Role of Transients in Two-Bounce Non-Line-of-Sight Imaging

Siddharth Somasundaram[1]   Akshat Dave[2]   Connor Henley[1]
Ashok Veeraraghavan[2]   Ramesh Raskar[1]
[1]Massachusetts Institute of Technology   [2]Rice University
{sidsoma,co24401,raskar}@mit.edu, {akshat,vashok}@rice.edu

(a) Two-Bounce NLOS   **(b) Application Settings**   **(c) Experimental Setup**   **(d) Reconstruction with 10 Measurements**
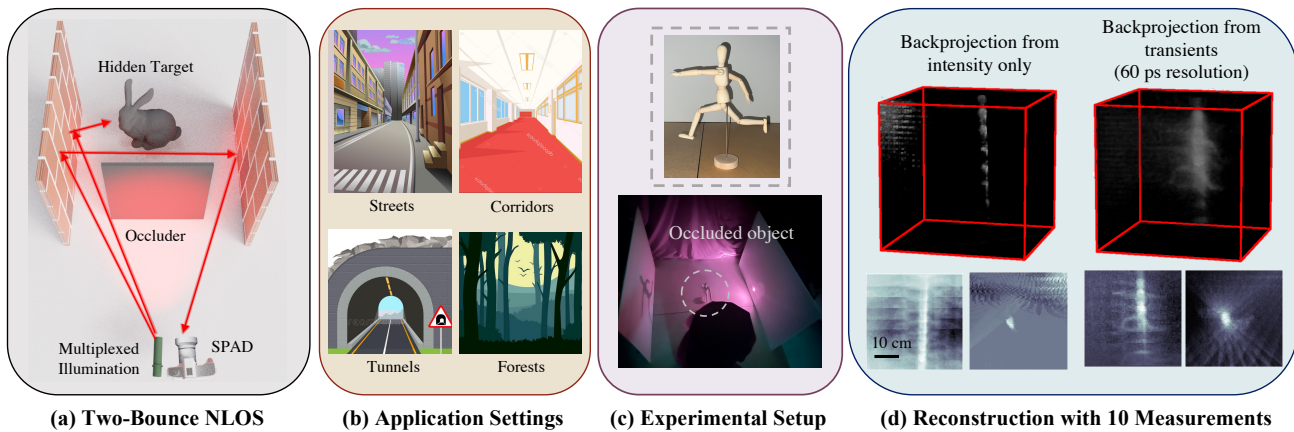
Figure 1. **Two-bounce Transient NLOS Imaging.** (a) We propose using two-bounce transients for NLOS imaging. Two-bounce signals capture information about the shadows in a scene, and use of transient information can reduce the number of measurements needed for reconstruction by using multiplexed illumination. (b) Our work can be applicable in many scenarios, such as tunnels, corridors, streets, and forests. (c)-(d) We experimentally validate the use of two-bounce transients on real objects. We show that time-of-flight information enables more robust reconstructions of occluded objects while using fewer measurements.

Abstract

The goal of non-line-of-sight (NLOS) imaging is to image objects occluded from the camera's field of view using multiply scattered light. Recent works have demonstrated the feasibility of two-bounce (2B) NLOS imaging by scanning a laser and measuring cast shadows of occluded objects in scenes with two relay surfaces. In this work, we study the role of time-of-flight (ToF) measurements, i.e. *transients, in 2B-NLOS under multiplexed illumination. Specifically, we study how ToF information can reduce the number of measurements and spatial resolution needed for shape reconstruction. We present our findings with respect to trade-offs in (1) temporal resolution, (2) spatial resolution, and (3) number of image captures by studying SNR and recoverability as functions of system parameters. This leads to a formal definition of the mathematical constraints for 2B lidar. We believe that our work lays an analytical groundwork for design of future NLOS imaging systems, especially as ToF sensors become increasingly ubiquitous.*

1. Introduction

Non-line-of-sight (NLOS) imaging aims to reconstruct objects occluded from direct line of sight and has the potential to be transformative in numerous applications across autonomous driving, search and rescue, and non-invasive medical imaging [24]. The key approach is to measure light that has undergone multiple surface scattering events and computationally invert these measurements to estimate hidden geometries. Recent work used two-bounce light, as shown in Fig. 2a, to reconstruct high quality shapes behind occluders [15]. The key idea is that two-bounce light captures information about the shadows of the occluded object. By scanning the laser source at different points l on a relay surface and measuring multiple shadow images, it is possible to reconstruct the hidden object by computing the visual hull [20] of the measured shadows.

Benefits of Two-Bounce   *Two-bounce* (2B) light can be captured using two relay surfaces on opposite sides of the hidden scene (Fig. 1a). This can occur in a variety of real-world settings such as tunnels, hallways, streets, and clut-
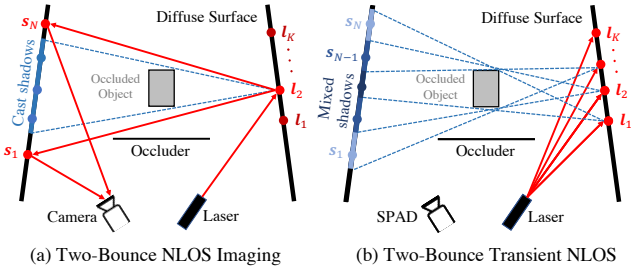
Figure 2. **Two-Bounce NLOS Imaging.** (a) Two-bounce NLOS imaging performs 3D reconstruction by using information contained in the shadows of the hidden object when illuminated by one virtual source l at a time. (b) In this work, we study the benefits of using transient information for two-bounce NLOS imaging under the presence of multiplexed illumination.

tered environments like forests (Fig. 1b). For example, imagine an autonomous vehicle in a tunnel being able to see ahead of the car in front of it by using two-bounce signals from the sides of the tunnel. Two-bounce light also helps alleviate the signal-to-noise ratio (SNR) issue in 3B-NLOS imaging, since 3B signals experience significant signal attenuation at each surface scatter event. Furthermore, 2B light captures information about the *cast shadows* of the hidden object instead of reflectance, as in the 3B case. As a result, the SNR of 2B signals is independent of the object's albedo and can therefore robustly image darker objects.

Contributions In this work, we propose using two-bounce transients for NLOS imaging to enable the use of multiplexed illumination. 3B-NLOS transient imaging systems have been analyzed previously, but we believe we are the first to analyze the use of multiplexed illumination and transient measurements for 2B-NLOS (Fig. 2b). Our key insight is that multiplexed illumination enables measurement of occluded objects with fewer image captures, and 2B transients enable demultiplexing shadows, as shown in Fig. 3. The robust SNR and promise of few-shot capture afforded by 2B transients makes the idea an important direction for few-shot NLOS imaging in general environments. We summarize our contributions below.

- *Model*: We present new insights into the algebraic formulations of two-bounce lidar for NLOS imaging.

- *Analysis*: We analyze tradeoffs with spatial resolution, temporal resolution, and number of image captures by using quantitative metrics such as SNR and recoverability, as well as qualitative validation from real and simulated results.

Although we envision the ideas introduced here to inspire future works in single-shot NLOS imaging, we do *not* claim this as a contribution in this paper.

Scope of This Work All data used for this work is from simulation and a single-pixel SPAD sensor. We do not physically realize few-shot results in this paper due to limited availability of high-resolution SPAD arrays presently. Instead, we emulate SPAD array measurements by scanning a single-pixel SPAD across the field of view and and emulate multiplexed illumination by scanning a laser spot, acquiring individual transient images, and summing the images in post-processing. These measurements, however, are more conducive to our goal of performing analysis that explores the landscape of possibilities in NLOS imaging with advances in ToF sensors. We believe that the ideas and analysis introduced in this work will be increasingly relevant as SPAD array technology matures [19, 26, 44].

2. Related Work

Non-Line-of-Sight Imaging

ToF Methods: The *non-line-of-sight imaging* (NLOS) problem aims to see around a corner by measuring light reflected off intermediary surfaces. A large body of work models the response of a hidden scene to a temporal pulse of light and generally perform a variant of ellipsoidal tomography or solve a linear inverse problem [3, 5, 11, 12, 18, 39]. The temporal profiles can probe features of the hidden space, such as Fermat paths [42], scene impulse response [23], and individual voxels [28]. Other works derive efficient algorithms using a confocal setup [2, 22, 27] and data-driven techniques [8]. NLOS has also been demonstrated with amplitude modulated continuous wave (AMCW) ToF by exploiting sparsity and fitting measurements to a Lambertian reflectance model [13, 16].

Other Active Methods: Chen et al. use steady-state (*i.e.* time-integrated) intensity information by leveraging partial directionality of diffuse reflectors [7]. Radio wavelengths can also be used to image through walls and/or around them [1, 32, 43]. There has also been significant progress in exploiting speckle correlations [9, 17, 25, 34] and synthetic wave holography [40, 41] for NLOS imaging. Henley et al. demonstrate the use of two-bounce light (*i.e.* shadows) for NLOS using intensity sensors [15]. An extension of this work uses ToF constraints on two-bounce pathlengths to isolate shadows from ambient background and higher-order inter-reflections [14]. However multiplexing was not used.

Passive Methods: Passive methods make use of optical information already present in the scene without an active source. Many passive approaches exploit polarization cues [10, 36], ambient thermal light [21, 36], shadows [35], multi-view reflections [38], and neural networks [33, 37] to reconstruct or make inferences about the scene outside the camera's field of view. *Occlusion-assisted imaging* makes use of information present at sharp discontinuities in LOS and NLOS scenes by turning them into virtual pinholes [4, 30] or

NLOS Imaging Domain	Key Components			Key Features		
	Assisting Surfaces	Sensor Type	Light Source	Number of Captures	Recovered Dimensions	Signal Quality
Passive Occlusion-Aided [4,31,33]	Relay Wall/Floor	Intensity	Passive	Low	2D	Low (strong background)
	Edge Occluders					
Three-Bounce Transient [18,27,30,42]	Relay Wall/Floor	ToF	Active (Scan)	High	3D	Medium (3-bounce)
	Edge Occluders					
Two-Bounce Transient (Ours)	Two LOS Walls	ToF	Active (Flash)	Low	3D	High (2-bounce)

Table 1. **Comparison of NLOS Methods.** Passive approaches require fewer measurements but can either recover up to two dimensions or are strongly affected by ambient noise. Works exploiting three-bounce light paths demonstrate 3D imaging around the corner but require a bounce on the hidden surface, reducing the signal quality. We analyze exploiting two-bounce transients for NLOS imaging.

virtual pinspecks [31]. Intensity-only methods are difficult to scale to scenarios with arbitrary lighting conditions.

3. Forward and Adjoint Operators

3.1. Problem Setup

We first introduce terminology and notation for our problem formulation. The object of interest lies behind an occluder and is flanked on two sides by planar Lambertian relay surfaces. Without loss of generality, the right wall is the *illumination wall* $\mathcal{L}$ and the left wall is the *observation wall* $\mathcal{O}$. Light incident on the illumination wall at a point l is a *virtual source*. Each virtual source diffusely radiates light towards the observation wall. Light will arrive at a *virtual detector* s from l if the object doesn't lie along the ray connecting s and l, as shown in Fig. 2a. If the object obstructs this ray, s is said to lie in the *shadow* of the object. We assume that every virtual source and virtual detector is visible to the laser and camera field of view, respectively. The hidden scene is discretized into a voxel grid described by a binary occupancy (or transparency) function. This implicity assumes that we are reconstructing opaque surfaces. We neglect the effect of higher order bounces (e.g. interreflections, 3B light) because SPADs can temporally gate out the longer ToF of 3B light. Furthermore, the intensity of higher order bounces will be much lower than that of two-bounce light because all surfaces are assumed to be diffuse.

We denote *light transients* as ToF measurements for light that arrives at the virtual detector and *shadow transients* as ToF measurements for light that didn't arrive at the virtual detector because of the object's shadow. Both of these are *two-bounce transients*. We show that modeling the problem in terms of shadow transients enables reconstruction of the hidden objects shape by solving a linear inverse problem.
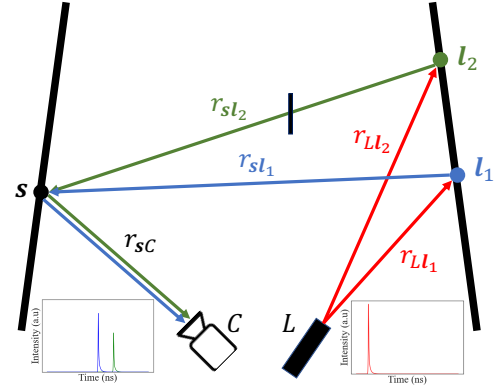

Figure 3. **Time-of-Flight Information for Demultiplexing Shadows.** Assume the point s lies in shadow. When using multiplexed illumination, there exists an ambiguity as to whether s lies in the shadow of l_1, l_2, or both. However, the paths L-l_1-s-C and L-l_2-s-C have different pathlengths, suggesting the usefulness of ToF information for demultiplexing shadows.

3.2. Linear Forward Model

Light Transients Consider the case where the occluded scene is empty. The measured transient is then given by

$$i_0(t; \mathbf{s}) = \alpha \iint_{\mathcal{L}} \frac{\delta(\|\mathbf{l} - \mathbf{g}\| + \|\mathbf{l} - \mathbf{s}\| - ct)}{\|\mathbf{l} - \mathbf{s}\|^2} d\mathbf{l}, \quad (1)$$

where **g** is the location of the laser, c is the speed of light, and α is the intensity of the laser. We refer to i_0 as the *empty transient*. For simplicity, we ignore α for the rest of the derivations in this section and assume that the length between the virtual detector and camera is calibrated or known. The numerator of the integrand

$$i(t; \mathbf{s}, \mathbf{l}) = \delta(\|\mathbf{l} - \mathbf{g}\| + \|\mathbf{l} - \mathbf{s}\| - ct) \quad (2)$$

describes the equation of an ellipsoid, with foci at **g** and **s** and major axis length ct. When an object is present in the
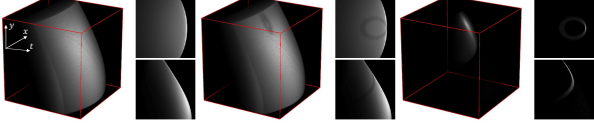
Figure 4. **Light and Shadow Transients.** Three spacetime curves are depicted. (left) Empty transient i_0 with no hidden object, (middle) light transient i_l with an object, and (right) shadow transient $i_s = i_0 - i_l$. Switching from light transients i_l to shadow transients i_s enables derivation of a linear forward model.

occluded scene, the measured light transient is given by

$$i_l(t; \mathbf{s}) = \iint_{\mathcal{L}} V_{\mathbf{s},\mathbf{l}} \cdot \frac{i(t; \mathbf{s}, \mathbf{l})}{\|\mathbf{l} - \mathbf{s}\|^2} d\mathbf{l}, \tag{3}$$

where $V_{\mathbf{s},\mathbf{l}}$ is the visibility between $\mathbf{s}$ and $\mathbf{l}$.

$$V_{\mathbf{s},\mathbf{l}} = \exp\left(\oint_{\mathbf{l}}^{\mathbf{s}} \ln(f_t(\mathbf{x})) d\mathbf{x} \right) \tag{4}$$

where Eq. (4) computes the product integral along the line connecting $\mathbf{l}$ and $\mathbf{s}$ and $f_t(\mathbf{x}) \in \{0,1\}$ is the transparency function of voxel $\mathbf{x}$ in the occluded region.

Shadow Transients The non-linearity of Eq. (4) makes model inversion challenging with light transients. Casting the formulation from light space-time to *shadow space-time* enables a linear formulation of the problem. A shadow transient at $\mathbf{s}$ is given by

$$i_s(t; \mathbf{s}) = i_0(t; \mathbf{s}) - i_l(t; \mathbf{s}). \tag{5}$$

Fig. 4 illustrates the difference between light transients and shadow transients. Shadow transients directly probe voxel occupancy, whereas light transients probe for voxel transparency. For example, if the occluded scene is empty except for voxel $\mathbf{x}$, the shadow transient is given by

$$i_s(t; \mathbf{s}, \mathbf{x}) = \frac{\delta(\|\mathbf{l}(\mathbf{x}, \mathbf{s}) - \mathbf{g}\| + \|\mathbf{l}(\mathbf{x}, \mathbf{s}) - \mathbf{s}\| - ct)}{\|\mathbf{l}(\mathbf{x}, \mathbf{s}) - \mathbf{s}\|^2}, \tag{6}$$

where $\mathbf{l}(\mathbf{x}, \mathbf{s})$ is a point on the illumination wall that lies along the ray connecting $\mathbf{x}$ and $\mathbf{s}$. If the wall doesn't intersect this ray, $i_s(t; \mathbf{x}, \mathbf{s}) = 0$.

Linear Relaxation Now, consider the general case where a single opaque object lies in the hidden scene. We can describe the shadow transient as

$$i_s(t; \mathbf{s}) = \iint_{\mathcal{L}} O_{\mathbf{s},\mathbf{l}} \cdot \frac{i(t; \mathbf{s}, \mathbf{l})}{\|\mathbf{l} - \mathbf{s}\|^2} d\mathbf{l}, \tag{7}$$

where $O_{\mathbf{s},\mathbf{l}} \in \{0,1\}$ is the occlusion function (denotes if the ray between $\mathbf{l}$ and $\mathbf{s}$ is occluded). It can be expressed as the indicator function

$$O_{\mathbf{s},\mathbf{l}} = \begin{cases} 1, & \text{if } \oint_{\mathbf{l}}^{\mathbf{s}} f_o(\mathbf{x}) d\mathbf{x} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$
$$\approx \oint_{\mathbf{l}}^{\mathbf{s}} f_o(\mathbf{x}) d\mathbf{x},$$

where $f_o(\mathbf{x})$ is the voxel occupancy function of the hidden scene. The linear relaxation in Eq. (8) is possible because NLOS reconstructions typically only consider ToF information, but not radiometric information. The above formulation will preserve the structure of the streak images in spacetime, even if the intensities are overestimated. Combining Eq. (5), (7), and (8), we obtain

$$i_0(t; \mathbf{s}) - i_l(t; \mathbf{s}) = \iint_{\mathcal{L}} \frac{i(t; \mathbf{s}, \mathbf{l})}{\|\mathbf{l} - \mathbf{s}\|^2} \cdot \left(\oint_{\mathbf{l}}^{\mathbf{s}} f_o(\mathbf{x}) d\mathbf{x} \right) d\mathbf{l}, \tag{9}$$

where we solve for $f_o(\mathbf{x})$. We can estimate i_0 from the scene geometry and i_l is the measured transient.

Discrete Beam Illumination Discretizing $\mathcal{L}$ as K discrete laser spots and the hidden space into a voxel grid, we can write Eq. (9) as

$$i_s[t; \mathbf{s}] = \sum_{k=1}^{K} \frac{\delta[\|\mathbf{l}_k - \mathbf{g}\| + \|\mathbf{l}_k - \mathbf{s}\| - ct]}{\|\mathbf{l}_k - \mathbf{s}\|^2} \cdot \left(\sum_{\mathbf{x} \in \mathcal{P}_{\mathbf{l}_k, \mathbf{s}}} f_o[\mathbf{x}] \right), \tag{10}$$

where $\mathcal{P}_{\mathbf{l}, \mathbf{s}}$ denotes the set of voxels lying along the path between $\mathbf{l}$ and $\mathbf{s}$. Eq. (10) can be expressed in matrix form

$$\mathbf{I}_s = \mathbf{PBf} = \mathbf{Af}, \tag{11}$$

where $\mathbf{I}_s \in \mathbb{R}^{n_u n_v n_t}$ is the vectorized space-time measurements, $\mathbf{f} \in \mathbb{R}^{n_x n_y n_z}$ is the voxel occupancies, and $\mathbf{A} \in \mathbb{R}^{n_u n_v n_t \times n_x n_y n_z}$ is the measurement operator mapping voxel occupancies to space-time measurements. n_u and n_v are the number of pixels along the u and v direction, and n_t is the number of timing bins. n_x, n_y, and n_z are the number of voxels in the x, y, and z directions.

$\mathbf{P}$ and $\mathbf{B}$ model the temporal and spatial relationships, respectively, between virtual sources, virtual detectors, and voxels. $\mathbf{P} \in \mathbb{R}^{n_u n_v n_t \times n_u n_v K}$ is a block diagonal matrix, with the column space of the ith block $\mathbf{P}^i \in \mathbb{R}^{n_t \times K}$ containing the corresponding $\delta(\cdot)$ functions from Eq. (2) for virtual detector $\mathbf{s}_i$ and all K virtual sources. $\mathbf{B} \in \mathbb{R}^{n_u n_v K \times n_x n_y n_z}$ is composed of $n_u n_v$ vertical sub-blocks. Each entry $\mathbf{B}^i_{kj} \in \{0,1\}$ of the ith sub-block $\mathbf{B}^i \in \mathbb{R}^{K \times n_x n_y n_z}$ of sub-block i indicates whether voxel j lies along the line connecting virtual detector $\mathbf{s}_i$ and virtual source $\mathbf{l}_k$. The multiplication $\mathbf{Bf}$ performs the line integral in Eq. (8). The kth column vector of $\mathbf{P}$ is the space-time impulse response (STIR) to the kth illumination source when the hidden scene is completely occupied (note that we are dealing with shadow transients, not light transients). The ith column of $\mathbf{A}$ forms the expected space-time measurements

if only the ith voxel is occupied (*i.e.* the STIR to a single voxel). We neglect the r^2 falloff term in our reconstructions. Additional explanation of the $\mathbf{A}$ matrix is provided in the supplementary. The linear model in Eq. (11) resembles that of 3B-bounce NLOS, where $\mathbf{A}$ is a measurement operator that maps voxels $\mathbf{f}$ to measurements $\mathbf{I}$. However, in 3B-NLOS, the information is encoded into the *reflectance* of the object, whereas in 2B-NLOS, the information is encoded into the *shadows*. Therefore, the former is linear with respects to light transients, and the latter is linear with respect to shadow transients.

3.3. Reconstruction Algorithm

Reconstruction of $\mathbf{f}$ can be performed by backprojection

$$\mathbf{f}_{bp} = \mathbf{A}^\top \mathbf{I}_s. \tag{12}$$

Backprojection produces a low-frequency reconstruction of $\mathbf{f}$. In practice, filtering $\mathbf{f}_{bp}$ with a Laplacian filter

$$\mathbf{f}_{fbp} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \mathbf{f}_{bp} \tag{13}$$

has been shown to yield reasonably good results in ellipsoidal tomography in 3-bounce NLOS and line-based tomography in medical imaging. Our linear formulation also enables reconstruction with priors. We recognize that these optimization-based methods would improve reconstruction over unfiltered backprojection, which is equivalent to a low-frequency version of voxel carving in our setup. However, we restrict our focus to unfiltered backprojection because of its convenient relationship with the Gram matrix, which enables further analysis of our method in Sec. 4.2.

4. Design and Analysis

In this section, we analyze how the performance of the proposed two-bounce transient system varies with system parameters and the geometry of LOS relay surfaces. We also analyze the impact of temporal resolution, spatial resolution, and number of measurements on recoverability. The amount of multiplexing in the scene is inversely related to the number of measurements. For example, if 60 laser spots are used and 15 measurements are captured, then each image contains 4 multiplexed shadows. We use SNR, coherence, and FWHM of the PSF to quantify performance.

4.1. Voxel SNR Analysis

We quantify the noise characteristics of the measured shadow two-bounce transients using per-voxel SNR. Consider the case where only voxel $\mathbf{x}$ is occupied in the hidden region. We derive the SNR of the number of "shadow" photons received in the measured shadow transient $\mathbf{I}_s^{\mathbf{x}}$. In vector
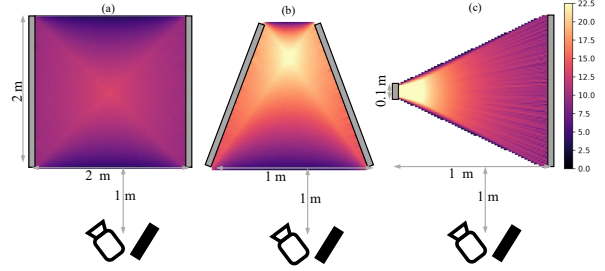

Figure 5. **SNR Analysis.** The goal of this analysis is to determine how scene geometry affects the SNR of the two-bounce measurement at each voxel in the hidden scene. The first two cases consider the orientations of the relay surfaces, and the third considers the field of view of the camera. SNR is highest in the middle of the hidden scene for parallel surfaces, orienting the walls closer together increases SNR due to reduced r^2 falloff, and SNR is highest closer to the virtual detectors for small FOV cameras.

form, the occupancy function can be described as $\boldsymbol{\delta}^{\mathbf{x}}$. From Eq. (11), the shadow transient is given by

$$\mathbf{I}_s^{\mathbf{x}} = \mathbf{A}\boldsymbol{\delta}^{\mathbf{x}} \tag{14}$$

and the light transients measured by the SPAD sensor are given by $\mathbf{I}_l^{\mathbf{x}}$. For this analysis, we account for r^2 falloff because it is a physically accurate phenomenon that impacts SNR. From Eq. (5),

$$\mathbf{I}_l^{\mathbf{x}} = \mathbf{I}_0 - \mathbf{A}\boldsymbol{\delta}^{\mathbf{x}} \tag{15}$$

where $\mathbf{I}_0$ is the empty transient and assumed to be calibrated. Assuming no SPAD pile-up distortion, the number of photons arriving at the sensor for each time bin of $\mathbf{I}_l^{\mathbf{x}}$ follow an independent Poisson distribution [29]. Thus, the total number of photons $P^{\mathbf{x}}$ measured follows a Poisson distribution with the rate parameter being the sum of all the photon counts for all time bins for all pixels.

$$P^{\mathbf{x}} \sim \text{Poisson}(\mathbf{1}^T(\mathbf{I}_0 - \mathbf{A}\boldsymbol{\delta}^{\mathbf{x}})) \tag{16}$$

where $\mathbf{1} \in \mathbb{R}^{n_u n_v n_t}$ is the one vector. Using the fact that SNR of a Poisson distribution is $\mu/\sigma = \sqrt{\lambda}$ with rate parameter λ, the per-voxel SNR $S^{\mathbf{x}}$ is given by

$$S^{\mathbf{x}} = \sqrt{\mathbf{1}^T(\mathbf{I}_0 - \mathbf{A}\boldsymbol{\delta}^{\mathbf{x}})} \tag{17}$$

The per-voxel SNR is plotted in Fig. 5 for three different scenarios. The first two scenarios deal with the orientation of the walls, and the third considers the field of view of the camera. In Fig. 5a, we can see that the SNR is higher for center voxels, as more rays connecting the illumination wall and observation wall pass through the center voxels. When the relay walls are oriented towards each other (Fig. 5b), the SNR increases due to reduced r^2 falloff. When the camera's field of view is small compared to the laser (Fig. 5c), regions closer to the observation wall have the highest SNR.
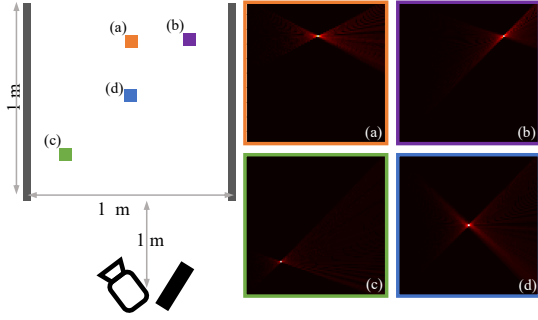
Figure 6. **Spatially Varying PSF.** In two-bounce lidar, the PSF has a double cone shape. The orientation of the double cone varies strongly as a function of the orientation and length of the relay surfaces (*i.e.* the virtual apertures) and the voxel position.

4.2. Recoverability Analysis

Resolution We now study the resolution limits of our method. In particular, we consider two key metrics in our measurement operator: mutual coherence and full width at half maximum (FWHM) of the PSF. We are interested in how (1) temporal resolution, (2) spatial resolution N, and (3) number of measurements M impact each of our metrics. From Eq. (12), we know that the backprojected occupancy can be obtained from the occupancy function by

$$\mathbf{f}_{bp} = (\mathbf{A}^\top \mathbf{A})\mathbf{f}, \tag{18}$$

where $\mathbf{G} = \mathbf{A}^\top \mathbf{A}$ is the Gram matrix. We can interpret the columns (or rows) of the Gram matrix to be the spatially varying point spread function (PSF) of the imaging system, as shown in Fig. 6. Each column $\mathbf{G}_i$ describes how a voxel $\mathbf{x}_i$ is blurred during backprojection reconstruction.

We can see the effect of different imaging parameters on the PSF in Fig. 7, where we sweep different values of (1) M and (2) N. In particular, we note that the lobes of the PSF are suppressed when using temporal information, even with low spatial resolution and few number of image measurements. When using no temporal information, the PSF is still roughly centered at the correct location, but has large FWHM. The sharpness of the PSF when using temporal information indicates that it is better suited to applications that require finer voxel resolution with few-shot capture. In case (1) without transients, we see that the FWHM of the PSF is large along both the x and y direction for small values of M. As M approaches the number of illumination spots, the FWHM approaches the theoretical lower bound. In case (2), increased spatial resolution helps better localize the center of the PSF, but the FWHM does not improve. The transient cases are quite robust to both low number of measurements and low spatial resolution. Intuitively, higher values of N in the transient case slightly help accentuate the center voxel

by projecting more diverse rays through the center voxel.

Mutual Coherence We wish to further quantify the impact of spatio-temporal resolution and number of measurements on the recoverability of the hidden scene. We know from compressive sensing theory that sparse recovery of the hidden scene is possible when the mutual coherence of $\mathbf{A}$ is sufficiently small. The mutual coherence is defined as

$$\mu = \max_{1 \le i \ne j \le n} \frac{|\mathbf{A}_i^\top \mathbf{A}_j|}{|\mathbf{A}_i||\mathbf{A}_j|}. \tag{19}$$

Intuitively, the coherence measures the largest "similarity" or correlation between any two column vectors in measurement matrix $\mathbf{A}$ [6]. Physically, this quantifies how similar the STIRs of two different voxels are. If $\mu = 1$, then two voxels have the exact same STIR, which means that there exists an ambiguity as to which voxel resulted in a certain spatio-temporal response. Alternatively, μ can be defined as the maximum off-diagonal entry of the Gram matrix. Fig. 8 illustrates how the Gram matrix (and subsequently the mutual coherence) is affected by coarser timing resolutions. From compressive sensing theory, we know that a K-sparse scene can be recovered if $K \le 0.5(1 + 1/\mu)$, which motivates our empirical analysis of the effect of spatial resolution and temporal resolution on μ.

From Fig. 9, we can see the impact that spatial and temporal resolution have on mutual coherence in a single-shot setup with multiplexed illumination. Consider a scene similar to that of Fig. 6, where the walls are 2 m apart, the walls are 3 m long, the axial distance to the walls is 2 m, and the voxel resolution is discretized to 2 cm. Increasing the spatial resolution at a fixed temporal resolution appears to have little to no effect on the mutual coherence. As Kadambi et al. also point out, improvements of coherence through higher spatial resolution are limited by the size of the virtual aperture (*i.e.* the wall) [16]. On the other hand, the difference between 100 ps resolution and 10 ps resolution reduces the coherence from 0.9 to 0.3 for a fixed spatial resolution. Any timing resolution coarser than $\sim$ 120 ps results in a coherence of 1 for multiplexed measurements because pathlength differences for different shadows are no longer resolvable. This analysis motivates the importance of temporal resolution when using multiplexed illumination.

5. Experiments and Results

We scan a single-pixel SPAD across the field of view on the left wall to mimic the output of a SPAD array. We scan a single laser spot across the right wall, and sum up the corresponding transient images in post-processing to mimic the effect of multiplexing. Further details are provided in the supplementary. We first validate our model using simulated data in Fig. 10, where we demonstrate that we can localize four letters and reconstruct their 2D shape with just a single

(Most multiplexing) — Number of measurements — **(No multiplexing)**

Number of Pixels

$M = 1$ $M = 10$ $M = 30$ $M = 60$ $N = 10$ $N = 20$ $N = 30$ $N = 40$

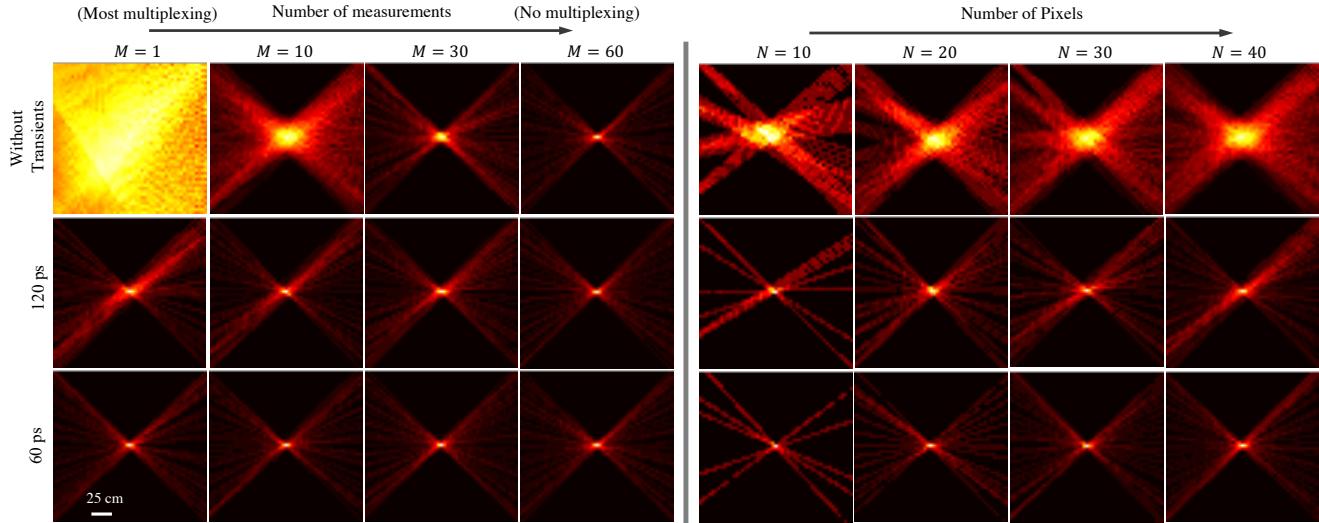Without Transients

120 ps

60 ps

25 cm

Figure 7. **PSF as a Function of Number of Measurements and Spatial Resolution.** The PSF of the center voxel is shown in the transient and intensity-only case when multiplexed with 60 laser spots. (1) the number of pixels $N = 40$ are kept fixed with the number of measurements M being swept (left) and (2) $M = 10$ is kept fixed and N is swept (right). Transient measurements result in PSFs with narrower FWHMs, even with low spatial resolution and low number of measurements.
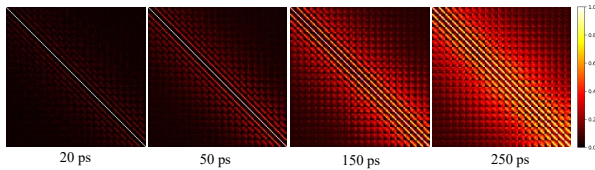

20 ps   50 ps   150 ps   250 ps

Figure 8. **Gram Matrix as a Function of Temporal Resolution.** For a defined voxel resolution and fixed spatial resolution. The columns (or rows) of the Gram matrix form the spatially varying PSFs shown in Fig. 6. The off-diagonal streaks describe the coherence between different voxels. As expected, higher temporal resolution decreases the coherence of the Gram matrix, which increases the likelihood of guaranteed sparse recovery.
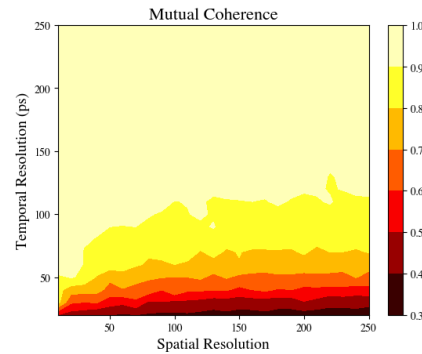

Figure 9. **Tradeoffs Between Spatial Resolution and Temporal Resolution.** We quantify the tradeoff using mutual coherence. We use the same setup as Fig. 6, but the walls have length 3 m, the walls are space 2 m apart, and the distance to the walls is 2 m. We discretize the voxel resolution to 2 cm (~ 67 ps). The bottom right of the figure, where spatial and temporal resolution are highest is where mutual coherence is the lowest. For a fixed temporal resolution, improvements in coherence quickly plateau. Improvements in temporal resolution dominate improvements in spatial resolution when reducing coherence, suggesting the importance of ToF for demultiplexing illumination in two-bounce NLOS imaging.

capture. We render i_o and i_l to obtain a noiseless version of the shadow transient for this experiment. Then, we reconstruct the shapes of real mannequins and must estimate i_0. Fig. 11 shows that ToF measurements provide a noticeable improvement in image reconstruction. When there's no multiplexing (*i.e.* the number of laser spots is the same as number of measurements), the intensity-based method is similar to the ToF-based method. However, in the intensity-only case, the object becomes increasingly blurry until it is no longer visible as the amount of multiplexing increases. In the ToF case for single-shot, we can see a blurry image roughly localized where the center of the mannequin is as well as some high-frequency detail. Furthermore, the top view of the ToF reconstructions show that the high-frequency shape of the arms can be reconstructed even with multiplexed illumination, indicating the promise of single-

shot capture with two-bounce ToF information. Fig. 12 shows that poorer temporal resolution results in aliasing artifacts (repeated arms and legs), whereas the rough shape of the object can still be seen with lower spatial resolution. Finally, we compare our method to the probabilistic voxel carving method used in [15] in Fig. 13. Voxel carving results in sharper reconstructions than backprojection.
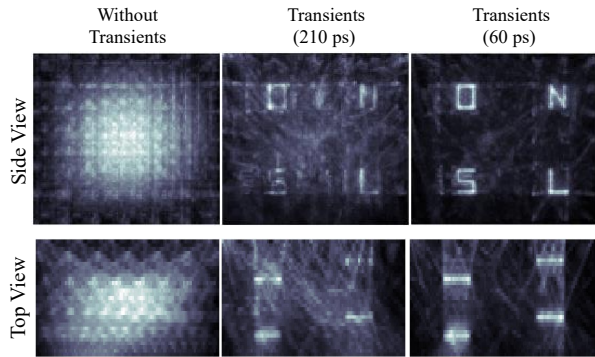
Figure 10. **Single-Shot Localization.** We obtain single-shot reconstructions of the "N", "L", "O", and "S" letters in the hidden scene with and without transient information. We see that using transient information, we are able to effectively localize the hidden object and even get a sharp outline of the letters, demonstrating the promise of single-shot capture with two-bounce transients.
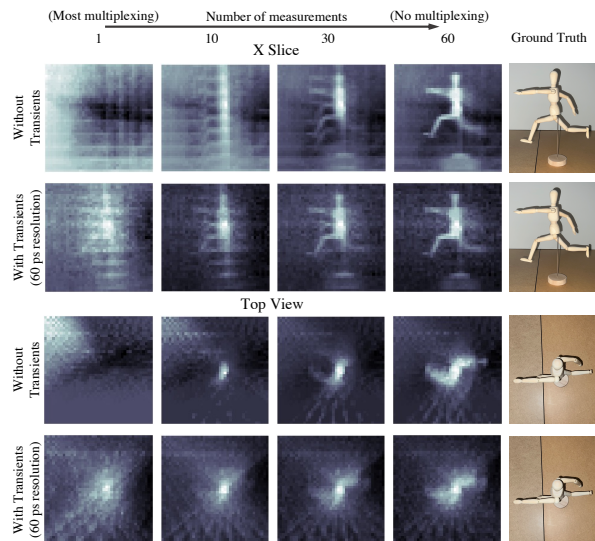

Figure 11. **Transients Enable Scene Recovery from Multiplexed Illumination.** We perform backprojection of measurements with varying levels of multiplexing. Plotted in the top two rows are x slices of the reconstructed volume. Under no multiplexing (right), intensity is sufficient to carve the shadow along detected pixel. But reconstruction quality sharply deteriorates as we move towards the multiplexing regime (left). When measuring with transients, time-of-flight information allows us to disambiguate between different shadows and recover a rough silhouette of the object even with just a single measurement.

However, our method can be improved by filtering or incorporating priors, and is also generalizable to multiplexed illumination by utilizing ToF information.
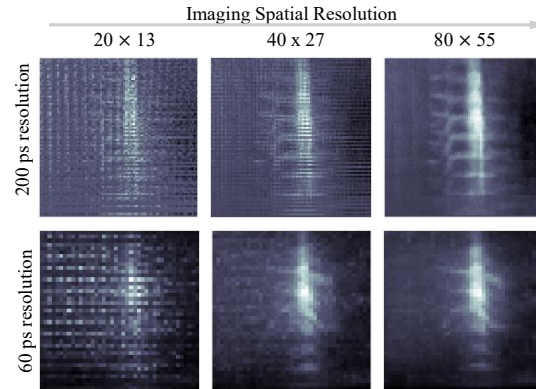

Figure 12. **Spatial Resolution vs. Timing Resolution.** We plot reconstructions with different spatio-temporal resolutions. Lower timing resolution results in aliasing artifacts (*i.e.* repeated arms and legs). Lower spatial resolution introduces hole artifacts.
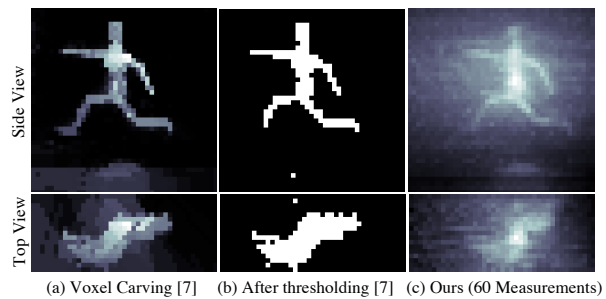

Figure 13. **Scanning Baseline Comparison.** We compare to the voxel carving method used in [15] (a) before and (b) after thresholding when scanning 1 laser spot. Our method is more general, but voxel carving yields sharper results with laser scanning.

6. Discussion

In this paper, we introduced the use of two-bounce transients for NLOS imaging. Compared to 3B-NLOS, 2B-NLOS enables more robust imaging of occluded objects in the presence of scenes with arbitrary lighting conditions and low albedo objects. Transient 2B-NLOS is promising for few-shot capture of occluded objects by using multiplexed illumination, especially as SPAD array technology matures. When comparing tradeoffs in temporal resolution, spatial resolution, and number of measurements, we qualitatively and quantitatively verify with real and simulated results that temporal resolution is most important for demultiplexing shadows. We anticipate future work to extend the analysis in this paper to SPAD arrays for real-time NLOS imaging.

Acknowledgements

References

[1] Fadel Adib and Dina Katabi. See through walls with wifi! In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pages 75–86, 2013. 2

[2] Byeongjoo Ahn, Akshat Dave, Ashok Veeraraghavan, Ioannis Gkioulekas, and Aswin C Sankaranarayanan. Convolutional approximations to the general non-line-of-sight imaging operator. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7889–7899, 2019. 2

[3] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. In *ACM SIGGRAPH 2017 Posters*, pages 1–2. 2017. 2

[4] Katherine L Bouman, Vickie Ye, Adam B Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. Turning corners into cameras: Principles and methods. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2270–2278, 2017. 2, 3

[5] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 2

[6] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008. 6

[7] Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. Steady-state non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6790–6799, 2019. 2

[8] Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz, and Felix Heide. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Transactions on Graphics (ToG)*, 39(6):1–18, 2020. 2

[9] Akshat Dave, Muralidhar Madabhushi Balaji, Prasanna Rangarajan, Ashok Veeraraghavan, and Marc P Christensen. Foveated non-line-of-sight imaging. In *Computational Optical Sensing and Imaging*, pages CTh5C–6. Optica Publishing Group, 2020. 2

[10] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. Pandora: Polarization-aided neural decomposition of radiance. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 538–556. Springer, 2022. 2

[11] Genevieve Gariepy, Francesco Tonolini, Robert Henderson, Jonathan Leach, and Daniele Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1):23–26, 2016. 2

[12] Felix Heide, Matthew O'Toole, Kai Zang, David B Lindell, Steven Diamond, and Gordon Wetzstein. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Transactions on Graphics (ToG)*, 38(3):1–10, 2019. 2

[13] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014. 2

[14] Connor Henley, Joseph Hollmann, and Ramesh Raskar. Bounce-flash lidar. *IEEE Transactions on Computational Imaging*, 8:411–424, 2022. 2

[15] Connor Henley, Tomohiro Maeda, Tristan Swedish, and Ramesh Raskar. Imaging behind occluders using two-bounce light. In *ECCV*, pages 573–588. Springer, 2020. 1, 2, 7, 8

[16] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics (ToG)*, 35(2):1–12, 2016. 2, 6

[17] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature photonics*, 8(10):784–790, 2014. 2

[18] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009. 2, 3

[19] Oichi Kumagai, Junichi Ohmachi, Masao Matsumura, Shinichiro Yagi, Kenichi Tayu, Keitaro Amagawa, Tomohiro Matsukawa, Osamu Ozawa, Daisuke Hirono, Yasuhiro Shinozuka, et al. 7.3 a 189× 600 back-illuminated stacked spad direct time-of-flight depth sensor for automotive lidar systems. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, volume 64, pages 110–112. IEEE, 2021. 2

[20] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2):150–162, 1994. 1

[21] Di Lin, Connor Hashemi, and James R Leger. Passive non-line-of-sight imaging using plenoptic information. *JOSA A*, 37(4):540–551, 2020. 2

[22] David B Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics (ToG)*, 38(4):1–13, 2019. 2

[23] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572(7771):620–623, 2019. 2

[24] Tomohiro Maeda, Guy Satat, Tristan Swedish, Lagnojita Sinha, and Ramesh Raskar. Recent advances in imaging around corners. *arXiv preprint arXiv:1910.05613*, 2019. 1

[25] Christopher A Metzler, Felix Heide, Prasana Rangarajan, Muralidhar Madabhushi Balaji, Aparna Viswanath, Ashok Veeraraghavan, and Richard G Baraniuk. Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging. *Optica*, 7(1):63–71, 2020. 2

[26] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Megapixel time-gated spad image sensor for 2d and 3d imaging applications. *Optica*, 7(4):346–354, 2020. 2

[27] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, 2018. 2, 3

[28] Adithya Pediredla, Akshat Dave, and Ashok Veeraraghavan. Snlos: Non-line-of-sight scanning through temporal focus-

ing. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–13. IEEE, 2019. 2

[29] Adithya K Pediredla, Aswin C Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal processing based pile-up compensation for gated single-photon avalanche diodes. *arXiv preprint arXiv:1806.07437*, 2018. 5

[30] Joshua Rapp, Charles Saunders, Julián Tachella, John Murray-Bruce, Yoann Altmann, Jean-Yves Tourneret, Stephen McLaughlin, Robin Dawson, Franco NC Wong, and Vivek K Goyal. Seeing around corners with edge-resolved transient imaging. *Nature communications*, 11(1):1–10, 2020. 2, 3

[31] Charles Saunders, John Murray-Bruce, and Vivek K Goyal. Computational periscopy with an ordinary digital camera. *Nature*, 565(7740):472–475, 2019. 3

[32] Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jurgen Dickmann, Klaus Dietmayer, Bernhard Sick, et al. Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2068–2077, 2020. 2

[33] Prafull Sharma, Miika Aittala, Yoav Y Schechner, Antonio Torralba, Gregory W Wornell, William T Freeman, and Frédo Durand. What you can learn by staring at a blank wall. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2330–2339, 2021. 2, 3

[34] Brandon M Smith, Matthew O'Toole, and Mohit Gupta. Tracking multiple objects outside the line of sight using speckle imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6258–6266, 2018. 2

[35] Tristan Swedish, Connor Henley, and Ramesh Raskar. Objects as cameras: Estimating high-frequency illumination from shadows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2593–2602, 2021. 2

[36] Kenichiro Tanaka, Yasuhiro Mukaigawa, and Achuta Kadambi. Polarized non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2136–2145, 2020. 2

[37] Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018. 2

[38] Kushagra Tiwary, Askhat Dave, Nikhil Behari, Tzofi Klinghoffer, Ashok Veeraraghavan, and Ramesh Raskar. Orca: Glossy objects as radiance field cameras. *arXiv preprint arXiv:2212.04531*, 2022. 2

[39] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3(1):1–8, 2012. 2

[40] Florian Willomitzer, Fengqiang Li, Muralidhar Madabhushi Balaji, Prasanna Rangarajan, and Oliver Cossair. High resolution non-line-of-sight imaging with superheterodyne re-

mote digital holography. In *Computational Optical Sensing and Imaging*, pages CM2A–2. Optica Publishing Group, 2019. 2

[41] Florian Willomitzer, Prasanna V Rangarajan, Fengqiang Li, Muralidhar M Balaji, Marc P Christensen, and Oliver Cossairt. Fast non-line-of-sight imaging with high-resolution and wide field of view using synthetic wavelength holography. *Nature communications*, 12(1):6647, 2021. 2

[42] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and Ioannis Gkioulekas. 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*, pages 6800–6809, 2019. 2, 3

[43] Shichao Yue, Hao He, Peng Cao, Kaiwen Zha, Masayuki Koizumi, and Dina Katabi. Cornerradar: Rf-based indoor localization around corners. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–24, 2022. 2

[44] Chao Zhang, Ning Zhang, Zhijie Ma, Letian Wang, Yu Qin, Jieyang Jia, and Kai Zang. A 240× 160 3d-stacked spad dtof image sensor with rolling shutter and in-pixel histogram for mobile devices. *IEEE Open Journal of the Solid-State Circuits Society*, 2:3–11, 2021. 2