Keyhole Imaging: Non-Line-of-Sight Imaging and Tracking of Moving Objects Along a Single Optical Path

Christopher A. Metzler, Member, IEEE, David B. Lindell, Member, IEEE, and Gordon Wetzstein, Senior Member, IEEE

Abstract—Non-line-of-sight (NLOS) imaging and tracking is an emerging technology that allows the shape or position of objects around corners or behind diffusers to be recovered from transient, time-of-flight measurements. However, existing NLOS approaches require the imaging system to scan a large area on a visible surface, where the indirect light paths of hidden objects are sampled. In many applications, such as robotic vision or autonomous driving, optical access to a large scanning area may not be available, which severely limits the practicality of existing NLOS techniques. Here, we propose a new approach, dubbed keyhole imaging, that captures a sequence of transient measurements along a single optical path, for example, through a keyhole. Assuming that the hidden object of interest moves during the acquisition time, we effectively capture a series of time-resolved projections of the object’s shape from unknown viewpoints. We derive inverse methods based on expectation-maximization to recover the object’s shape and location using these measurements. Then, with the help of long exposure times and retroreflective tape, we demonstrate successful experimental results with a prototype keyhole imaging system.

Index Terms—Non-line-of-sight, time-of-flight, unknown-view tomography.

I. INTRODUCTION

IMAGING or tracking objects outside a camera’s direct line of sight has important applications in autonomous driving, robotic vision, and many other areas. By analyzing the time-of-flight of indirectly scattered light, transient non-line-of-sight (NLOS) imaging approaches (e.g., [1]–[9]) are a particularly promising approach to seeing around corners at large scales.

However, all of these techniques rely on time-resolved measurements of indirect light transport that are captured by sampling a large area of a surface within the line of sight of the imaging system. The sampled area acts as a synthetic aperture and it was shown that the resolution of estimated hidden scenes, and also the accuracy of NLOS tracking, is primarily limited by both the size of the sampling area and the temporal resolution of the imaging system [4]. Unfortunately, in many applications, such as navigation in tight quarters or at long standoffs, optical access to a large sampling surface may not be feasible, making current NLOS techniques impractical.1

Here, we introduce a new form of NLOS imaging where transient measurements are recorded along a single optical path. As shown in Fig. 1, one example of an optical configuration that is enabled by our approach is that of sampling a single point on the wall of a room through a keyhole. Although imaging through a keyhole may be possible by moving a small camera very close to the keyhole, our approach operates at a standoff distance from the keyhole and only requires a single optical path from the light source–detector pair to the sampling point inside the room.

1Another barrier to eye-safe, real-time NLOS imaging is light throughput. The challenging radiometry associated with keyhole imaging with diffuse objects is discussed in the supplement.
Trying to recover the shape or to track the location of objects in a static room with such a configuration is a highly ill-posed problem. However, as was recognized by Bouman et al. [10], object motion can make otherwise ill-posed NLOS imaging problems tractable. Unlike their work, however, keyhole imaging only has access to a single optical light path. To account for this, our work makes use of the intuition that measurements of a moving object captured with a fixed sensor are equivalent to measurements of a fixed object captured with a moving “virtual” sensor. This is the same principle that underlies problems in other domains, such as inverse synthetic aperture radar (ISAR) [11]. Just as NLOS imaging is analogous to SAR and tomography [12], keyhole imaging of a moving object is analogous to ISAR and unknown-view tomography. Like ISAR and unknown-view tomography, keyhole imaging requires solving a challenging and highly non-convex inverse problem: jointly estimating both the shape of the hidden object and its unknown motion. We address this problem using a variant of the expectation-maximization (EM) algorithm.

To our knowledge, this is the first work to study the keyhole imaging problem, which could enable an entirely new class of NLOS imaging problems to be addressed. In doing so, this work makes several distinct contributions:

- We introduce the keyhole imaging problem: NLOS imaging or tracking where the relay wall consists of just a single visible point.
- We develop a EM-based algorithm that uses the unknown motion of the object to solve the keyhole imaging problem.
- We evaluate the performance of this method for various types of motion, which model both translation and rotation, in extensive simulations.
- We build a prototype system and experimentally demonstrate NLOS imaging and tracking through a keyhole.

Assumptions and Limitations: Our present work relies on several assumptions and simplifications that help in tackling the challenging keyhole imaging problem. First, we assume the hidden object exhibits only rigid body motion. Second, we assume the object does not self-occlude nor produce interreflections. Third, to greatly reduce computational complexity, our present reconstructions of the trajectories and the objects (even those that are non-planar, like the mannequin in Fig. 1) are 2D. Fourth, it is worth pointing out that our reconstructions have an inherent flip ambiguity. For example, measurements from an object moving from left to right are indistinguishable from measurements of horizontally flipped but otherwise identical object moving from right to left. Similarly, our method cannot distinguish between an object rotating in place and an object translating along a circular trajectory. Finally, because of the limited power of our laser source, we covered the hidden objects in retroreflective tape and set the total exposure time to a few minutes so as to increase the strength of the returning signals. As a result, our reconstructions have access to $2\sim5 \times$ more photons than one would detect when imaging a diffuse, white, one square meter hidden objects at a two meter standoff with a barely eye-safe 9.9 mW 1550 nm laser over a five second exposure time. See the supplement for more information about eye-safe laser power limits and the radiometry associated with room-scale keyhole imaging of diffuse and retroreflective objects.

II. RELATED WORK

NLOS Imaging and Tracking: While long theorized [13], [14], NLOS imaging of static objects was first successfully demonstrated in 2012 [15]. Since then, much progress has been made on developing more efficient data acquisition setups and inverse methods [8]. The majority of these works reconstruct hidden scenes from transient measurements captured using pulsed [1], [2], [4], [4]−[7], [9], [16]−[25] or continuous-wave [5], [26], [27] active illumination, or using a coherent source that creates speckle [28]−[32]. Acoustic NLOS imaging has also been explored [33], [34], as have methods that forgo active illumination and instead rely opportunistically on occlusions or illumination sources in the scene itself [10], [35]−[38]. All these works require a large relay wall to sample the indirect light transport inside their line of sight.

More recently, several groups have developed methods to image or track moving hidden objects. This feat can be accomplished by either ignoring motion and using powerful illumination to form frame-by-frame triangulations [39] or reconstructions [7], [40] in real-time or by leveraging the differences between subsequent frames with object motion to track them [3], [10], [41]−[44]. These methods too rely upon large relay walls.

Keyhole imaging is a new NLOS imaging modality. Unlike other NLOS approaches, it only requires access to a single optical light path. The keyhole imaging inverse problem is significantly more challenging than conventional NLOS imaging because the locations of the sampling points are unknown.

Unknown-View Tomography: The problem of reconstructing a 3D or 2D object from a series of 2D or 1D projections, respectively, taken at unknown view angles is known as unknown-view tomography [45]. This problem occurs in a number of applications; most notably inverse synthetic aperture radar (ISAR) [11] and single-particle cryogenic electron microscopy [46], [47].

A host of methods exist to solve the unknown-view tomography problem: Moment methods compute perspective invariant features and use these features to directly reconstruct the scene [48]−[50]. Low-dimensional embedding methods estimate the perspective associated with each measurement by embedding it into a low-dimensional space, this estimated perspective then allows for the application of standard recovery algorithms [51], [52]. Bayesian methods, which are the most popular and successful approach to unknown-view tomography, compute maximum likelihood (ML) or maximum a posteriori (MAP) estimates of the scene under priors on the distribution of the scene, the measurements, and the noise [53], [54]. Bayesian methods are sample efficient and highly robust to noise, but can be computationally demanding.

Keyhole imaging is a particularly challenging unknown-view tomography problem as it reconstructs an image of a 3D object from unregistered 1D measurements. Nevertheless, we found Bayesian algorithms could be adapted to solve this problem. To
overcome their high computational cost we took advantage of GPU computing.

III. KEYHOLE IMAGING

The keyhole imaging setup is illustrated in Fig. 1. A confocal pulsed light source–detector pair, placed at a standoff distance from the door, illuminates and images a visible point through a small aperture (such as a keyhole). The time-resolved measurements, captured with a single-photon avalanche diode (SPAD) or other detector, contain the temporal response of the emitted light pulse that travels to a visible point through the keyhole, and scatters back from a hidden object. A series of measurements, captured as the hidden object undergoes rigid motion, are used to recover the shape and motion of the hidden object.

In this section, we describe the keyhole measurement model in detail and derive a Bayesian estimation method for recovering the shape and motion of hidden objects.

A. Observation Model

Assume that the hidden object is defined by a volumetric albedo \( \rho(x') \), where \( x' = [x', y', z']^T \) is the object’s local coordinate system. Assuming the object exhibits only rigid body motion, a transform \( \Theta_l = [R_l | t_l] \in \mathbb{R}^{3 \times 4} \) describes the motion of the object at some time \( l \) and transforms it into the global coordinate system, using rotation and translation, as \( x = R_l x' + t_l \). Each time-resolved measurement \( y_l \in \mathbb{R}^T \), represented by a histogram with \( T \) bins counting the number of photon-arrival events within a certain time window, is defined by a formation model

\[
y_l = f(\rho, \Theta_l) + \eta_l, \tag{1}
\]

where \( \eta_l \) denotes (potentially signal dependent) noise and \( f(\rho, \Theta_l) \) is the NLOS imaging forward model

\[
f(\rho, \Theta_l) = \int \frac{1}{g(x)} \rho(x') \delta(2\|x\|_2 - \tau c) \, dx'. \tag{2}
\]

Here, \( c \) is the speed of light, \( \tau \) is the time relative to an emitted laser pulse, the point at which we record the scene is located in the origin of the global coordinate system, and \( g(x) \) describes the falloff of light with distance. For example, hidden objects with Lambertian reflectance exhibit a falloff of \( g_{\text{diffuse}}(x) = \|x\|_2^2 \) whereas perfectly retroreflective objects experience \( g_{\text{retro}}(x) = \|x\|_2^2 \) [4]. Moreover, \( g \) can also include angle-dependent factors that are influenced by the surface normals, the normal of the visible wall, or other angle-dependent reflectance characteristics. Using experimental measurements, we found that \( g_{\text{exp}}(x) = \|x\|_2^2 \cos^{-4}(\phi) \), where \( \phi \) is the angle between the wall’s normal and \( x \), best modeled the falloff associated with a patch, oriented parallel to the visible wall, of the store-bought retroreflective tape that we used in our captured results. See the supplement for more information about the origin of a quartic falloff models with retroreflective objects, how we fit the model to our data, and the effects of model mismatch on our reconstructions.

Note that (2) is the standard confocal NLOS model [4], with the hidden object’s position transformed by \( \Theta_l \). This formulation does not model self-occlusion and in this work we assume the hidden object does not self-occlude. The keyhole imaging reconstruction problem uses a series of measurements \( y_1, \ldots, y_L \) to reconstruct the albedo \( \rho \) in its local coordinate system. This problem is challenging because the measurements are parameterized by latent hidden object locations, \( \Theta_1, \ldots, \Theta_L \), corresponding to each captured measurement. (In practice, we parameterize each measurement by the position of an isotropic virtual sensor during each measurement. This parameterization allows us to represent rotation and translation of the hidden object as translation of the virtual sensor.)

B. Reconstruction With Expectation-Maximization

We seek to recover an estimate of the unknown albedo, \( \rho \), from the observations \( y_1, \ldots, y_L \). This can be done by maximizing the log likelihood of the observed measurements

\[
\mathcal{L}(\rho; y) = \log p(y | \rho) = \log \int_\Theta p(y, \Theta | \rho) \, d\Theta, \tag{3}
\]

where \( y = [y_1, \ldots, y_L] \) and \( \Theta = [\Theta_1, \ldots, \Theta_L] \).

Maximizing the objective (3) directly is numerically unstable; without an excellent initialization, \( p(y | \rho) \) will start near 0 and one is tasked with maximizing the log of 0. Instead, in this work we utilize EM, which efficiently maximizes (3) by solving a series of easy-to-solve least squares problems. In particular, EM iteratively applies two steps: (1) an expectation step in which an estimated conditional distribution of \( \Theta \) is used to form a lower bound to the log likelihood, and (2) a maximization step, which estimates the albedo given the current conditional distribution of \( \Theta \).

EM Algorithm for Keyhole Imaging

Given an estimate \( \rho^{(n)} \) of the hidden object’s albedo at iteration \( n \) of the EM algorithm, we perform the expectation step by finding a lower bound on the log of the likelihood given by (3). Using Jensen’s inequality, it can be shown that a lower bound is given as \( Q(\rho, \rho^{(n)}) \) [55], where, up to additive constants,

\[
Q(\rho, \rho^{(n)}) = \mathbb{E}_{\Theta | y, \rho^{(n)}}[\log p(y, \Theta | \rho)],
\]

\[
= \int_\Theta p(\Theta | y, \rho^{(n)}) \log[p(y | \Theta, \rho)p(\Theta | \rho)] \, d\Theta,
\]

\[
= \sum_{l=1}^L \int_{\Theta_l} p(\Theta_l | y_l, \rho^{(n)}) \log[p(y_l | \Theta_l, \rho)p(\Theta_l | \rho)] \, d\Theta_l,
\]

\[
\approx \sum_{l=1}^L \sum_{\Theta_{l,k} \in \Omega} p(\Theta_{l,k} | y_l, \rho^{(n)}) \log[p(y_l | \Theta_{l,k}, \rho)p(\Theta_{l,k} | \rho)],
\]

\[
= \sum_{l=1}^L \sum_{\Theta_{l,k} \in \Omega} p(\Theta_{l,k} | y_l, \rho^{(n)}) \log[p(y_l | \Theta_{l,k}, \rho)]. \tag{4}
\]

Here, \( \Omega \) is the domain of possible hidden object positions (parameterized as translations of a virtual sensor); \( p(\Theta_{l,k} | y_l, \rho^{(n)}) \) is the probability the object is at position \( \Theta_{l,k} \) during the \( l^{th} \).
Algorithm 1: EM for Keyhole Imaging.

1: Initialize: $\rho^{(0)}$
2: for $n=0, 1, \ldots, N-1$ do
3: \hspace{1em} Compute $w_{\theta_{l,k}} = \exp\left(-\frac{1}{2\sigma^2\rho} \left\| y_l - f(\rho^{(n)}, \theta_{l,k}) \right\|^2\right)$ \forall $i, k$.
4: \hspace{1em} $\rho^{(n+1)} = \arg \max_{\rho} Q(\rho, \rho^{(n)})$
5: \hspace{1em} Return $\rho^{(N)}$

Adding Priors: EM can be extended to computing MAP estimates by redefining $Q(\rho, \rho^{(n)})$ as

$$Q(\rho, \rho^{(n)}) = \left( \sum_{l=1}^{L} \sum_{\theta_{l,k} \in \Omega} -w_{\theta_{l,k}} \left\| y_l - f(\rho, \theta_{l,k}) \right\|^2 \right) + \lambda \log p(\rho),$$

where $p(\rho)$ is a prior on $\rho$. In this work, we use $\log p(\rho) = -\|L\rho\|_1 - \|\rho\|_1$, where $L$ is a Laplacian filter. This expression corresponds to a prior that the hidden object’s albedo is smooth and sparse.

Implementation Details: We developed an implementation of EM using the GPU-accelerated and free-to-use PyTorch library [56]. We ran EM for 30 iterations. Each maximization step was accomplished by running the ADAM optimizer [57] $n + 1$ times, where $n$ is the EM iteration number. During the early iterations of the algorithm, when our estimate of $w$ is unreliable, we do not spend much time optimizing $\rho$. We set ADAM’s learning rate to 0.1 and set its $\beta$ values, which control the momentums’ decay, to 0.5 and 0.999. We set the regularization parameter $\lambda$ to 2000 for the experimental data and 0 for the simulated data. The noise variance $\sigma$ was set to 200 for both simulated and experimental data; this value is far larger than the true noise variance, but helped account for model mismatch and the discretization of the trajectory support set $\Omega$.

During simulations, $\Omega$ was a $33 \times 33$ equispaced grid spanning 1 m by 1 m. For experiments, $\Omega$ was a $33 \times 33$ grid spanning 1 m along the axis parallel to the wall and 15 cm along the axis perpendicular to the wall; this was the range of our translation stages. We enforced positivity on $\rho$ by parameterizing it with the variable $\nu$, with $\rho := \nu^2$ where the square is taken elementwise. The variable $\nu$ was initialized with an i.i.d. Gaussian vector with mean 0 and variance 1, which is roughly the same as the variance of $\nu$ after the algorithm has converged.

We found that deterministic annealing helped EM avoid local minima [58]. With deterministic annealing

$$w_{\theta_{l,k}} = \left[ \exp\left(-\frac{\left\| y_l - f(\rho^{(n)}, \theta_{l,k}) \right\|^2}{2\sigma^2}\right) \right]^{\beta},$$

where $\beta \in (0, 1]$ is an inverse temperature parameter that increases iteration to iteration. We set $\beta^{(n+1)} = 1.3 \cdot \beta^{(n)}$, with $\beta^{(0)}$ set such that $\beta^{(N-1)} = 1$. Annealing serves to make the distribution of the estimated object locations more uniform during the earlier iterations of the algorithm, when the algorithm has a less accurate estimate of $\rho$. Accordingly, as demonstrated in the supplement, with annealing EM is insensitive to how it is initialized.

IV. EVALUATION AND ANALYSIS

A. Simulation Setup

We first investigate the keyhole imaging problem in simulation. The keyhole measurements are simulated from nine different binary objects, drawn from the HaSyV2 dataset [59], which are illustrated in the top row of Fig. 3. Each of these objects has a resolution of $64 \times 64$ and is 50 cm tall and wide in the simulator. Our simulated SPAD measurements have a temporal resolution of 16 ps. We apply Poisson noise to the measurements.
such that they have the desired signal-to-noise ratio (SNR) for each test.

For each of the nine objects, we simulate measurements from nine distinct trajectories of varying lengths. The virtual sensor locations corresponding to each of these trajectories is shown in Fig. 2. Recall measurements of a moving object with a fixed sensor location are equivalent to measurements of a fixed object with a moving virtual sensor. Three of these trajectories (left column) have the virtual sensor locations restricted to a constant $z$ plane; this corresponds to object roll and translation up-and-down and side-to-side. Another three of these trajectories (middle column) have the virtual sensor locations restricted to a constant $x$ plane; this corresponds to object pitch and translation up-and-down and forward-and-backward. The last three of these trajectories (right column) have the virtual sensor locations restricted to a constant $y$ plane; this corresponds to object yaw and translation side-to-side and forward-and-backward. The horizontal motion in the last of these trajectories is arguably the most realistic; for instance, this is the type of motion exhibited by cars, rolling chairs, and shopping carts. The last row of Fig. 2 presents the $33 \times 33$ grids $\Omega$ of allowable virtual sensor locations that we use with EM.

Because the EM algorithm does not know the object trajectories beforehand, the algorithm’s reconstructions have certain ambiguities/invariances. Constant $z$ trajectories are rotation invariant: given a reconstructed trajectory and an object that fits the measurements, one could rotate both clockwise without changing the fit. Similarly, constant $x$ trajectories are invariant to vertical flips of the trajectory and object and constant $y$ trajectories, which are what we test with real data in the next section, are invariant to horizontal flips.

When considered as object/trajectory pairs, none of the reconstructions are invariant to translations. However, an object/trajectory pair can be expressed as an equivalent pair where the object is translated one direction and the trajectory is translated equally in the opposite direction. Thus, in comparing our reconstructed objects to the ground truth, we need to compare across translations as well.

### B. Simulation Results

In order to form a performance baseline, we first reconstruct the hidden objects assuming their trajectories are known. This is accomplished by maximizing the log-likelihood

$$
\max_{\rho} \left( \sum_{l=1}^{L} -\|y_l - f(\rho, \theta_l)\|_2^2 + \lambda \log p(\rho) \right),
$$

where the objects’ locations over time, $\theta_1, \ldots, \theta_L$, are given. We use 200 iterations of gradient descent (GD), along-side the ADAM optimizer, to maximize (9). As in the EM case, the non-negative albedo $\rho$ was parameterized as $\rho = \nu^2$ and the variable $\nu$ was initialized with an i.i.d. Gaussian vector with mean 0 and variance 1.

We compare EM, which does not know the trajectories, and GD, which does, qualitatively in Fig. 3 and quantitatively in Table I. Table I reports reconstruction accuracy in terms of disambiguated Structural Similarity (SSIM) [60], which we define as

$$
\max_{rtf \in rtf} \text{SSIM}(\rho, rtf(\hat{\rho})),
$$

such that they have the desired signal-to-noise ratio (SNR) for each test.

For each of the nine objects, we simulate measurements from nine distinct trajectories of varying lengths. The virtual sensor locations corresponding to each of these trajectories is shown in Fig. 2. Recall measurements of a moving object with a fixed sensor location are equivalent to measurements of a fixed object with a moving virtual sensor. Three of these trajectories (left column) have the virtual sensor locations restricted to a constant $z$ plane; this corresponds to object roll and translation up-and-down and side-to-side. Another three of these trajectories (middle column) have the virtual sensor locations restricted to a constant $x$ plane; this corresponds to object pitch and translation up-and-down and forward-and-backward. The last three of these trajectories (right column) have the virtual sensor locations restricted to a constant $y$ plane; this corresponds to object yaw and translation side-to-side and forward-and-backward. The horizontal motion in the last of these trajectories is arguably the most realistic; for instance, this is the type of motion exhibited by cars, rolling chairs, and shopping carts. The last row of Fig. 2 presents the $33 \times 33$ grids $\Omega$ of allowable virtual sensor locations that we use with EM.

Because the EM algorithm does not know the object trajectories beforehand, the algorithm’s reconstructions have certain ambiguities/invariances. Constant $z$ trajectories are rotation invariant: given a reconstructed trajectory and an object that fits the measurements, one could rotate both clockwise without changing the fit. Similarly, constant $x$ trajectories are invariant to vertical flips of the trajectory and object and constant $y$ trajectories, which are what we test with real data in the next section, are invariant to horizontal flips.

When considered as object/trajectory pairs, none of the reconstructions are invariant to translations. However, an object/trajectory pair can be expressed as an equivalent pair where the object is translated one direction and the trajectory is translated equally in the opposite direction. Thus, in comparing our reconstructed objects to the ground truth, we need to compare across translations as well.

### B. Simulation Results

In order to form a performance baseline, we first reconstruct the hidden objects assuming their trajectories are known. This is accomplished by maximizing the log-likelihood

$$
\max_{\rho} \left( \sum_{l=1}^{L} -\|y_l - f(\rho, \theta_l)\|_2^2 + \lambda \log p(\rho) \right),
$$

where the objects’ locations over time, $\theta_1, \ldots, \theta_L$, are given. We use 200 iterations of gradient descent (GD), along-side the ADAM optimizer, to maximize (9). As in the EM case, the non-negative albedo $\rho$ was parameterized as $\rho = \nu^2$ and the variable $\nu$ was initialized with an i.i.d. Gaussian vector with mean 0 and variance 1.

We compare EM, which does not know the trajectories, and GD, which does, qualitatively in Fig. 3 and quantitatively in Table I. Table I reports reconstruction accuracy in terms of disambiguated Structural Similarity (SSIM) [60], which we define as

$$
\max_{rtf \in rtf} \text{SSIM}(\rho, rtf(\hat{\rho})),
$$

such that they have the desired signal-to-noise ratio (SNR) for each test.

For each of the nine objects, we simulate measurements from nine distinct trajectories of varying lengths. The virtual sensor locations corresponding to each of these trajectories is shown in Fig. 2. Recall measurements of a moving object with a fixed sensor location are equivalent to measurements of a fixed object with a moving virtual sensor. Three of these trajectories (left column) have the virtual sensor locations restricted to a constant $z$ plane; this corresponds to object roll and translation up-and-down and side-to-side. Another three of these trajectories (middle column) have the virtual sensor locations restricted to a constant $x$ plane; this corresponds to object pitch and translation up-and-down and forward-and-backward. The last three of these trajectories (right column) have the virtual sensor locations restricted to a constant $y$ plane; this corresponds to object yaw and translation side-to-side and forward-and-backward. The horizontal motion in the last of these trajectories is arguably the most realistic; for instance, this is the type of motion exhibited by cars, rolling chairs, and shopping carts. The last row of Fig. 2 presents the $33 \times 33$ grids $\Omega$ of allowable virtual sensor locations that we use with EM.

Because the EM algorithm does not know the object trajectories beforehand, the algorithm’s reconstructions have certain ambiguities/invariances. Constant $z$ trajectories are rotation invariant: given a reconstructed trajectory and an object that fits the measurements, one could rotate both clockwise without changing the fit. Similarly, constant $x$ trajectories are invariant to vertical flips of the trajectory and object and constant $y$ trajectories, which are what we test with real data in the next section, are invariant to horizontal flips.

When considered as object/trajectory pairs, none of the reconstructions are invariant to translations. However, an object/trajectory pair can be expressed as an equivalent pair where the object is translated one direction and the trajectory is translated equally in the opposite direction. Thus, in comparing our reconstructed objects to the ground truth, we need to compare across translations as well.

### B. Simulation Results

In order to form a performance baseline, we first reconstruct the hidden objects assuming their trajectories are known. This is accomplished by maximizing the log-likelihood

$$
\max_{\rho} \left( \sum_{l=1}^{L} -\|y_l - f(\rho, \theta_l)\|_2^2 + \lambda \log p(\rho) \right),
$$

where the objects’ locations over time, $\theta_1, \ldots, \theta_L$, are given. We use 200 iterations of gradient descent (GD), along-side the ADAM optimizer, to maximize (9). As in the EM case, the non-negative albedo $\rho$ was parameterized as $\rho = \nu^2$ and the variable $\nu$ was initialized with an i.i.d. Gaussian vector with mean 0 and variance 1.

We compare EM, which does not know the trajectories, and GD, which does, qualitatively in Fig. 3 and quantitatively in Table I. Table I reports reconstruction accuracy in terms of disambiguated Structural Similarity (SSIM) [60], which we define as

$$
\max_{rtf \in rtf} \text{SSIM}(\rho, rtf(\hat{\rho})),
$$

such that they have the desired signal-to-noise ratio (SNR) for each test.

For each of the nine objects, we simulate measurements from nine distinct trajectories of varying lengths. The virtual sensor locations corresponding to each of these trajectories is shown in Fig. 2. Recall measurements of a moving object with a fixed sensor location are equivalent to measurements of a fixed object with a moving virtual sensor. Three of these trajectories (left column) have the virtual sensor locations restricted to a constant $z$ plane; this corresponds to object roll and translation up-and-down and side-to-side. Another three of these trajectories (middle column) have the virtual sensor locations restricted to a constant $x$ plane; this corresponds to object pitch and translation up-and-down and forward-and-backward. The last three of these trajectories (right column) have the virtual sensor locations restricted to a constant $y$ plane; this corresponds to object yaw and translation side-to-side and forward-and-backward. The horizontal motion in the last of these trajectories is arguably the most realistic; for instance, this is the type of motion exhibited by cars, rolling chairs, and shopping carts. The last row of Fig. 2 presents the $33 \times 33$ grids $\Omega$ of allowable virtual sensor locations that we use with EM.
where \( \hat{\rho} \) denotes the reconstructed albedo and RTF denotes the set of all possible rotations, translations, and flips of the object. We test over all rotations in \( 5^\circ \) increments and all translations in 1 pixel increments.

Fig. 3 compares reconstructions between GD and EM when the object follows trajectory (f) from Fig. 2 and the measurements have an SNR of 15. The results show EM performs nearly as well as GD. However, because there is a flip ambiguity, that is an object moving left would produce the same measurements as a flipped version of an object moving right, EM occasionally finds two mirrored objects/trajectories and it struggles to decide between the two. Scenarios like these (which are more common at lower resolutions) result in artifacts like those demonstrated for objects 6 and 8 in Fig. 3, where we can see the traces of two flipped versions of the object overlaid on top of one another.

Table I demonstrates EM performs only a little worse on average than GD across a variety of SNRs and trajectories. Additional simulation results, including rotation-only trajectories and 3D trajectories, are provided in Appendix B.

V. EXPERIMENTAL VALIDATION

A. Experimental Setup

Our prototype system is illustrated in Fig. 4. The optical setup consists of a 670 nm pulsed laser source (ALPHALAS PICOPower-LD-670-50) operating with a 10 MHz repetition rate, an average power of approximately 0.1 mW, and a pulse width of 30 ps. This laser is in a confocal configuration with a fast-gated single-pixel SPAD detector (Micro Photon Devices PDM series SPAD, 50 \( \mu \)m \( \times \) 50 \( \mu \)m active area), which allows us to gate out direct bounce photons, capturing only indirect photons from the hidden object. A time-correlated single photon counter (PicoQuart PicoHarp 300) takes as input a trigger signal from the laser and photon detection event triggers from the SPAD and forms time-stamped histograms of photon arrival times with 16ps bin widths.

We captured keyhole measurements of the objects illustrated in the top row of Fig. 5. The objects are each 50 cm tall and covered in retroreflective tape. The objects were affixed to two Zaber translation stages: a 1m linear stage (X-BLQ1045-E01) and a 15cm stage (T-LSR150 A). Using the stages, objects can be translated to any point within a 1m by 15cm horizontal plane, where the long axis (\( x \)-axis) is aligned parallel to the wall and the short axis (\( z \)-axis) is aligned perpendicular to the wall such that at their closest point the objects are .62cm from the wall. In the captured measurements, objects are translated to between 66 and 165 discrete locations. To capture adequate signal given the limited laser power of our prototype, the objects hold position at each location for 10 seconds. Example trajectories are shown in the last column of Fig. 5.

B. Experimental Results

As in the previous section, in order to form a performance baseline, we first reconstruct the hidden objects using GD assuming their trajectories were known. The resulting reconstructions are presented in the second row of Fig. 5. The reconstructions
illustrate that \((x, z)\) translations contain the measurement diversity necessary to perform NLOS reconstructions with real data. The mannequin’s pose is recognizable and the letters are all readable. The mannequin’s reconstruction is likely blurrier than the letters’ because the mannequin is 3D while the model fit to the data is 2D.

For the case where object trajectories are not known, we use EM to recover both the shape and trajectory of the hidden object. For each measurement, EM also produces an estimate of the probability distribution of each object’s location. We select the highest probability location as the location estimate. Note that we found that by the time our algorithm has converged, the posterior probability estimates of the object locations are essentially one hot. That is, for each instance in time \(l\), \(w_{y|0}\) assigns a probability near one that the object is at a certain position and a probability near 0 that it is at any other position. (During the early iterations of the algorithm, when it does not have an accurate estimate for the shape of the object, the assigned probabilities are nearly uniform.)

The reconstructed objects and trajectories are presented in the third and fourth rows of Fig. 5. Again the mannequin’s pose is recognizable and the letters are all readable. Moreover, EM’s estimated trajectories closely match the ground truth and in all four examples the general path of the object is clearly visible.

The EM reconstruction of the “K” serves to illustrate the translation ambiguity that was discussed in the previous section. Relative to the true trajectory, the reconstructed trajectory is displaced by a positive shift in the \(x\) direction. Meanwhile, the object reconstruction is offset by a negative shift in the \(x\) direction. Considered jointly, the object’s location is recovered accurately. That is, when shifted to the right \((+x)\) to undo the error in its trajectory, the EM “K” in row three of Fig. 5 overlaps with the GD “K” in row two.

The EM reconstructions were computed at a 256×256 resolution in five and a half minutes using an Nvidia Titan RTX GPU and a six core Intel CPU. The known-trajectory GD reconstructions took fifteen seconds on the same hardware.

Code and data to reproduce these results is available at https://github.com/computational-imaging/KeyholeImaging.

VI. DISCUSSION

This work proposes, develops, and experimentally validates keyhole imaging—a new technique to reconstruct the shape and location of a hidden object from NLOS measurement captured along a single optical path. Whereas in standard NLOS imaging one scans multiple points across a relay surface to form a synthetic aperture, in keyhole imaging one scans a single point and the trajectory of the object determines the size and shape of the synthetic aperture. This distinction has important implications.

Resolution Limits: Keyhole imaging performance is upper-bounded by the performance of conventional NLOS system with an equivalent synthetic aperture. Our simulation results from the previous section indicate this bound is fairly tight. The larger distinction between keyhole and conventional NLOS is that (with planar, horizontal object motion) keyhole’s synthetic aperture is a horizontal surface rather than vertical wall. This change results in significantly different resolution limits.

In the appendix, following the analysis conducted in [4], we derive resolution limits associated with the \(\omega_x \times \omega_z\) horizontal synthetic aperture illustrated in Fig. 6. In particular, we assume that the detector has Gaussian jitter with standard deviation \(\sigma_{SPAD}\) and derive lower bounds on the resolutions in the \(x\), \(y\), and \(z\) directions, defined simply as the minimum spacing at
which two nearby points would produce distinguishable measurements. Our analysis shows that

$$\Delta x \geq \frac{c}{2} \sqrt{\frac{y^2 + z^2}{(|x| + \omega_x/2)^2}} + 1, \quad (11)$$

$$\Delta y \geq \frac{c}{2} \sqrt{\frac{z^2}{\omega_x^2}} + 1, \quad (12)$$

$$\Delta z \geq \frac{c}{2} \sqrt{\frac{y^2}{(\omega_x + z)^2}} + 1, \quad (13)$$

where $c$ is the speed of light and $\gamma = 2\sqrt{2\ln 2}\sigma_{SPAD}$ is the full-width at half maximum of the temporal jitter. The $x$, $y$, and $z$ coordinates are given assuming the origin is at the center of the synthetic aperture with respect to the $x$ direction and on the edge nearest the object with respect to the $z$ direction. See Figures 6 and 2(l) for reference.

These results imply that the resolution in any dimension is lower bounded by $\frac{c}{2}$ and that the resolution in the $x$ and $z$ dimensions improves as the synthetic aperture gets larger with respect to $x$ and $z$ respectively. That is, the resolution in $x$ and $z$ improves as the object translates more in the $x$ and $z$ directions. The above results also imply that near the height of sampling point on the wall, that is as $y$ approaches 0, the system has very limited resolution in the vertical ($\Delta y$) direction; the bound described by (12) approaches infinity.

**Scanning Through a Small Aperture:** Imaging through a small aperture/keyhole causes a reduction in light throughput, as compared to standard confocal NLOS imaging. The size of this reduction is determined by the size of the keyhole as well as the distances between the detector, the keyhole, and the relay surface. In a configuration like ours, the keyhole is significantly larger than both the laser beam width and the SPAD’s focus size on the wall. In this context, the keyhole only serves to restrict the maximum usable aperture size of the detector. If the detector is $d_1$ away from the keyhole and the keyhole is $d_2$ away from the wall then a $\Gamma$ diameter circular keyhole restricts the usable aperture diameter to $A = \frac{d_1^2 + d_2}{d_2} \Gamma$. With $d_1 = 50\text{cm}$, $d_2 = 1\text{m}$, and $\Gamma = 5\text{mm}$ this works out to 7.5 $\text{mm}$, which corresponds to a minimum effective f-number of 6.6 for our 50 $\text{mm}$ lens.

**Laser Power:** In this work we imaged retroreflective objects whose trajectories’ total exposure times were one to three minutes. However, as observed in [4], at even 1 $\text{m}$ standoff distances diffuse objects return 100× less light than high-quality retroreflective hidden objects and this discrepancy grows quadratically with standoff distance. Large-scale real-time keyhole imaging of non-retroreflective objects is possible with more powerful lasers, such as those used in [6], [7], which are 10,000× brighter than ours. However, doing so with eye-safe lasers will require operating at lower SNRs than our current results. See the second table in the supplement for predicted photons counts under various operating conditions.

**Future Work:** Several avenues exist to improve the proposed EM-based reconstruction scheme. First, the present algorithm ignores the fact that motion is generally continuous; that is it treats a motion trajectory that bounces back and forth across the room just a likely as a smooth straight path. Situations like this can result in artifacts as two mirrored versions of the object are reconstructed on top of one another, as described in Section IV-B. To address this problem, motion-continuity and other priors on $\theta$, which fit naturally into the EM framework, could be incorporated. Second, while our current method imposes smoothness and sparsity on the reconstruction, more advanced priors, particularly learned priors, could greatly aid in the reconstruction. Third, our present results restrict the trajectories to one of three planes. While our implementation supports three dimensional trajectories (a very simple 3D trajectory is tested in Appendix B), they also require significantly more computation. Developing more efficient algorithms and implementations, perhaps by restricting the search space using motion continuity priors or by using neural network based priors [61], could make our method more generalizable. Finally, extending our work to handle non-rigid body motion and self-occlusion are important open problems.

**VII. CONCLUSION**

NLOS imaging has emerged as an important research direction in the computational imaging and optics communities and is widely recognized for enabling capabilities that would have been impossible only a few years ago. While seeing around corners has long required imaging a large visible surface, we demonstrate imaging and tracking using a single visible point by exploiting object motion. Moreover, combining time-of-flight imaging with object motion may be useful for other 3D imaging applications. We envision that keyhole imaging could unlock new applications, such as NLOS imaging in constrained and cluttered environments.

**APPENDIX A**

**Resolution Analysis**

We now derive the resolution limits associated with a $\omega_x$ by $\omega_z$ horizontal synthetic aperture whose corners are at the points $(-\omega_x/2, 0, -\omega_z), (\omega_x/2, 0, -\omega_z), (\omega_x/2, 0, 0)$, and $(-\omega_x/2, 0, 0)$. Following the analysis conducted in [4], we determine how close to each other two points can get in the $x$, $y$, and $z$ directions before they produce indistinguishable measurements. We say two sets of measurements are indistinguishable if they are separated in time by less that the full width at half maximum, $\gamma = 2\sqrt{2\ln 2}\sigma_{SPAD}$, of the temporal jitter of the system. This occurs if there is no sampling point $p$ within the synthetic aperture for which the distance between $p$ and the two points in the scene, $q_1$ and $q_2$, is greater than or equal to $c/2$, where $c$ is the speed of light.

**A. Resolution in $X$**

We first need to select a point on the synthetic aperture for which two points $q_1 = (x - \Delta x/2, y, z)$ and $q_2 = (x + \Delta x/2, y, z)$ would produce the most easily distinguishable measurements. This is the point $p$ for which the absolute difference...
between \(\|p - q_1\|\) and \(\|p - q_2\|\) is maximized. Assuming \(\Delta x\) is small, this point can be found by selecting the point \(p\) for which the inner product between the vectors \(p - q_1\) and \(\frac{y - q_1}{\|y - q_1\|}\) is maximized. That is, by selecting the point for which \(p - q_1\) is best aligned with \(q_1 - q_2\). Assuming \(z \geq 0\), this point is \(p = (-\text{sign}(x)\omega_x/2, 0, 0)\). See the left side of Fig. 7.

The resulting distances between \(p\) and \(q_1\) and \(q_2\) are described by

\[
\min(\|q_1 - p\|, \|q_2 - p\|) = \sqrt{(\omega_x/2 + |x| - \Delta x/2)^2 + y^2 + z^2}
\]

and

\[
\max(\|q_1 - p\|, \|q_2 - p\|) = \sqrt{(\omega_x/2 + |x| + \Delta x/2)^2 + y^2 + z^2}.
\]

The points \(q_1\) and \(q_2\) produce distinguishable measurements when \(\|q_1 - p\| - \|q_2 - p\| \geq c\gamma/2\). This implies that \(q_1\) and \(q_2\) are resolvable when

\[
\Delta x \geq \frac{c\gamma \sqrt{c^2 \gamma^2 - 4 \omega_x^2 - 16 \omega_x |x| - 16|x|^2 - 16y^2 - 16z^2}}{\sqrt{4c^2 \gamma^2 - 64 \omega_x^2 - 64 \omega_x |x| - 64|x|^2}}.
\]

If we assume \(c^2 \gamma^2 \approx 0\), the above simplifies to

\[
\Delta x \geq \frac{c\gamma}{2} \sqrt{\frac{y^2 + z^2}{|x| + \omega_x^2}} + 1.
\]  \hspace{1cm} (14)

B. Resolution in \(Y\)

As before, we first need to select a point on the synthetic aperture for which two points \(q_1 = (x, y, z - \Delta z/2)\) and \(q_2 = (x, y, z + \Delta z/2)\) would produce the most easily distinguishable measurements. Assuming \(z \geq 0\) and \(|x| \leq \omega_x/2\), this point is \(p = (x, 0, -\omega_x)\). See the middle of Fig. 7.

The resulting distances between \(p\) and \(q_1\) and \(q_2\) are

\[
\|q_1 - p\| = \sqrt{(y - \Delta y/2)^2 + z^2}
\]

and

\[
\|q_2 - p\| = \sqrt{(y + \Delta y/2)^2 + z^2}.
\]

The points \(q_1\) and \(q_2\) produce distinguishable measurements when \(\|q_1 - p\| - \|q_2 - p\| \geq c\gamma/2\). This implies that \(q_1\) and \(q_2\) are resolvable when

\[
\Delta y \geq \frac{c\gamma \sqrt{c^2 \gamma^2 - 16y^2 - 16z^2}}{\sqrt{4c^2 \gamma^2 - 64y^2}}.
\]

If we assume \(c^2 \gamma^2 \approx 0\), the above simplifies to

\[
\Delta y \geq \frac{c\gamma}{2} \sqrt{\frac{z^2}{y^2}} + 1.
\]  \hspace{1cm} (15)

C. Resolution in \(z\)

As before, we first need to select a point on the synthetic aperture for which two points \(q_1 = (x, y, z - \Delta z/2)\) and \(q_2 = (x, y, z + \Delta z/2)\) would produce the most easily distinguishable measurements. Assuming \(z \geq 0\) and \(|x| \leq \omega_x/2\), this point is \(p = (x, 0, -\omega_x)\). See the middle of Fig. 7.

The resulting distances between \(p\) and \(q_1\) and \(q_2\) are

\[
\|q_1 - p\| = \sqrt{y^2 + (\omega_x + z - \Delta z/2)^2}
\]

and

\[
\|q_2 - p\| = \sqrt{y^2 + (\omega_x + z + \Delta z/2)^2}.
\]

The points \(q_1\) and \(q_2\) produce distinguishable measurements when \(\|q_1 - p\| - \|q_2 - p\| \geq c\gamma/2\). This implies that \(q_1\) and \(q_2\) are resolvable when

\[
\Delta z \geq \frac{c\gamma \sqrt{c^2 \gamma^2 - 16 \omega_x^2 - 32 \omega_x z - 16y^2 - 16z^2}}{\sqrt{4c^2 \gamma^2 - 64 \omega_x^2 - 128 \omega_x z - 64z^2}}.
\]

If we assume \(c^2 \gamma^2 \approx 0\), the above simplifies to

\[
\Delta z \geq \frac{c\gamma}{2} \sqrt{\frac{y^2}{(\omega_x + z)^2}} + 1.
\]  \hspace{1cm} (16)
We simulated reconstructing objects which followed the rotation-only trajectory on the left and the densely sampled 3D trajectory on the right. For the 3D result, the object moved to every point on the grid.

![Fig. 8](image1)

**Fig. 8. Rotation-only and 3D trajectory simulation setup.** We simulated reconstructing objects which followed the rotation-only trajectory on the left and the densely sampled 3D trajectory on the right. For the 3D result, the object moved to every point on the grid.

![Fig. 9](image2)

**Fig. 9. Rotation-only and 3D trajectory simulation results.** Reconstructions of the star object with rotation-only and 3D trajectories from measurements with an SNR of 5. These trajectories are more challenging than the planar, translation-based trajectories from Section IV.

**APPENDIX B**

**ADDITIONAL SIMULATION RESULTS**

In this section we test EM on the more challenging rotation-only and densely-sampled 3D trajectories. Because, under our forward model, the measurements formed by rotating a hidden object are equivalent to the measurements formed by moving the virtual sensor along a circular trajectory around the axis of rotation, we form and represent the rotation-only trajectory using the virtual sensor trajectory illustrated in Fig. 8(a). Our 3D trajectory is illustrated in Fig. 8(b). The rotation-only trajectory uses the $33 \times 33$ sampling grid illustrated in Fig. 2(1). In order for EM to fit in the GPU’s memory, the 3D sampling grid is restricted to $10 \times 10 \times 10$ points; these are the same points that are densely sampled in Fig. 8(b).

We tested these trajectories first under the same conditions as described in Section IV-A and then again where we simplified the problem by reducing the size of the simulated images from $50 \text{ cm} \times 50 \text{ cm}$ to $15 \text{ cm} \times 15 \text{ cm}$. (We kept the resolution at $64 \times 64$.) The smaller objects make it easier for the algorithm to distinguish which measurement belongs to which point on the grid. As Table II illustrates, EM struggled to accurately reconstruct the hidden objects when they are $50 \text{ cm} \times 50 \text{ cm}$, but is, on average, roughly as successful as it was with the previous trajectories when the hidden objects are restricted to $15 \text{ cm} \times 15 \text{ cm}$.

**TABLE II**

|                | $SNR = 5$ | $SNR = 15$ | $SNR = 50$ |
|----------------|-----------|------------|------------|
| **Rotation-only Traj.** |           |            |            |
| 50 cm $\times$ 50 cm | 0.54      | 0.43       | 0.66       | 0.46 | 0.77 | 0.46 |
| 3D Traj.        | 0.60      | 0.43       | 0.71       | 0.44 | 0.84 | 0.44 |
| **Rotation-only Traj.** |           |            |            |
| 15 cm $\times$ 15 cm | 0.51      | 0.52       | 0.60       | 0.53 | 0.68 | 0.54 |
| 3D Traj.        | 0.60      | 0.54       | 0.68       | 0.55 | 0.75 | 0.55 |

**ACKNOWLEDGMENT**

We would like to thank the reviewers for their feedback and help improving this manuscript. We would also like to thank Prasanna Rangarajan for help with the radiometry analysis.

**REFERENCES**

[1] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar, “Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging,” *Nat. Commun.*, vol. 3, p. 745, 2012.

[2] M. Buttavasa, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, “Non-line-of-sight imaging using a time-gated single photon avalanche diode,” *Opt. Express*, vol. 23, no. 16, pp. 20 997–21 011, 2015.

[3] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio, “Detection and tracking of moving objects hidden from view,” *Nat. Photon.*, vol. 10, pp. 23–26, 2016.

[4] M. O’Toole, D. B. Lindell, and G. Wetzstein, “Confocal non-line-of-sight imaging based on the light cone transform,” *Nat.*, pp. 338–341, 2018.

[5] S. Xin, S. Nousias, K. N. Kutulakos, A. C. Sankaranarayanan, S. G. Narasimhan, and I. Gkioulakas, “A theory of fermat paths for non-line-of-sight shape reconstruction,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6800–6809.

[6] X. Liu et al., “Non-line-of-sight imaging using phasor-field virtual wave optics,” *Nat.*, vol. 572, pp. 620–623, 2019.

[7] D. B. Lindell, G. Wetzstein, and M. O’Toole, “Wave-based non-line-of-sight imaging using fast f-k migration,” *ACM Trans. Graph. (SIGGRAPH)*, vol. 38, no. 4, p. 116, 2019.

[8] D. Faccio, A. Velten, and G. Wetzstein, “Non-line-of-sight imaging,” *Nat. Rev. Phys.*, pp. 1–10, 2020.

[9] S. I. Young, D. B. Lindell, B. Giord, D. Taubman, and G. Wetzstein, “Non-line-of-sight surface reconstruction using the directional light-cone transform,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1407–1416.

[10] K. L. Bouman et al., “Turning corners into cameras: Principles and methods,” in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 2270–2278.

[11] M. Pritchett and C. Chen, “Principles of inverse synthetic aperture radar/isar/imaging,” in *Proc. EASCON.80*; Elect. Aerospace Syst. Conf., Jan. 1980, pp. 340–345.

[12] X. Liu, S. Bauer, and A. Velten, “Analysis of feature visibility in non-line-of-sight measurements,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 140–10 148.

[13] I. Freund, “Looking through walls and around corners,” *Physica A: Stat. Mechanics Its Appl.*, vol. 168, no. 1, pp. 49–65, 1990.

[14] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, “Looking around the corner using transient imaging,” in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 159–166.

[15] A. Velten et al., “Femto-photography: Capturing and visualizing the propagation of light,” *ACM Trans. Graph.*, vol. 32, 2013.
[16] R. Pandharikar, A. Velten, A. Bardagyi, E. Lawson, M. Bawendi, and R. Raskar, “Estimating motion and size of moving non-line-of-sight objects in cluttered environments,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 265–272.

[17] D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, Q. Dai, and R. Raskar, “Ultra-fast lensless computational imaging through 3d frequency analysis of time-resolved light transport,” *Int. J. Comput. Vis.*, vol. 110, no. 2, pp. 128–140, 2014.

[18] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar, “Reconstruction of hidden 3 d shapes using diffuse reflections,” *Opt. Express*, vol. 20, no. 17, pp. 19 096–19 108, Aug. 2012.

[19] M. Laurentzis, J. Klein, E. Bacher, and N. Metzger, “Multiple-return single-photon counting of light in flight and sensing of non-line-of-sight objects at shortwave infrared wavelengths,” *Opt. Lett.*, vol. 40, no. 20, pp. 4815–4818, 2015.

[20] C.-Y. Tsai, K. N. Kutulakos, S. G. Narasimhan, and A. C. Sankaranarayanan, “The geometry of first-returning photons for non-line-of-sight imaging,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7216–7224.

[21] V. Arelano, D. Gutierrez, and A. Jarabo, “Fast back-projection for non-line of sight reconstruction,” *Opt. Express*, vol. 25, no. 10, pp. 11 574–11 583, 2017.

[22] A. K. Pediredla, M. Buttafava, A. Tosi, O. Cossairt, and A. Veeraraghavan, “Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner,” in *IEEE Int. Computational Photography*, 2017, pp. 1–12.

[23] F. Xu et al., “Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging,” *Opt. Express*, vol. 26, no. 8, pp. 9945–9962, 2018.

[24] C. Jin, Z. Song, S. Zhang, J. Zhai, and Y. Zhao, “Recovering three-dimensional shape through a small hole using three laser scatterings,” *Opt. Lett.*, vol. 40, no. 1, pp. 52–55, 2015.

[25] F. Heide, M. O’Toole, K. Zang, D. B. Lindell, S. Diamond, and G. Wetzstein, “Non-line-of-sight imaging with partial occluders and surface normals,” *ACM Trans. Graph.*, vol. 38, no. 3, pp. 1–22, 2019.

[26] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, “Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3222–3229.

[27] A. Kadambi, H. Zhao, B. Shi, and R. Raskar, “Occluded imaging with time-of-flight sensors,” *ACM Trans. Graph.*, vol. 35, no. 2, p. 15, 2016.

[28] J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, “Non-invasive imaging through opaque scattering layers,” *Nat.*, vol. 491, no. 7423, p. 232, 2012.

[29] O. Katz, E. Small, and Y. Silberberg, “Looking around corners and through thin turbid layers in real time with scattered coherent light,” *Nat. Photon.*, vol. 6, no. 8, pp. 549–553, 2012.

[30] O. Katz, P. Heidrich, M. Fink, and S. Gigan, “Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations,” *Nat. Photon.*, vol. 8, no. 10, p. 784, 2014.

[31] A. Viswanath, P. Rangarajan, D. MacFarlane, and M. P. Christensen, “Indirect imaging using correlography,” in *Proc. COSI*, 2018.

[32] C. A. Metzler et al., “Deep-inverse correlography: Towards real-time high-resolution non-line-of-sight imaging,” *Optica*, vol. 7, no. 1, pp. 63–71, 2020.

[33] D. B. Lindell, G. Wetzstein, and V. Koltun, “Acoustic non-line-of-sight imaging,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6780–6789.

[34] I. An, D. Lee, J.-w. Choi, D. Manocha, and S. E. Yoon, “Diffraction-aware sound localization for a non-line-of-sight source,” in *Proc. Conf. Int. Conf. Robot. Autom.*, 2019, pp. 4061–4067.

[35] A. Torralba and W. T. Freeman, “Accidental pinhole and pinspreek cameras: Revealing the scene outside the picture,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 374–381.

[36] M. Batarseh, S. Sukho, Z. Shen, H. Geman, R. Rezvani, and A. Dogar, “Passive sensing around the corner using spatial coherence,” *Nat. Commun.*, vol. 9, no. 1, pp. 3629, 2018.

[37] C. Thrampoulidou et al., “Exploiting occlusion in non-line-of-sight active imaging,” *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, Sep. 2018.

[38] C. Saunders, J. Murray-Bruce, and V. K. Goyal, “Computational periscopy with an ordinary digital camera,” *Nat.*, vol. 565, no. 7740, p. 472, 2019.

[39] J. Brooks and D. Faccio, “A single-shot non-line-of-sight range-finder,” *Sensors*, vol. 21, no. 19, pp. 4820, 2020.

[40] M. O’Toole, D. B. Lindell, and G. Wetzstein, “Real-time non-line-of-sight imaging,” in *ACM SIGGRAPH Emerging Technologies*, 2018.
David B. Lindell (Member, IEEE) is currently working toward the Ph.D. degree in the Electrical Engineering Department at Stanford University Engineering Department, Stanford University, Stanford, CA, USA. His research interests include developing computational algorithms for non-line-of-sight imaging, single-photon imaging, and 3-D imaging with sensor fusion.

Gordon Wetzstein (Senior Member, IEEE) received the graduation (with Hons.) degree from the Bauhaus-Universität Weimar, Weimar, Germany and the Ph.D. degree in computer science from the University of British Columbia, BC, Canada, in 2011. He is currently an Assistant Professor of Electrical Engineering and, by courtesy, of Computer Science, with Stanford University, Stanford, CA, USA. He is the Leader of Stanford Computational Imaging Lab and a Faculty Co-Director of the Stanford Center for Image Systems Engineering. At the intersection of computer graphics and vision, computational optics, and applied vision science, his research has a wide range of applications in next-generation imaging, display, wearable computing, and microscopy systems. Prior to joining Stanford in 2014, he was a Research Scientist with the Camera Culture Group, MIT. He is the recipient of an NSF CAREER Award, an Alfred P. Sloan Fellowship, an ACM SIGGRAPH Significant New Researcher Award, a Presidential Early Career Award for Scientists and Engineers (PECASE), an SPIE Early Career Achievement Award, a Terman Fellowship, an Okawa Research Grant, the Electronic Imaging Scientist of the Year 2017 Award, an Alain Fournier Ph.D. Dissertation Award, Laval Virtual Award, and the Best Paper and Demo Awards at ICCP 2011, 2014, and 2016 and at ICIP 2016.