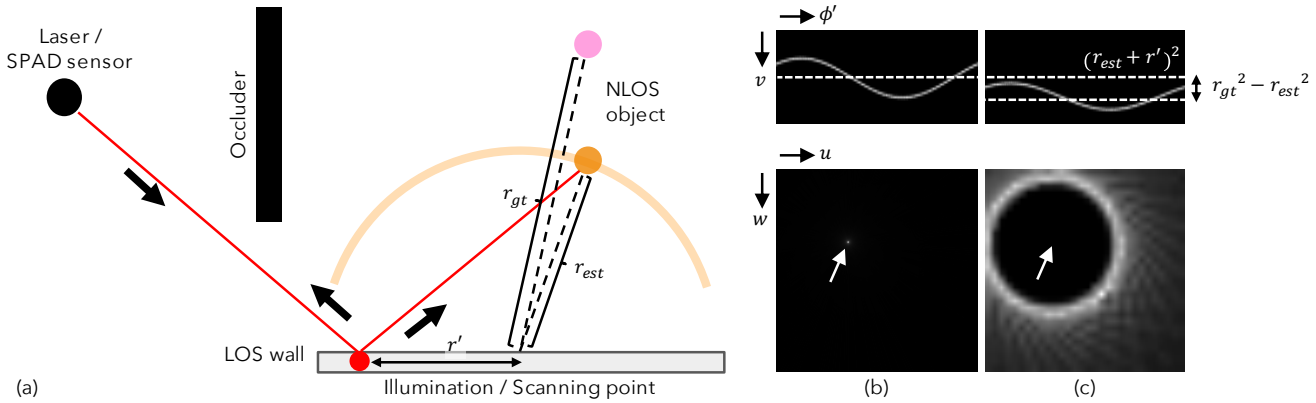


Figure 3. 2D imaging via the inverse Radon transform. **(a)** An infinitesimally small NLOS scatterer positioned at $[x, y, z]$ is a distance r_{gt} away from the center of the scanning circle with radius r'. **(b)** If the estimated distance r_{est} perfectly matches r_{gt} (*i.e.*, $r_{\text{est}} = r_{\text{gt}}$), the reconstruction contains a single point $2r'[x, y]$ representing the position of the scatterer. **(c)** Suppose the distance is incorrectly estimated (*i.e.*, $r_{\text{est}} \neq r_{\text{gt}}$). Then, the sinusoid corresponding to the scatterer will be shifted by $r_{\text{gt}}^2 - r_{\text{est}}^2$ in the transient sinogram. After applying the inverse Radon transform, the reconstruction is a circle centered at $2r'[x, y]$ with a radius proportional to the error in distance. We refer to this circle as the *Radon circle of confusion*.

rather than corresponding to a single point in the output 2D image from an inverse Radon Transform, the scatterer instead maps to a circle with radius Δ centered at $2r'[x, y]$. The radius Δ of this *Radon circle of confusion* does not change with the scanning circle radius r'.

Under this analysis, the 2D Radon NLOS image formation model for estimated distance r_{est} in the Fourier domain is given by

$$O_{r_{\text{est}}}(u, v) = 2r' \sum_d e^{-i\sqrt{u^2+v^2}(r_d^2 - r_{\text{est}}^2)} F_d(2r'u, 2r'v), \tag{3}$$

where $F_d(u, v)$ is the 2D Fourier Transform of the orthographic projection onto the wall of all elements of the scene r_d away from the center of the scanning circle. The $2r'$ term accounts for the scaling given by Equation (13).

This model has the following implications:

1. An object that perfectly lies upon the surface of a sphere with known radius yields a perfect reconstruction via an inverse Radon Transform, because every object point will be perfectly orthographically projected in the output image.

2. In the more general case, object points r_d away that do not satisfy the sphere constraint will generate circular patterns with radius $r_d^2 - r_{\text{est}}^2$ centered at their scaled orthographic projections.

We explore these effects in further detail in the rest of this section. In Section 3.2.2, we demonstrate the effectiveness of inverse Radon reconstruction for scenes that perfectly lie upon the surface of a sphere. In practice, because many common scenes do not satisfy this spherical constraint, we propose either empirical undistortion (Section 3.2.3) or larger scanning circles (Section 3.2.4) to remedy the artifacts that arise from a direct application of inverse Radon reconstruction. We also show manual refocusing via changing the estimated radius r_{est} in Section 3.2.5.

To implement our inverse Radon reconstruction, we use MATLAB's *iradon* functionality, which uses a backprojection operation. We discuss the empirical effects of filtered backprojection in 2D NLOS imaging in Section 3.2.6.

3.2.2 2D Image Reconstruction of Spherically-Constrained Scenes

Section 5.2 in the main paper shows that even though our inverse Radon reconstruction approach was not designed for large planar objects, it still recovers an approximate reconstruction of the 2D scene. With the above in mind, one might be interested in the quality of reconstruction when the **scene actually satisfies the sphere constraint** (*i.e.*, when the hidden object lies on the surface of a sphere of radius r with the same center as the scanning circle).

Figure 4 shows 2D reconstruction results for a simulated Z shape $1.0\,\text{m}$ from the wall, in which the sphere constraint is fully satisfied (Figure 4(b)) or not satisfied (Figure 4(c)). With a $1.0\,\text{m}$ scanning diameter, the spherically-constrained scene is much more accurately reconstructed compared to the planar version, as expected.

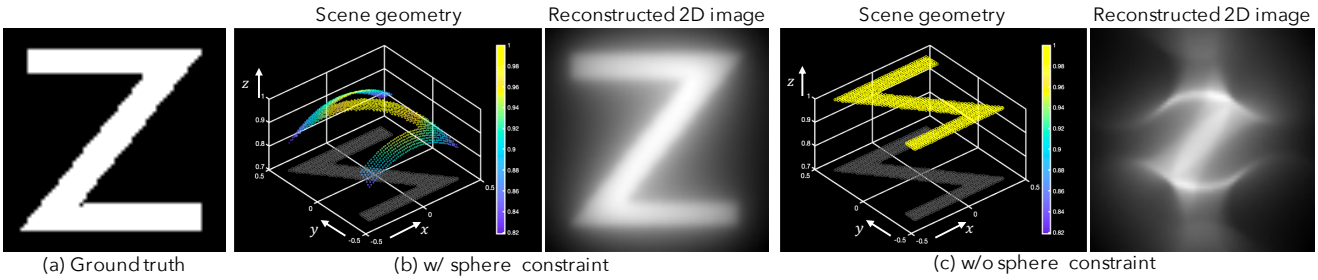


Figure 4. 2D imaging spherically-constrained scenes. **(a)** Ground truth projection. **(b)** If every point of the hidden scene lies within the surface of a sphere, our method reconstructs an accurate 2D image. **(c)** Even though our inverse Radon reconstruction approach was not designed for large planar objects that violate the sphere constraint, it still recovers an approximate reconstruction of the 2D scene.

3.2.3 2D Image Undistortion when the Scene is not Spherically-Constrained

As shown in Figure 5(a), when the NLOS object does not satisfy the sphere constraint, the inverse Radon transform produces a distorted version of the Z object. For every point on the Z object, the inverse Radon transform produces a circle with a radius proportional to the distance of that point from the surface of the sphere. This results in large circular distortions when

a planar object is too large relative to the size of the sphere, as shown in Figure 5(b). To compensate for this distortion, we empirically found that a simple fisheye lens undistortion operation and a cropping operation helps to produce a clearer image in such cases.

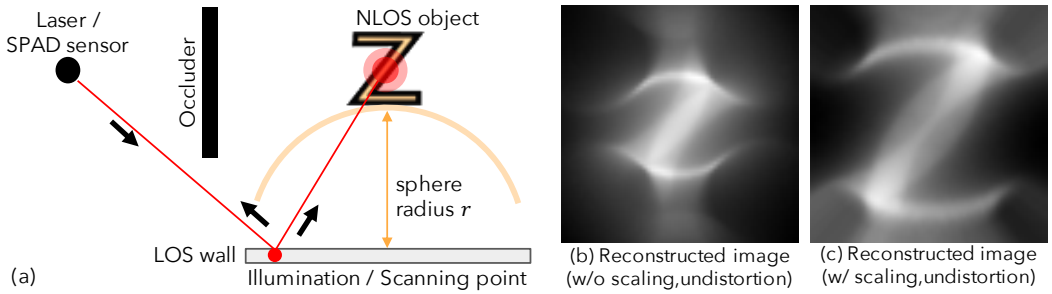


Figure 5. Undistorting 2D images of a scene that does not satisfy the sphere constraint. **(a)** Illustration of C^{2}NLOS scene geometry. **(b)** When the object does not perfectly lie on the surface of a sphere, the output of an inverse Radon reconstruction suffers from circular artifacts. **(c)** Empirically, we observe that applying a fisheye lens undistortion procedure improves the quality of the reconstruction.

3.2.4 Effect of the Scanning Circle's Radius

As shown in Section 3.2.1, the size of the *Radon circle of confusion* does not depend on the size of the scanning circle. However, because the rest of the image scales linearly with scanning radius r' from Equation (13), the effective size of the Radon circle of confusion in relation to the other features of the hidden scene therefore decreases with r'. Therefore, to minimize the effects of an incorrect r_{est}, we should maximize r'.

Figure 6 shows reconstructed 2D images with different circle scanning sizes (right shows results with larger scanning circles). Figures 6(i) to (iii) show reconstructed 2D images for a scene with a single infinitesimally small scatterer at $(x, y, z) = (0.5, 0.5, 1.0)[m]$, $(0.3, 0.3, 1.0)[m]$, and $(0.1, 0.1, 0.1)[m]$, respectively. Figure 6(iv) and (v) contain three scatterers at different depths. In Figure 6(iv), the scatterers share the same x, y location $(0.4, 0.4)$ but are positioned at different depths $z = 0.8$, 0.4, and 1.2. In Figure 6(iv), the scatterers are positioned at $(x, y, z) = (0.4, 0.4, 1.0)$, $(0.0, 0.0, 0.8)$, and $(-0.4, -0.4, 1.2)$. As shown in these figures, scatterers at different depths produce circles of confusion of different sizes, but centered at the scatterers' x, y locations. Using a larger scanning radius r' reduces the relative sizes of the circles of confusion.

Figure 6(vi) shows reconstruction results for a simulated Z shape $1.0\,\text{m}$ from the wall. **The larger the scanning circle, the smaller the effect of the circles of confusion, resulting in clearer images.** In practice, light falloff severely reduces the quality of the signal as r' increases. Thus, a C^{2}NLOS imaging system should aim to find the right balance between SNR and scanning circle size.

3.2.5 Synthetic Refocusing

As mentioned in Section 5.2 of the main paper, changing the value for the estimated sphere radius r can be used to focus on different parts of the hidden scene. Figure 7 shows refocusing with three scatterers at different depths, located at $(x, y, z) = (0.2, 0.2, 0.8)$, $(0.0, 0.0, 1.0)$, and $(-0.2, -0.2, 1.2)$. As shown in the figure, we can adjust the value of r to *"focus"* the image at a particular radius, which *"blurs"* points at other radii with a larger circle of confusion.

3.2.6 Frequency-Domain Filtering for Inverse Radon Reconstruction

An inverse Radon transform procedure often uses frequency-domain filtering to attenuate low-frequency components that are over-represented in the measurements. In the absence of noise with an object that perfectly satisfies the spherical constraint, a ramp filter (also known as a Ram-Lak filter) can perfectly reconstruct the hidden scene [2]; see Figure 8(c). However, as shown in Figure 8(e), we empirically find that these filters typically enhance the high-frequency circular artifacts generated by objects that do not satisfy the spherical constraint. As a result, we choose to use an unfiltered version of the inverse Radon transform for our 2D reconstructions, as demonstrated in Figure 8(d).

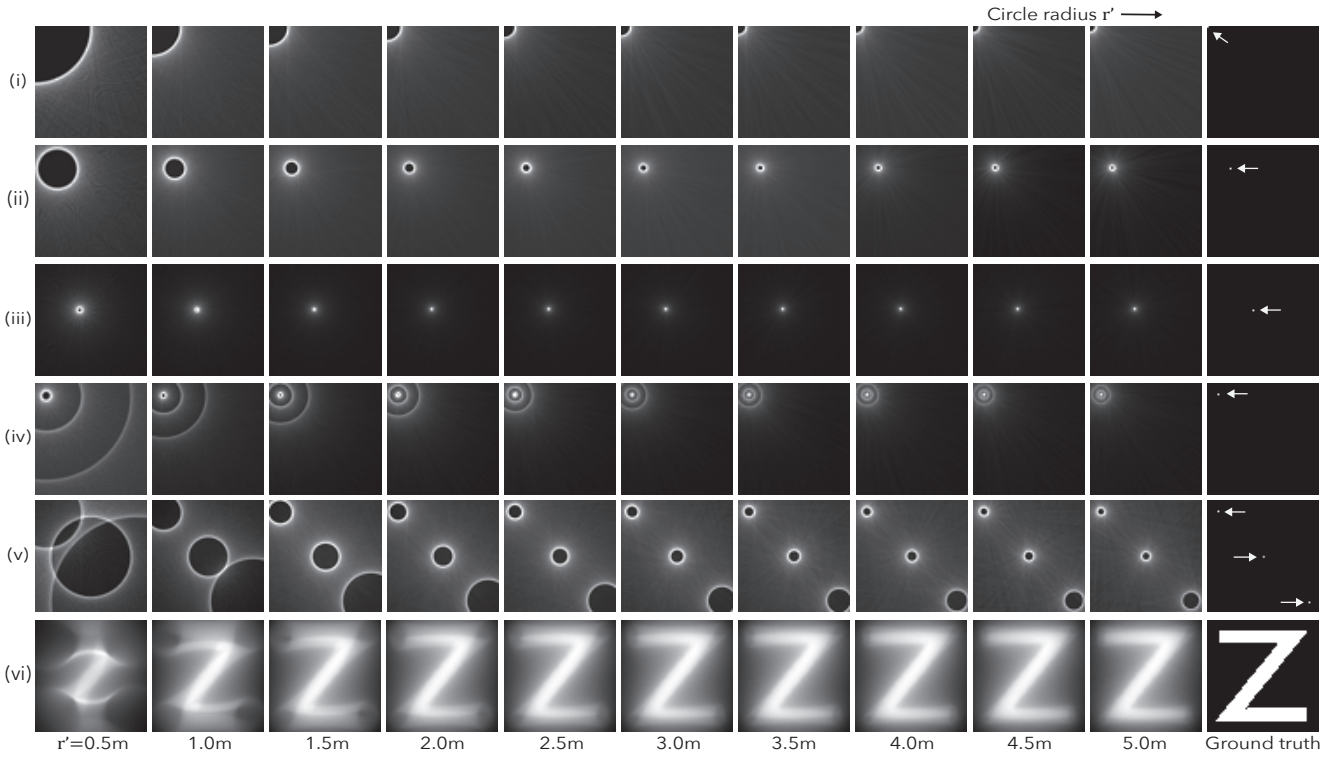


Figure 6. 2D NLOS imaging with different scanning circle sizes (right shows results with larger scanning circles). The effective size of the Radon circle of confusion decreases with larger scanning circle radius r', resulting in higher image quality.

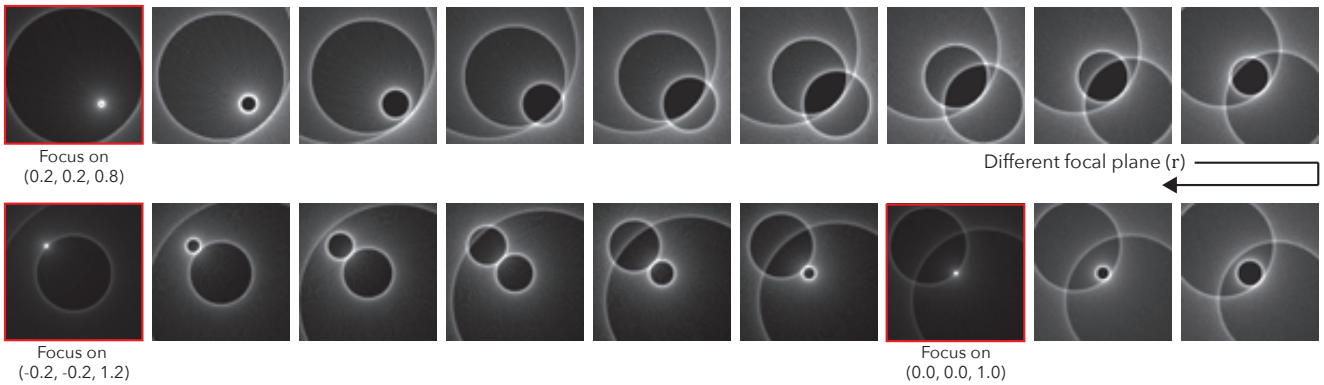


Figure 7. Manual refocusing in 2D NLOS imaging on a 3-scatterer scene. By changing the estimated radius r, we can manually refocus to one of the scatterers, while making the circles of confusion for the other scatterers larger.

3.2.7 Relationship between reconstruction quality and number of samples compared with 2D grid scanning

To investigate the impact on number of samples for the Radon reconstruction when objects satisfy the sphere constraint, we refer to [2] (end of Section 5.1.1), which states that "for a well-balanced N x N reconstructed image, [...] the total number of projections should also be roughly N", corresponding to N samples on the scanning circle. C^2NLOS requires far fewer samples when compared to the $N \times N$ grid required by LCT [4] and FK [3]. This also implies that more circle samples allow for higher resolution reconstructions; however, too many samples yield diminishing returns, because the transients have limited temporal resolution.

As we describe in the main paper, our C^2NLOS data consist of 360 samples and are used to reconstruct 2D images at a resolution of 360×360. However, for a budget of 360 samples, LCT and FK would be limited to 19×19 spatial samples, resulting in only a 19×19 reconstruction. For a fixed budget of 360 samples, we show a qualitative comparison in Figure 9

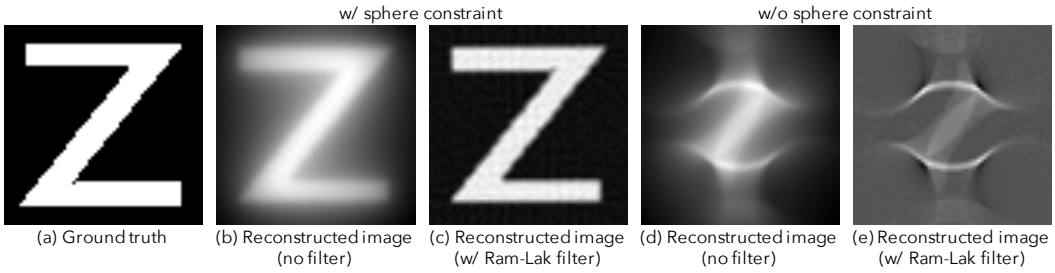


Figure 8. While traditional inverse Radon reconstructions rely on high pass filters, we avoid them in 2D NLOS imaging because such filters empirically enhance the effects of the Radon circle of confusion, making the artifacting much more visible. When the hidden object satisfies the sphere constraint, a Ram-Lak filter generates a high quality reconstruction as shown in **(c)**. However, when the object does not perfectly satisfy the spherical constraint as in **(d)** and **(e)**, we find that the non-filtered version empirically produces nicer results.

for 2D reconstruction and Figure 12 for 3D reconstruction. In both 2D and 3D, we believe C^{2}NLOS is comparable in quality to LCT and FK.

However, it is important to note that the number of samples is not the bottleneck. Rather, acquisition speeds are fundamentally limited by the scanning path, and how quickly the mirror galvanometers can follow this path. A single row of a coarse grid requires the same capture time as an entire circular scan; therefore, sampling a 19×19 grid is still 19 times slower than a C^{2}NLOS scan.

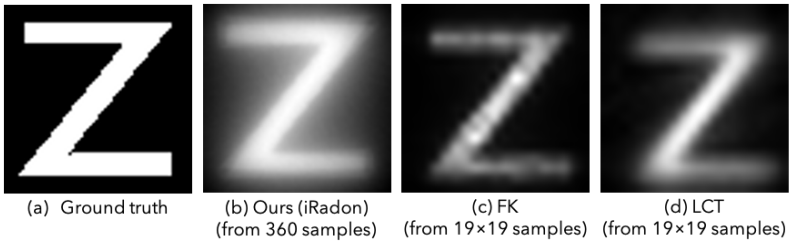


Figure 9. Qualitative comparison of 2D reconstruction quality for a fixed budget of 360 samples. For FK/LCT, we perform reconstructions from a 19×19 grid of samples (361 total samples). In this scenario, while both scanning patterns use the same number of samples, a C^{2}NLOS pattern can be acquired 19 times faster.

4. Reconstruction Procedure Details

This section explains various implementation details for the algorithms described in the paper. Section 4.1 explains Hough voting for object localization in further detail. Section 4.2 describes the 3 Points algorithm [4], a baseline for single object localization used in Section 5.1. In Section 4.3, we show a full derivation of the proximal operators used for our 3D volume imaging described in Section 5.3 in the main submission.

4.1. Sinusoid Parameter Estimation based on the Hough Transform

As mentioned in Section 4.1 of the main paper, our method estimates the best matching amplitude α, phase β, and offset γ for every sinusoid in a transient sinogram. This section aims to provide more detail for this procedure.

The key challenge is the presence of noise in a measured transient sinogram as shown in Figure 10(a). Similar to [5], we perform sinusoid fitting by using the Hough transform, which is commonly used as a robust parameter estimation approach. The Hough transform relies on a voting procedure to estimate the most likely parameters, which tends to be computationally expensive when there are many parameters to estimate. However, we can speed up the procedure in two ways. First, we can use a single sinusoid with a fixed temporal offset and phase shift as a template image for each amplitude (see Figure 10(b)). Thus, we only need to prepare templates for the number of candidate amplitudes, which greatly reduces the computational cost. Second, we can also perform fast Hough transforms through a convolution, which can be efficiently computed in the Fourier domain. Here, we can define our sinusoidal 2D image template as

$$T_\alpha(\theta, v) = \begin{cases} 1 & \text{if } v = \alpha \cos(\theta) + N/2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where N represents temporal resolution of T_α. Convolving this 2D image template with the transient sinogram produces a slice of 2D parameter space image A_α:

$$A_\alpha = \mathcal{F}^{-1}(\mathcal{F}(T_\alpha(\theta, v)) * \mathcal{F}(\tau(\theta, v))) \tag{5}$$

where $\mathcal{F}, \mathcal{F}^{-1}$ are the Fourier and inverse Fourier transform operations. Computing A_α for each amplitude α produces a three dimensional parameter space volume $A_{\{0:N/2\}}$. The element with the highest value indicates the sinusoidal parameters that best represents the sinusoid in the input transient.

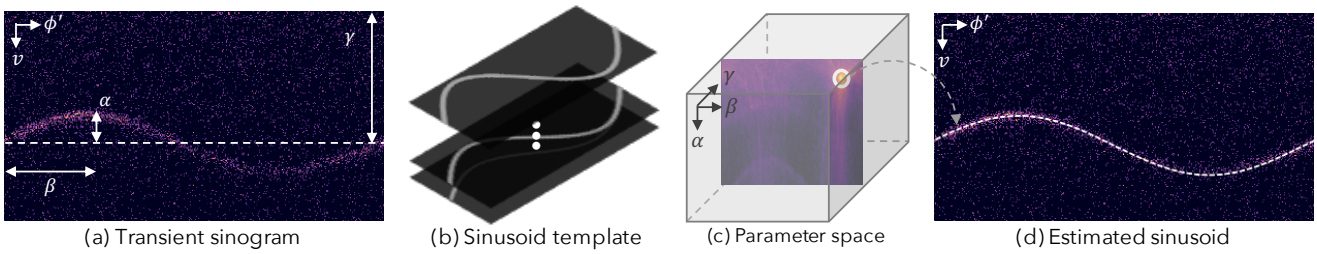


Figure 10. Our Hough transform-based object localization. Given a transient sinogram **(a)**, we generate a 3D parameter space **(c)** by convolving a sinusoid template **(b)** with the input transient sinogram. The peak of the parameter space represents the sinusoid parameters that best fit the input sinogram **(d)**.

4.2. Object Localization with 3 Point Scanning

We used 3 scanning point trilateration [4] as a baseline in Section 5.1 for single object localization. We describe this procedure in further detail in this section. As illustrated in Figure 11, let x denote the position of the NLOS object, and x'_1, x'_2, and x'_3 the positions of each scanning point. Assuming that there is only a single object in the hidden scene, the temporal peaks of the transient measurements for each scanning point t_1, t_2, t_3 directly give the distance between the scanning points and the hidden object. These distances can be calculated from the temporal peaks as $r_1 = t_1 c/2$, $r_2 = t_2 c/2$, $r_3 = t_3 c/2$, where c denotes the speed of light. Since the object exists at the intersection of the 3 spheres centered at x'_1, x'_2, and x'_3 with radius r_1, r_2, and r_3 respectively, the object position x is obtained by solving the three following simultaneous equations:

$$(x'_1 - x)^2 = r_1^2 \tag{6}$$

$$(x'_2 - x)^2 = r_2^2 \tag{7}$$

$$(x'_3 - x)^2 = r_3^2 \tag{8}$$

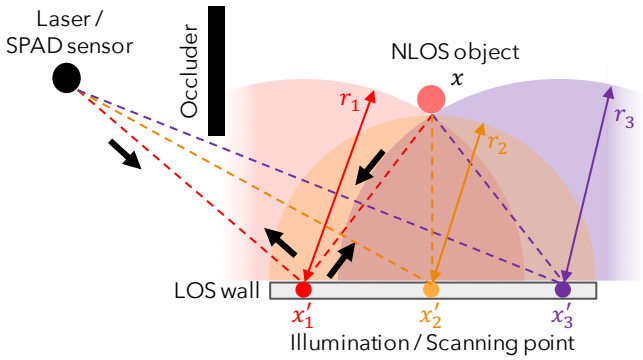


Figure 11. Object localization with just 3 scanning points [4]

4.3. 3D Volume Imaging via ADMM

As mentioned in the main paper, we recover a full 3D volume reconstruction $\boldsymbol{\rho}$ of the hidden scene from C^{2}NLOS measurements $\boldsymbol{\tau}_{\text{circ}}$, by using a modified version of the iterative light cone transform (LCT) procedure used in confocal NLOS imaging [4]. We decide to use non-negativity, sparsity, and total variation priors to compensate for the very sparsely sampled input data. With these priors, the optimization problem can be written as:

$$\min_{\boldsymbol{\rho}} \quad \frac{1}{2}\left\|\boldsymbol{\tau}_{\text{circ}} - \mathbf{M}\mathbf{A}\boldsymbol{\rho}\right\|_2^2 + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\rho}) + \lambda_s\left\|\boldsymbol{\rho}\right\|_1 + \lambda_{TV}\left(\left\|\mathbf{D}_x\boldsymbol{\rho}\right\|_1 + \left\|\mathbf{D}_y\boldsymbol{\rho}\right\|_1 + \left\|\mathbf{D}_z\boldsymbol{\rho}\right\|_1\right) \tag{9}$$

For simplicity, we modify the above optimization to operate in the light cone domain [4]:

$$\min_{\boldsymbol{\rho}_u} \quad \frac{1}{2}\left\|\boldsymbol{\tau}_{\text{circ}} - \mathbf{M}\mathbf{H}\boldsymbol{\rho}_u\right\|_2^2 + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\rho}_u) + \lambda_s\left\|\boldsymbol{\rho}_u\right\|_1 + \lambda_{TV}\left(\left\|\mathbf{D}_x\boldsymbol{\rho}_u\right\|_1 + \left\|\mathbf{D}_y\boldsymbol{\rho}_u\right\|_1 + \left\|\mathbf{D}_z\boldsymbol{\rho}_u\right\|_1\right) \tag{10}$$

where $\mathbf{H}$ is a convolutional matrix, $\boldsymbol{\rho}_u$ is a resampled version of $\boldsymbol{\rho}$ following the LCT procedure, and $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$ implement finite difference operators along the x, y, and z directions respectively. For notational simplicity, let $\boldsymbol{\rho} = \boldsymbol{\rho}_u$ and $\boldsymbol{\tau} = \boldsymbol{\tau}_{\text{circ}}$. In order to apply ADMM, we can rewrite the above equation:

$$\min_{\boldsymbol{\rho}} \quad \underbrace{\frac{1}{2}\left\|\boldsymbol{\tau} - \mathbf{z}_1\right\|_2^2}_{g_1(\mathbf{z}_1)} + \underbrace{\mathcal{I}_{\mathbb{R}_+}(\mathbf{z}_2)}_{g_2(\mathbf{z}_2)} + \underbrace{\lambda_s\left\|\mathbf{z}_3\right\|_1}_{g_3(\mathbf{z}_3)} + \underbrace{\lambda_{TV}\left\|\mathbf{z}_4\right\|_1}_{g_4(\mathbf{z}_4)} + \underbrace{\lambda_{TV}\left\|\mathbf{z}_5\right\|_1}_{g_5(\mathbf{z}_5)} + \underbrace{\lambda_{TV}\left\|\mathbf{z}_6\right\|_1}_{g_6(\mathbf{z}_6)} \tag{11}$$

$$\text{s.t.} \quad \underbrace{\begin{bmatrix}\mathbf{M}\mathbf{H}\\ \mathbf{I}\\ \mathbf{I}\\ \mathbf{D}_x\\ \mathbf{D}_y\\ \mathbf{D}_z\end{bmatrix}}_{\mathbf{C}}\boldsymbol{\rho} - \underbrace{\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\\ \mathbf{z}_4\\ \mathbf{z}_5\\ \mathbf{z}_6\end{bmatrix}}_{\mathbf{z}} = \mathbf{0} \tag{12}$$

The Augmented Lagrangian for this objective function can be written as:

$$L_\mu(\boldsymbol{\rho}, \mathbf{z}, \mathbf{y}) = \sum_{i=1}^{6} g_i(\mathbf{z}_i) + \mathbf{y}^T(\mathbf{C}\boldsymbol{\rho} - \mathbf{z}) + \frac{\mu}{2}\left\|\mathbf{C}\boldsymbol{\rho} - \mathbf{z}\right\|_2^2 \tag{13}$$

We operate on the scaled form, with $\mathbf{u} = \mathbf{y}/\mu$:

$$L_\mu(\boldsymbol{\rho}, \mathbf{z}, \mathbf{u}) = \sum_{i=1}^{6} g_i(\mathbf{z}_i) + \frac{\mu}{2}\left\|\mathbf{C}\boldsymbol{\rho} - \mathbf{z} + \mathbf{u}\right\|_2^2 - \frac{\mu}{2}\left\|\mathbf{u}\right\|_2^2 \tag{14}$$

We are now ready to write out the proximal operator update rules:

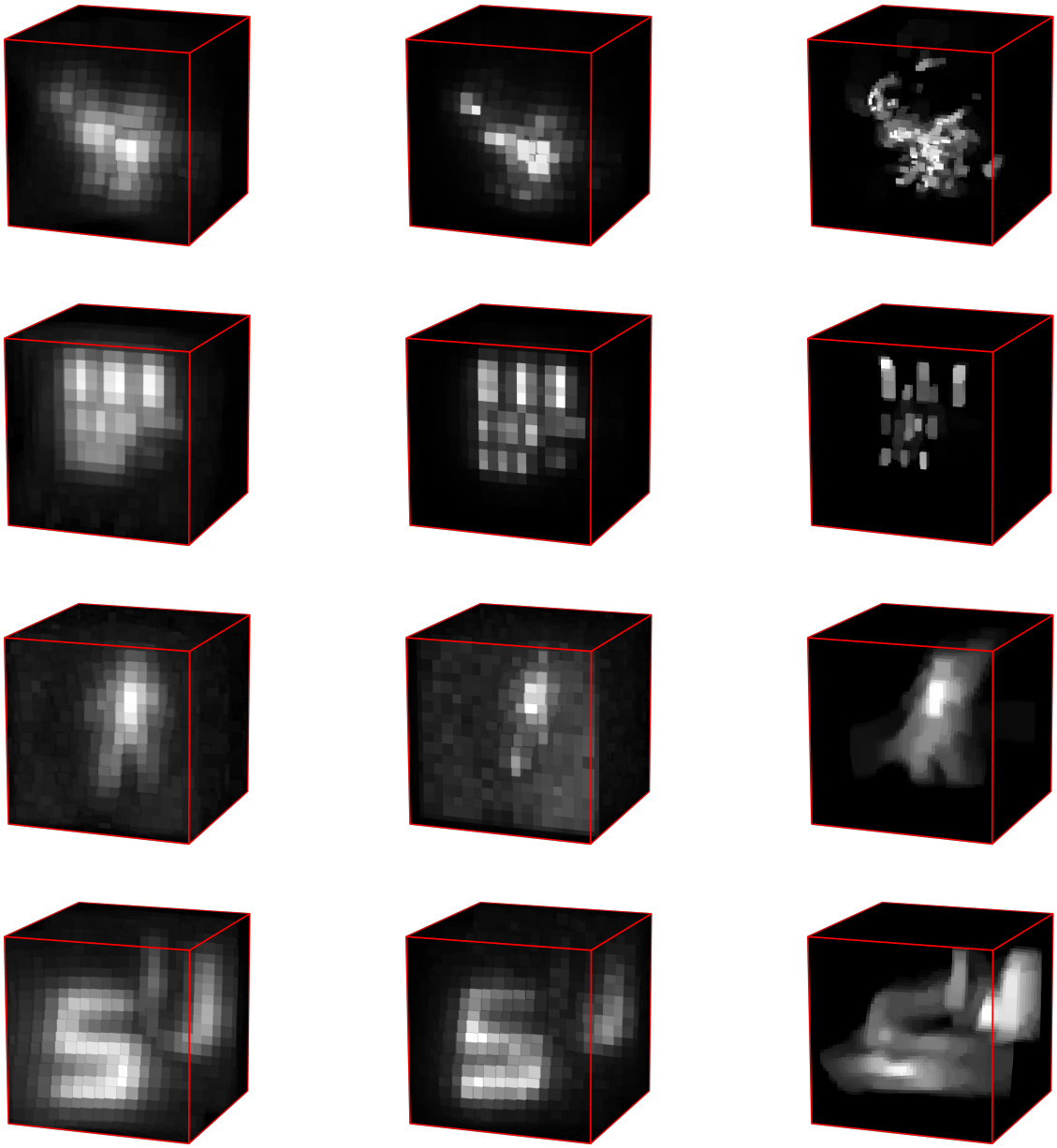
$$
\begin{aligned}
\mathbf{z}_1 &\leftarrow \arg\min_{\mathbf{z}_1} g_1(\mathbf{z}_1) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_1\right\|_2^2, \quad \mathbf{v} = \mathbf{M}\mathbf{H}\boldsymbol{\rho}_u + \mathbf{u}_1 \\
&= \arg\min_{\mathbf{z}_1} \frac{1}{2}\left\|\boldsymbol{\tau}-\mathbf{z}_1\right\|_2^2 + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_1\right\|_2^2 \\
&= \frac{\boldsymbol{\tau}+\mu\mathbf{v}}{1+\mu} \\
\mathbf{z}_2 &\leftarrow \arg\min_{\mathbf{z}_2} g_2(\mathbf{z}_2) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_2\right\|_2^2, \quad \mathbf{v} = \boldsymbol{\rho} + \mathbf{u}_2 \\
&= \arg\min_{\mathbf{z}_2} \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{z}_2) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_2\right\|_2^2 \\
&= \max(0, \mathbf{v}) \\
\mathbf{z}_3 &\leftarrow \arg\min_{\mathbf{z}_3} g_3(\mathbf{z}_3) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_3\right\|_2^2, \quad \mathbf{v} = \boldsymbol{\rho} + \mathbf{u}_3 \\
&= \arg\min_{\mathbf{z}_3} \lambda_s \left\|\boldsymbol{z}_3\right\|_1 + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_3\right\|_2^2 \\
&= S_{\lambda_s/\mu}(\mathbf{v}) \\
\mathbf{z}_4 &\leftarrow \arg\min_{\mathbf{z}_4} g_4(\mathbf{z}_4) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_4\right\|_2^2, \quad \mathbf{v} = \mathbf{D}_x\boldsymbol{\rho} + \mathbf{u}_4 \\
&= \arg\min_{\mathbf{z}_4} \lambda_{TV} \left\|\boldsymbol{z}_4\right\|_1 + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_4\right\|_2^2 \\
&= S_{\lambda_{TV}/\mu}(\mathbf{v}) \\
\mathbf{z}_5 &\leftarrow \arg\min_{\mathbf{z}_5} g_5(\mathbf{z}_5) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_5\right\|_2^2, \quad \mathbf{v} = \mathbf{D}_y\boldsymbol{\rho} + \mathbf{u}_5 \\
&= \arg\min_{\mathbf{z}_5} \lambda_{TV} \left\|\boldsymbol{z}_5\right\|_1 + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_5\right\|_2^2 \\
&= S_{\lambda_{TV}/\mu}(\mathbf{v}) \\
\mathbf{z}_6 &\leftarrow \arg\min_{\mathbf{z}_6} g_6(\mathbf{z}_6) + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_6\right\|_2^2, \quad \mathbf{v} = \mathbf{D}_z\boldsymbol{\rho} + \mathbf{u}_6 \\
&= \arg\min_{\mathbf{z}_6} \lambda_{TV} \left\|\boldsymbol{z}_6\right\|_1 + \frac{\mu}{2}\left\|\mathbf{v}-\mathbf{z}_6\right\|_2^2 \\
&= S_{\lambda_{TV}/\mu}(\mathbf{v}) \\
\mathbf{u} &\leftarrow \mathbf{u} + \mathbf{C}\boldsymbol{\rho} - \mathbf{z} \\
\boldsymbol{\rho} &\leftarrow \arg\min_{\boldsymbol{\rho}} \frac{1}{2}\left\|\mathbf{C}\boldsymbol{\rho}-\mathbf{v}\right\|_2^2, \quad \mathbf{v} = \mathbf{z} - \mathbf{u}
\end{aligned}
$$

It is difficult to solve the proximal operator for $\boldsymbol{\rho}$ in closed form with the lossy mapping term $\mathbf{M}$. Instead, we opt for a linearized ADMM approach. Our update rule for $\boldsymbol{\rho}$ now looks like the following:

$$
\boldsymbol{\rho} \leftarrow \boldsymbol{\rho} - \frac{\mu}{\nu}\mathbf{C}^*(\mathbf{C}\boldsymbol{\rho} - \mathbf{v}) \tag{15}
$$

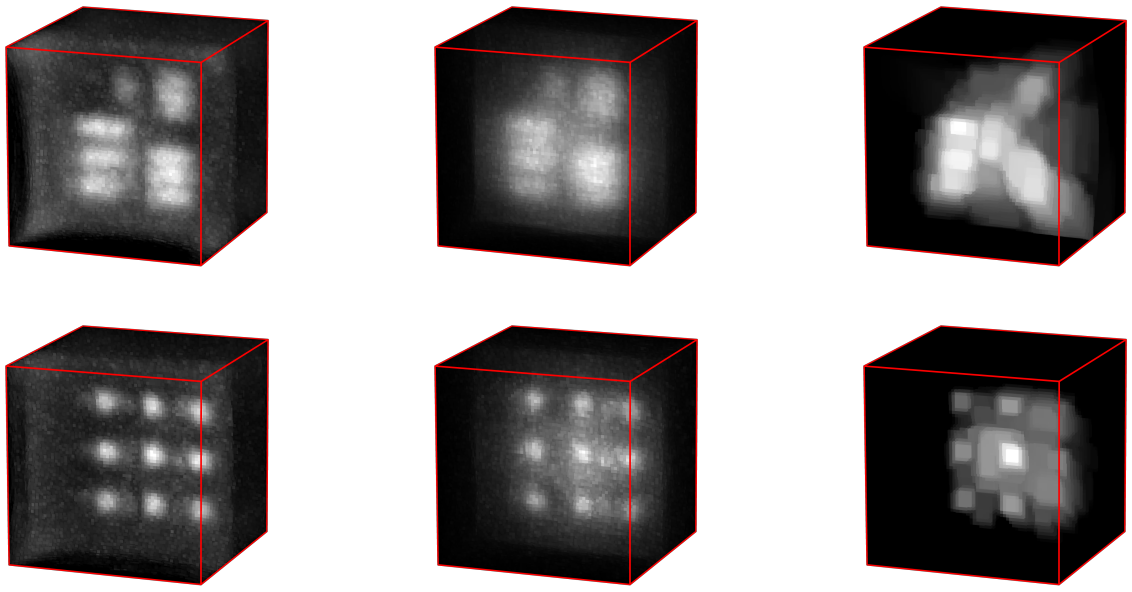
where ν controls the learning rate. Under this formulation, all of the above proximal operators can be efficiently solved, because $\mathbf{H}$, $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$ can all be expressed as elementwise multiplications in the Fourier domain.

We show additional 3D reconstruction results in Fig. 13 and Fig. 14. In general, a C^2NLOS scan is sufficient for recovering the important shape of the hidden scene, as shown by our reconstructions. However, especially in the case of simulated data, not enough measurements are provided to resolve possible ambiguities in the voxel volume, resulting in streaking artifacts that degrade the output quality.



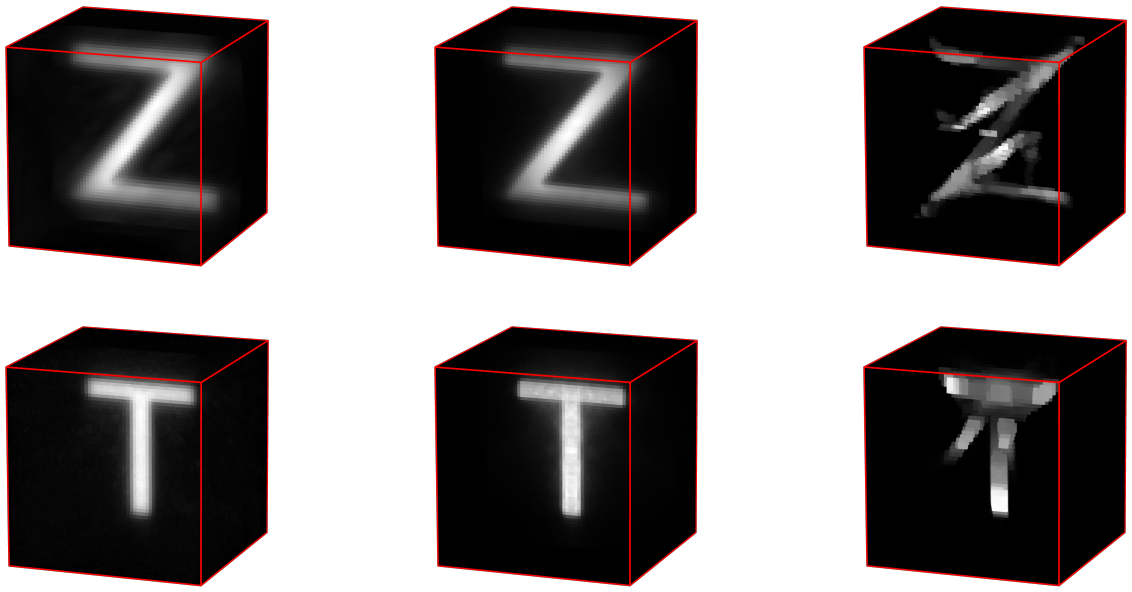
(a) LCT (b) FK (c) C^{2}NLOS

Figure 12. Qualitative comparison of 3D reconstruction quality for a fixed budget of 360 samples. For FK/LCT, we used a 19×19 grid. In this scenario, while both scanning patterns use the same number of samples, a C^{2}NLOS pattern can be acquired 19 times faster.



(a) LCT (b) FK (c) C^2NLOS

Figure 13. 3D reconstruction examples on real data from O'Toole *et al*. [4]. Our circular scans contain significant information about the hidden scene, capturing the important features of the hidden scene.



(a) LCT (b) FK (c) C^2NLOS

Figure 14. 3D reconstruction examples on simulated data from the Z-NLOS Dataset [1]. Our circular scans contain significant information about the hidden scene, but can potentially suffer from artifacting thanks to ambiguity in the measurements.

References

[1] Galindo, M., Marco, J., O'Toole, M., Wetzstein, G., Gutierrez, D., Jarabo, A.: A dataset for benchmarking time-resolved non-line-of-sight imaging. In: ACM SIGGRAPH 2019 Posters, pp. 1–2 (2019) 12

[2] Kak, A.C., Slaney, M., Wang, G.: Principles of computerized tomographic imaging. Medical Physics **29**(1), 107–107 (2002) 2, 5, 6

[3] Lindell, D.B., Wetzstein, G., O'Toole, M.: Wave-based non-line-of-sight imaging using fast f-k migration. ACM Transactions on Graphics (TOG) **38**(4), 116 (2019) 6

[4] O'Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. Nature **555**(7696), 338 (2018) 1, 6, 7, 8, 9, 12

[5] Zou, C.C., Ge, S.: A Hough transform-based method for fast detection of fixed period sinusoidal curves in images. In: International Conference on Signal Processing (ICSP). pp. 909 – 912 vol.1 (2002) 7