Reconstructing Rooms Using Photon Echoes: A Plane based model and reconstruction algorithm for looking around the corner

Adithya Kumar Pediredla
Rice University
Houston, Texas, USA
adithya.k.pediredla@rice.edu

Mauro Buttafava and Alberto Tosi
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Milan, Italy
mauro.buttafava@polimi.it and alberto.tosi@polimi.it

Oliver Cossairt
Northwestern University
Evanston, Illinois, USA
ollie@eecs.northwestern.edu

Ashok Veeraraghavan
Rice University
Houston, Texas, USA
vashok@rice.edu

Abstract

Can we reconstruct the entire internal shape of a room if all we can directly observe is a small portion of one internal wall, presumably through a window in the room? While conventional wisdom may indicate that this is not possible, motivated by recent work on 'looking around corners', we show that one can exploit light echoes to reconstruct the internal shape of hidden rooms.

Existing techniques for looking around the corner using transient images model the hidden volume using voxels and try to explain the captured transient response as the sum of the transient responses obtained from individual voxels. Such a technique inherently suffers from challenges with regards to low signal to background ratios (SBR) and has difficulty scaling to larger volumes. In contrast, in this paper we argue for using a plane-based model for the hidden surfaces. We demonstrate that such a plane-based model results in much higher SBR while simultaneously being amenable to larger spatial scales. We build an experimental prototype composed of a pulsed laser source and a SPAD (single-photon-avalanche-detector) that can achieve a time resolution of about 30ps and demonstrate high-fidelity reconstructions both of individual planes in a hidden volume and for reconstructing entire polygonal rooms composed of multiple planar walls.

1. Introduction

Over the last few years researchers have demonstrated that using information contained in transient light echoes, one can reconstruct objects that are outside the line of sight [23, 37, 16]. Unfortunately, progress in 'looking around the corner' (LATC) has been slow since the early results by Velten et. al [37] and Kirmani et. al [23, 24]. Current techniques still rely on a voxel based representation to model
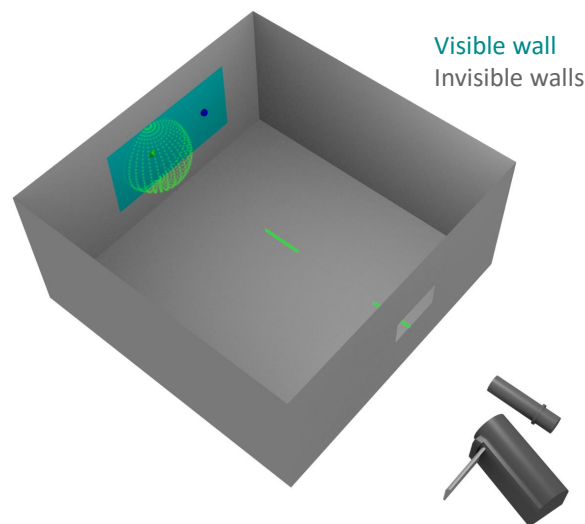

Figure 1. *Can we reconstruct the shape of a hidden closed room with a small peephole?* In this paper, we show an experimental set up, built with a transient camera and a pico-second laser, that can infer the shape of a hidden room. The transient camera consists of single SPAD detector (single photon avalanche diode) with 30 ps jitter. The pico-second laser has a jitter comparable to that a SPAD, and emits pulses at 530-570 nm wavelength. Coherence lengths at this bandwidth are too low to do interferometric measurements. Instead, we rely on the arrival times of photon echoes in our algorithm.

the hidden scene and attempt to reconstruct the hidden volume by expressing the observed transient response as a sum of the transients produced by individual voxels that make up the hidden volume. With such an approach, there are three main challenges that currently limit progress in practical and realistic application scenarios.

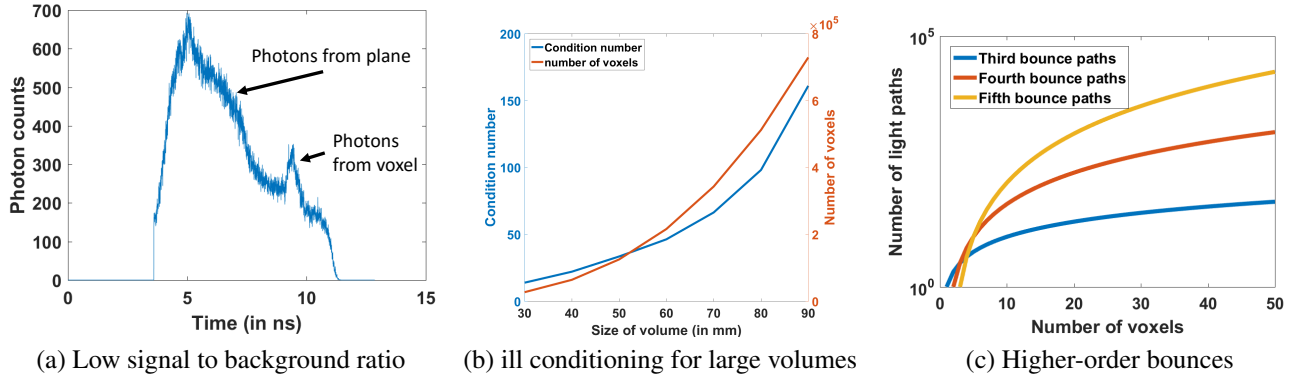| (a) Low signal to background ratio | (b) ill conditioning for large volumes | (c) Higher-order bounces |

Figure 2. **Challenges with current voxel based reconstructions:** (a) Third bounce photons from an individual voxel form a very small fraction of the returning third bounce photons, which by themselves are small in number. The plot shows the captured transient data when the scene consists of a large plane and a small voxel, clearly showing higher SBR for the plane. (b) voxel based reconstructions suffer from ill-conditioning as the volume being reconstructed increases as demonstrated by rapidly worsening condition number. (c) The light paths that correspond to higher order bounces rapidly overwhelm the third bounce light paths as the number of voxels increase.

1. **Low Signal to Background Ratio (SBR):** LATC using transient light echoes relies on the photons that bounced three times. The vast majority of the photons captured are direct reflection photons, or photons that have otherwise encountered multiple internal reflections and therefore act as a large background bias in which the third bounce photon signal is buried. This low SBR is further exacerbated by the use of a voxel-based modeling approach, wherein the photons that are scattered from a small hidden voxel only form a miniscule fraction of the already small number of third bounce photons.

2. **Scaling to Larger Hidden Volumes:** A voxel-based approach inherently means that the number of voxels to be reconstructed increases cubically with the size of the hidden scene. Consequently, the demultiplexing problem becomes exceedingly ill-conditioned as the size of the hidden volume increases.

3. **Higher-order bounces:** While the direct bounce photons can be potentially time-gated (as is the case in our experimental data), the higher-order bounces remain un-modeled and contribute as a source of noise and background. A model-based approach to model these higher-order bounces is intractable as can be seen from the fact that even just the number of fourth bounce light-paths increase quadratically in the number of voxels, which as indicated earlier increases cubically in the size of hidden volume.

As a consequence of these three reasons, existing reconstruction results for LATC typically show a two or three discrete, compact objects that are placed far away from each other to limit inter-reflections. Our goal in this paper, is to consider the toy-problem of reconstructing a hidden room and show that a plane-based model for transient light echoes has the potential to overcome some of these above-mentioned limitations.

Plane-based representation: Motivated by the fact that in many problems in computer vision, including stereo, multi-view stereo [12], RGB-D [22], and structure from motion [41] a plane-based model has been shown to outperform voxel-based representations, especially when encountering man-made scenes, urban landscapes and indoor environments we explore the utility of such a representation for LATC. Firstly, such a plane-based representation results in a higher SBR just by virtue of the fact that there are typically several large planar structures in many environments and the number of photons reflected from these large planar structures can now be modeled together as opposed to being divided among the many hundreds or thousands of voxels that make up this plane. Secondly, such plane-based models are eminently more suitable to scale to larger scenes since we are only modeling the hidden surface, rather than the entire hidden volume. These two facts together result in much better conditioning for the inverse problem when compared to the voxel-based representation. Finally, a plane-based model is a much more compact representation (in terms of number of planes that make up a hidden scene), and this compactness may provide us with opportunities to directly model and account for the higher-order bounces (not yet demonstrated in this paper).

Contributions: In this paper, we demonstrate that a plane-based representation is a promising approach for LATC by focusing on the following toy-problem as a motivating example. Can we reconstruct the entire internal shape of a room if all we can directly observe is a small portion of one internal wall, presumably through a window in the room? The key technical contributions of this paper are the following:

- **Plane-based Representation:** We propose a plane-

based representation for looking around the corner using transient light echoes and develop a dictionary based algorithm for reconstructing a hidden surface as a linear superposition of planes.

- **Symmetries:** We show that there exist several symmetries in the transient light echoes of a plane and demonstrate practical techniques to avoid these symmetries and ensure reliable unique reconstruction.
- **Analysis:** We perform a systematic analysis of the effect of various parameters such as number of virtual sources, time resolution of transients, SBR and the size of visible surface on the reconstruction error.
- **Real Experimental Prototype:** We build an experimental platform composed of a pulsed laser, a single-photon avalanche diode (SPAD) sensor and a time-correlated single photon counting module (TCSPC) and demonstrate state of art results in reconstructing hidden scenes including rooms.

Limitations: While a plane-based representation allows us to outerperform previous reuslts using voxel-based representations, the proposed method for LATC still suffers from many limitations. Firstly, the higher-order bounces have not yet been incorporated into our model and therefore remain a source of noise and background in all our experiments, potentially limiting reconstruction performance. Secondly, our entire approach assumes that the bidirectional reflectance distribution (BRDF) of the hidden subject/room is lambertian and when this assumption is violated, one will have to make appropriate changes to the model to accomodate varying reflectance of the hidden surfaces. Our actual experimental prototype also introduces a couple of important limitations. The overall time jitter of the system (SPAD detector, TCSPC module, pulsed laser source and broadening due to wall surface reflection) is around 100 ps. This time jitter limits the accuracy of our reconstruction. One can employ a more expensive sensors with lower jitter to increase the reconstruction accuracy as shown in our simulations. In addition, we use a supercontinuum laser source with pulse repetition rate of 77.78 MHz. Hence, the SPAD measurements phase wrap after 12.86 ns or around 3.86 m. If the hidden volume is larger than $3.86/2 \approx 1.9$ meters, this would result in errors due to phase wrapping. So all our real scenes were smaller than this limit. In practice, if we needed to accomodate larger scenes, then the pulse repetition rate of the laser source should be reduced, perhaps by using a pulse-picker.

2. Related work
2.1. Transient imaging

Nils Abramson [1] might have been the first to capture light-in-flight (or transient image) with holography. Around 30 years later, Velten et al [36, 38] designed a system based on streak camera and femto-second laser to image the transient interaction of light with matter. Streak camera sacri-

fices spatial resolution for temporal resolution. Hence, Velten et al. used scanning mirrors to acquire a 2D spatio-temporal light-in-flight image. Gao et al [13] extended velten et al.'s technique to capture transient images of dynamic, non-repeatable events by combining compressed sensing techniques with streak camera based light-in-flight imaging. As streak cameras are very expensive, Heide et al. [17] used inexpensive photonic mixer devices (PMD) to capture transient image by incorporating prior knowledge about the light/object interactions in an optimization framework. Lin et al. [28] identified Heide et al.'s approach as capturing Fourier transform of transient image, and proposed a framework to decrease the computational and memory cost of the reconstruction algorithm. Kadambi et al. [19] also captured transient images independently, in the same year as Heide et al., with PMDs by employing M-sequences as coded-illumination. Kadambi et al. deconvolve the images captured by PMD to get the transient response. Tadano et al. [34] increased the temporal resolution of the Kadambi et al.'s technique by employing a light source array. Tadano et al. [33] also built a depth selective camera by using M-sequences as coded illumination. Gariepy et al. [14] was the first to use SPADs to capture transient images directly at 67 ps temporal resolution. The spatial resolution of the camera they used is 32×32. The transient imagers like SPAD were also redesigned to capture photometric and geometric information from just one photon [25, 32], or image with out the need for optical elements [24].

2.2. Looking around the corners

To our knowledge, Kirmani et al. [23] first theorized the ability to look around the corners using 5D space-time-impulse-response (STIR) of the scene. They showed results on reconstructing hidden patches. Velten et al. [37] demonstrated real experiments for looking around corners using third-bounce reflections and elliptic-back projection technique. Gupta et al. [16], using the same set-up as Velten et al., modeled the forward problem as a linear system, and studied various inverse algorithms for reconstructing hidden shapes. Heide et al. [18] used low cost PMD sensors, and incorporated sparsity priors on the depth and volume in an optimization framework to reconstruct objects around the corner. Buttafava et al. [5] used single pixel SPADs with a temporal resolution of 30 ps. They captured a 3D STIR by scanning the wall with a 515 nm wavelength mode-locked pulsed laser source, running at 55 MHz. Tsai et al. [35] used the information from the first bounce photons alone to reconstruct objects around the corners. We have used a similar transient camera system whose details are given in hardware set up section. Kadambi et al. [20] derived resolution bounds with the help of array signal processing techniques and Pediredla et al. [30] derived the bounds using linear systems approach. Using the wave nature of the light, Katz et al. [21] proposed a transmission matrix based approach which Rangarajan et al. [31] extended to hologra-
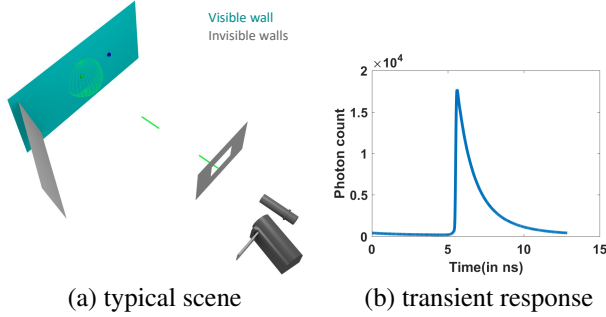
(a) typical scene   (b) transient response

Figure 3. **Forward model:** We use a Monte-Carlo based computational model to solve for the transient response of invisible plane. (a) shows a typical scene. The geometry consists of a visible wall (cyan color in the rest of the paper) and an invisible wall (gray color in the rest of the paper). The pulsed laser source emits photons that hit the visible wall at a location referred as virtual source (VS) and marked as a green dot in the scene which emits photons isotropically. Some of these photons hit the invisible wall, and return to the visible wall. The detector is focused on the wall on a location referred as virtual detector (VD) and marked with blue dot in the scene. The transient response is defined as the photon flux reaching the virtual detector from virtual source. (b) shows the transient response accounting for the jitter of the hardware we have built in the lab.

phy to solve the looking around corners problem.

The reconstruction problem is also extended to track around the corners by Gariepy et al. [15], who used a SPAD array to track objects beyond line of sight. The acquisition time is around 3 seconds for each position of the hidden object, and hence, this technique can only track slow moving objects. Recently, Kleint et al [26], used an ordinary 2D intensity camera, to track the translation and orientation of known objects in real time.

2.3. Hearing the shape of a room

Our problem statement is similar to the one addressed by Dokmanic et al. [10, 11] in the context of audio. By recording the acoustics echoes of the room with multiple microphones and a loud speaker located at fixed known positions inside the room, Dokmanic et al. proposed to reconstruct the shape of the room. The room is assumed to made of planes and convex in shape. The two important differences between the Dokmanic et al.'s research from ours is in the choice the location of transmitter/receiver and the waves used. Dokmanic et al placed the sensors inside the room where as our sensors are outside the room. Dokmanic et al. used sound waves whereas we chose to solve the problem with optical waves. As the surfaces are highly specular (mirror like) at the wavelength of sound (17 millimeters to 17 meters), the problem of room reconstruction is one of finding the correct echoes in each microphone corresponding to any given wall. Though not very easy, Dokmanic et al. used the rank property of Euclidean distance matrices,

popularly used in GPS literature [9], to sort the echoes and reconstruct the room. However, at the wavelength of light (400-700nm), the roughness of the materials is comparable enough to make the surfaces diffuse. Therefore the photon echoes have a smooth transient response and we cannot use the ideas developed by Dokmanic et al. directly.

3. Reconstructing a plane

Can we reconstruct the entire internal shape of a room if all we can directly observe is a small portion of one internal wall, presumably through a window in the room? We will address this question by first addressing a simpler problem of reconstructing a single infinitely large plane by observing its transient response. Once we have understood this problem, we will then model rooms as composed of many intersecting planes and extend our method to reconstruct a room – one plane at a time.

3.1. Forward Model

We parameterize the planes in (z_{int}, θ, ϕ) co-ordinates, where z_{int} is the z-intercept of the plane, θ is the angle between plane normal and the z-axis, ϕ is the angle between x-axis and the projection of plane normal in the xy plane. Note that this parameterization will result in $z_{int} = \pm\infty$ for all the planes parallel to z-axis (or $\theta = \frac{\pi}{2}$). In this section, we will only consider planes not parallel to z-axis.

Here, we will derive the transient response of plane. Figure 3 shows the setup of a typical scene. We will refer to the wall visible from the laser/camera pair as the "visible wall" and the other wall plane as the "invisible wall". The source emits a laser beam that hits the visible wall at a location referred as virtual source (VS). The detector is focused on a point on the visible wall, referred as virtual detector (VD). The analytical expression for the photon flux ($\Phi(t)$) received at VD at time t is given by

$$\Phi(t) = I \int_{\forall \bar{p} \in \mathcal{C}(t)} \frac{\cos \angle(\bar{v}_s - \bar{p}, N_i) \cos \angle(\bar{p} - \bar{v}_d, N_v)}{\|\bar{p} - \bar{v}_s\|^2 \|\bar{p} - \bar{v}_d\|^2},$$

(1)

where I is the photon flux emitted from VS, $\bar{v}_s$ and $\bar{v}_d$ are the position vectors of VS and VD, N_v and N_i are the unit normals of the visible wall and invisible wall. $\mathcal{C}(t)$ is the intersection of the invisible wall with the ellipsoid of revolution, whose focal points are $\bar{v}_s$ and $\bar{v}_d$ and major axis is $\frac{ct}{2}$, where c is the speed of light.

Unfortunately, the analytical formula for the transient response of a plane does not have closed form expression. This is because, the curve $\mathcal{C}(t)$ is an elliptical ring on the plane. The transient photon flux reaching the camera at time t is an integral of the fluxes contributed by points on this elliptic ring. Even the perimeter of an ellipse is not a tractable integral and does not have a closed form expression [6]. Hence, it is not possible to have a closed form expression
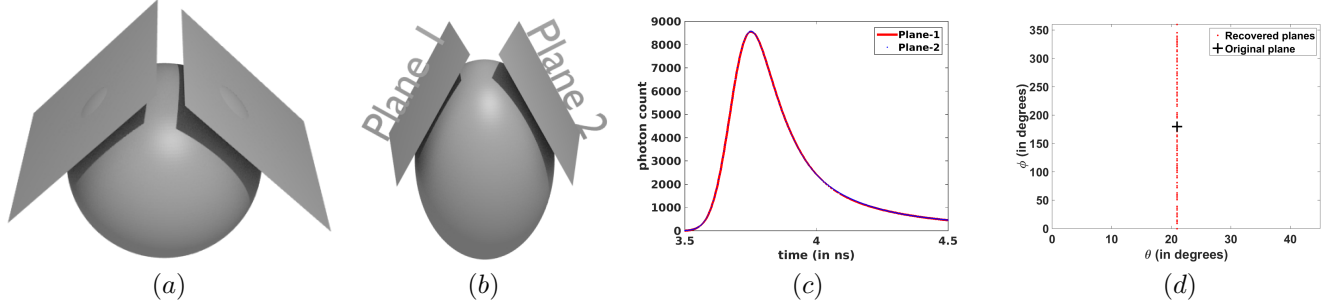
Figure 4. Symmetry for a single transient measurement (a) When VS and VD are chosen to be same point, all the planes that are at a fixed distance have the same transient response and we cannot recover the plane parameters (b) even when the VS and VD are not at the same point, we have a symmetry. All the planes that have same z_{int} and θ will have same the same transient response. (c) We show the transient response of two planes that has different ϕ but same z_{int} and θ. We can notice that the transient responses are identical. (d) For a given plane, we recover the plane parameters using sparse dictionary approach for various realizations of noise and plot the θ and ϕ values of the recovered plane. We can notice that the error in θ is negligible where as ϕ varies arbitrarily over $[0, 360°]$.

for the transient response of the plane. Therefore, we use a Monte-Carlo based computational model for computing the transient response of the plane given its geometry.

To estimate the transient response of a room, we generate a million virtual photons uniformly placed on the hemisphere facing the normal of the visible wall, and track them. Some of these photons will hit the plane of interest and the others are lost into vacuum. The number of photons that hit the visible wall is a function of the plane normal. For each photon that hits the plane of interest, we compute the travel time and intensity of the flux that reaches the VD by accounting for foreshortening due to the distance and orientation of VD. We compute the transient response by binning the photon flux based on the travel time. Note that the transient response corresponds to the time traveled by photons from virtual source to virtual detector and not from the source to detector. We will show in the experiments section that we can calibrate and remove the bias introduced by the photon travel from source to virtual source and detector to virtual detector.

A sample impulse response of a random plane is shown in Figure 3. We can notice that the response starts after a finite time t_0, the minimum round trip time from source to the visible plane, and then falls exponentially, reaching close to zero after a few nanoseconds. In practice, the tail has less strength compared to the background level of the transient imaging system. Due to the repetition rate of the laser, the transient response phase wraps after ~ 12.5 ns. The SPAD system we have used has a temporal jitter higher than the sampling time period. Hence, the transient measurements of the planes are smooth as shown in Figure 3.

3.2. Symmetry

Can observing the transient response from any single virtual source (VS)/virtual detector (VD) pair result in a unique and accurate estimate of the plane parameters?

The short answer is NO. There is an inherent symme-

try in the underlying transient responses that make unique plane identification impossible– i.e., multiple planes in the scene will result in the same transient response. As an illustrative example, lets consider the case when VS and VD are co-located, for simplicity at the origin. When a photon travels from the VS, to the object, back to the VD, the roundtrip travel time will depend on the 3D distance from the VS/VD to the point. As a result, there is a locus of scene points along a spherical surface whose distance or time traveled is constant (i.e., the iso-temporal surface). All planes that are equidistant from the origin (i.e, planes with the same perpendicular distance to the origin), are tangential to this spherical iso-temporal surface. The 2-D family of all these planes are symmetric with respect to the origin and will result in the same transient response due to this symmetry. Two such planes and their transient responses are shown in Figure 4(a). To re-state this symmetry explicitly – given a round trip travel time t_0, we can draw the corresponding iso-temporal surface which is a sphere of radius $r_0 = c\frac{t_0}{2}$. All planes that are tangential to this surface produce the exact same transient response.

Such a symmetry exists not just for the case when VS and VD are co-located, but also for arbitrary VS-VD combinations. For example, consider the case when both VS and VD are on the z-axis. The resulting iso-temporal surface is a spheroid that is symmetric around z-axis as shown in Figure 4(b). All the planes that make same angle with z-axis and have the same z-intercept will have the exact same impulse response due the geometric symmetry of the spheroid. This is a little harder to visualize but as shown in Figure 4(b), all of these planes result in the same transient response. Thus there is now a 1-D family of planes that result in exactly the same transient response. Hence, it is not possible to invert the observed transient response and estimate the plane parameters uniquely from the transient responses observed with *any single VS-VD pair*. In Figure 4(d), we show that it is not possible to recover the angle ϕ for this example case.

# virtual sources	SBR				# virtual sources	SBR				# virtual sources	SBR		
	1	3	10			1	3	10			1	3	10
4	30.52	14.43	12.43		4	6.13	2.78	2.48		4	27.37	3.01	1.24
8	30.21	10.45	9.73		8	6.37	2.08	2.03		8	17.01	1.59	0.75
16	24.30	9.33	9.90		16	5.43	2.12	2.14		16	18.70	1.49	0.80

(a) Reconstruction error in z_{int} (mm)　　　(b) Reconstruction error in $\theta°$　　　(c) Reconstruction error in $\phi°$

Table 1. Simulation: Effect of increasing the number of virtual sources on the reconstruction of a plane

scan area (mm × mm)	SBR				scan area (mm × mm)	SBR				scan area (mm × mm)	SBR		
	1	3	10			1	3	10			1	3	10
50 × 50	51.05	33.76	25.14		50 × 50	10.95	7.79	6.22		50 × 50	76.47	12.76	4.34
100 × 100	45.99	20.83	14.66		100 × 100	9.86	4.35	2.91		100 × 100	42.04	7.33	7.30
200 × 200	30.52	14.43	12.43		200 × 200	6.13	2.78	2.48		200 × 200	27.37	3.01	1.24
400 × 400	19.77	9.70	9.92		400 × 400	3.60	1.83	1.81		400 × 400	13.28	1.79	0.96

(a) Reconstruction error in z_{int} (mm)　　　(b) Reconstruction error in $\theta°$　　　(c) Reconstruction error in $\phi°$

Table 2. Simulation: Effect of increasing the scan area on the reconstruction of a plane

temporal jitter (ps)	SBR				temporal jitter (ps)	SBR				temporal jitter (ps)	SBR		
	1	3	10			1	3	10			1	3	10
25	15.33	12.43	12.38		25	3.04	2.48	2.48		25	8.34	1.37	1.28
50	20.69	12.43	12.43		50	4.44	2.48	2.48		50	14.15	1.86	1.38
100	30.52	14.43	12.43		100	6.13	2.78	2.48		100	27.37	3.01	1.24
200	45.86	15.74	12.43		200	9.83	3.24	2.48		200	46.69	6.24	1.21
400	71.40	23.32	12.89		400	13.96	4.88	2.55		400	78.79	13.86	2.35

(a) Reconstruction error in z_{int} (mm)　　　(b) Reconstruction error in $\theta°$　　　(c) Reconstruction error in $\phi°$

Table 3. Simulation: Effect of increasing the temporal jitter of the system on the reconstruction of a plane

3.3. Explicit symmetry breaking

While the symmetry inherent in any single VS-VD pair precludes the possibility of unique identification of a hidden plane, we can fortunately break the symmetry by carefully choosing at multiple VS-VD pairs that ensure that the cumulative transient responses for each plane is indeed unique. As an example, even when only two VS-VD pairs are chosen carefully, such that they are not collinear, it can be shown that these symmetries can be eliminated ensuring unique and accurate reconstruction of plane parameters. In practice though, we choose several VS-VD combinations for reconstruction since these additional measurements provide robustness to other un-modeled sources of errors such as inter-reflections and background. Due to practical challenges in the physical implementation, we fixed the location of VD and moved VS with the help of a pair of galvanometric mirrors introduced in the light path of the pulsed laser source. In Figure 5 , we show how choosing multiple VS locations, even with a fixed VD location breaks symmetry.

3.4. Characterization of single plane recovery

To recover a single hidden plane, we used a dictionary-based approach. Let D be the dictionary of transient responses of various hidden planes. For creating D, we first sampled multiple planes uniformly from $z_{int} \in$ [200mm, 800mm] in steps of 20 mm, $\theta \in [0°, 45°]$ in steps of 3°, $\phi \in [0°, 360°]$ in steps of 3°. We limit the choice of θ to a max of 45° as further increase creates planes that go between VS and VD. The transient response of each one of these planes will be a column of the matrix D, which is obtained by evaluating the value of the integral in equation 1 at these plane parameters.

Once we have the dictionary (D), and a transient measurement ($\bar{y}$), the hidden plane can be obtained by solving the following optimization problem.

$$\arg \min_{\bar{x}} \|\bar{y} - D\bar{x}\|,$$
$$\text{subject to } x_i \in \{0, 1\},$$
$$\sum_i x_i = n,$$

where n is the number of hidden planes ($n = 1$ for single hidden plane recovery). The above problem is computationally hard and we relax it with the following problem:

$$\arg \min_{\bar{x}} \|\bar{y} - D\bar{x}\| + \lambda \|\bar{x}\|_1.$$

The value of λ is chosen appropriately so that the sparsity
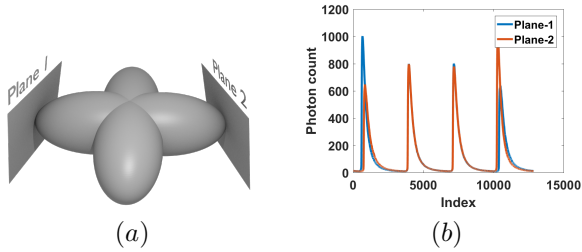
(a) (b)

Figure 5. **Symmetry breaking:** We break the symmetry by gathering data from more than one VS location. We can notice from the figure that each VS-VD combination has its own symmetry but the entire configuration is not have symmetric as the virtual sources are ordered. (b) The concatenated impulse response from all four virtual source locations on both the planes. Notice how the plane responses are different.

of $\bar{x}$ is n. The hidden planes are the ones corresponding to the values of $\bar{x}$ that are non-zero.

For simulation studies, we have to generate the transient response of an arbitrary plane (y) accounting for noise and background into the ideal transient response. First, the actual observed response is a Poisson process whose rate is given by equation 1 – therefore, the transient response of an arbitrary plane must be a realization of this random process so that the effect of photon counting noise is incorporated. Second, the actual observed response will always contain a background which can be modeled as a Poisson process with a constant rate (the constant depending upon the strength of unmodeled factors including wrap-around due to pulse repetition rate and higher order bounces). We account for both these effects in simulating the transient response y by appropriately sampling from Poisson processes. Other noises such as read noise are negligible for SPAD.

The reconstruction algorithm is influenced by various experimental parameters, namely signal-to-background ratio (SBR), number of virtual sources, the total scan area (where the virtual sources are placed), and temporal jitter of the transient camera. Through simulations, we have systematically evaluated the effect of these parameters in this section. The SPAD set up we have used in our experiments has an SBR ~ 5, temporal jitter of 100 ps. We varied the simulation parameters around these experimental settings and quantified the reconstruction accuracy on randomly generated hundred planes. All these random planes are sampled uniformly from $[200\text{ mm},\ 800\text{ mm}] \times [0°,\ 45°] \times [0°,\ 360°]$ and the transient response y is computed by appropriately sampling from the Poisson process on the ideal response.

The results of the simulation experiments are in Tables 1-3. The main conclusions of the simulation experiments are

1. Improving SBR beyond 10 has no significant effect.
2. Increasing number of VS has diminishing returns.

3. Increasing scan area and decreasing temporal jitter improves accuracy, especially at low SBR.

4. Reconstructing Rooms

In the previous section, we showed how to estimate plane parameters of an infinite plane from the transient response. Here, we will use those concepts to solve an interesting challenge problem of reconstructing the geometry of the room from the transient response with only one visible surface of the room.

4.1. Problem setup

Consider a polygonal room with a small peephole which gives us the ability to see one wall. We will call this wall as a visible wall. There are n different invisible walls, and the room is made up of all these (n+1) walls. Even on the visible wall, we can observe only a finite area. We want to reconstruct the entire geometry of the room, by which we imply the parameters of all the planes that constitute the room. We want to solve this inference from the transient response of the room obtained by placing virtual sources and detectors on the visible area of the visible wall.

4.2. Modeling inaccuracies

One simple method to approach this problem is to model the transient response as a sum of the transient responses that will be observed if each of those individual planes that compose the room were present separately. If such a linear model works, then the problem is solved as we can use linear inversion techniques to solve the problem. Unfortunately, this linear model is inaccurate in the following three ways:

Concave polygons: Assume that the room is a concave polygon from the viewpoint of the visible wall. Hence, there is a point in the room that is not in the direct line of sight of the visible wall. So, only fourth bounce photons are the ones we collect. However, the linear model takes care of only third bounce photons, and hence, we cannot estimate this wall. So, the only kind of rooms we are going to consider are the ones where the room is a convex polygon, which is not a very restrictive assumption as most rooms are convex polygons.

Finite planes: All our planes are assumed to be infinite. However, due to the occlusions caused by other walls, the planes are finite in nature. The finite extent of the walls affects the transient response observed. We will show in Section 5 on how to address this problem.

Higher order bounces: When we considered single planar surface, only third bounce photons reached the detector. Now, when we have multiple surfaces, the fourth, fifth and higher order bounces also reach the detector. These light paths are not modeled in the original planar model, which causes another source of error. We will show in Section 5 on how to handle both the finite plane and higher order

bounce errors together.

5. Handling non-linearities

Within the construction of convex polygons, now let us look at how the problems induced by finite planes and interreflections can be handled. We will use a two-step algorithm. First, we will create an initial estimate of the room, then we will refine the initial estimate of the room, by taking into account the finiteness and inter-reflections between the planar surfaces.

5.1. Initial estimate of the walls

Consider the case where the virtual source and detector on the visible wall are at least a distance 150 mm away from the nearest invisible wall. The third bounce photons reach the detector much before the fourth and other higher order bounces. Hence, we can solve the problem using third bounce photons.

5.2. Refining the estimate of the wall

When we have finite walls, we have two factors that cause the difference between the transient response of the finite wall to an infinite wall, namely non-linear attenuation factor, and the time extent. The time extent, where the transient response of the wall is non-zero, is smaller for a finite wall, compared to that of an infinite wall. The non-linear attenuation factor comes into picture due to the points that are within the time extent of the finite wall but are not part of the finite wall. We reflect the effect of these parameters on the dictionary and find a more accurate estimate of the walls. Once we obtain a more precise estimate of the shape of the room, we repeat the process till convergence.

6. Hardware setup

The block diagram of our setup is shown in Figure 6. The set up can be broken into three main components: (i) picosecond pulsed laser illumination, (ii) single-photon detection and TCSPC electronics and (iii) reconstruction algorithm. The laser source is used to synchronize the entire setup, triggering the SPAD-based detection module and acting as a STOP signal for the time-measurement unit.

6.1. Illumination

A picosecond supercontinuum laser (SuperK Extreme, EXW-12, NKT Photonics A/S) is used as illumination source. The laser generates light pulses with around 30 ps width at 77.78 MHz repetition rate, with a broad wavelength range (between 400 and 2400 nm). We used tunable band pass filter (SuperK Varia, NKT Photonics A/S) to select a wavelength range of 530-570 nm. The output from the filter is injected into an optical fiber and then is collimated with an achromatic aspherical collimator (PAF-X-11-PC-A, Thorlabs Inc.). The average output power of the free-space

beam reaches 50 mW and the beam diameter is approximately 3 mm. The position of the virtual source on the wall is controlled using a 2-axis galvanometric mirrors (GVS-012, Thorlabs Inc.) able to scan 40 degrees optical angle in both x and y directions.

6.2. Single-photon detection and TCSPC electronics

We used gated-mode single-photon detection module developed at Politecnico di Milano (Milano, Italy) as transient camera [4]. It is based on a silicon SPAD [8] with a circular active area of 50μm diameter [39]. The module exhibits a detection efficiency around 30% at 550 nm, a dark count rate lower than 50 counts per second (at $15°$C temperature) and a temporal response narrower than 35 ps Full-Width at Half Maximum (FWHM). SPAD count-rate saturates at around 8×10^6 counts per second. Hence, to prevent saturation, we exploit the gated-mode operation capability of the detection system, in order to reject the first bounce photons. The gate duration is set to 6 ns, during which the SPAD is able to detect photons. The time position of the gate respect to the laser pulse is controlled using an adjustable pico-second delayer, with 50 ns delay range and 10 ps resolution (Micro-photon Devices s.r.l.).

Every time a photon is detected, the module sends a NIM (Nuclear Instrument Module) compatible voltage signal to the TCSPC unit. TCSPC technique [29, 2] relies on the precise measurement of the time elapsed between each detected photon and the corresponding illumination pulse, building a histogram of photon arrival times in order to reconstruct the waveform of optical signal under investigation, with picosecond resolution. We used a PicoHarp 300 (PicoQuant GmbH) as TCSPC module. The NIM output pulses of the picosecond laser and SPAD module are used as input channels. The bin-width of the PicoHarp 300 is 4 ps and its maximum conversion rate is around 10^7 conversions per second, well above the SPAD saturation count rate. The resulting histograms are saved in a PC for further data processing. The measured jitter of the overall system is around 100 ps FWHM. This value is the sum of the contributions from the laser source, SPAD module, TCSPC unit and the broadening of the light pulse when reflected by the wall.

6.3. Reconstruction algorithm

For calibration, we remove the time bias introduced by SPAD electronics, time travel between source to VS and detector to VD. We first measure the transient response recorded by SPAD when the laser is pointing towards VD. From peak detection, we find the time that corresponds to the total travel time between the source to the VD, VD to the detector, and other delays in the electronic circuits. We physically measured the distance between the VS and the wall to remove the effect for travel time between source and VD from the above time measurement. To account for the travel time between VS to the source, for every scan position, we also measure the location of the VS from the VD
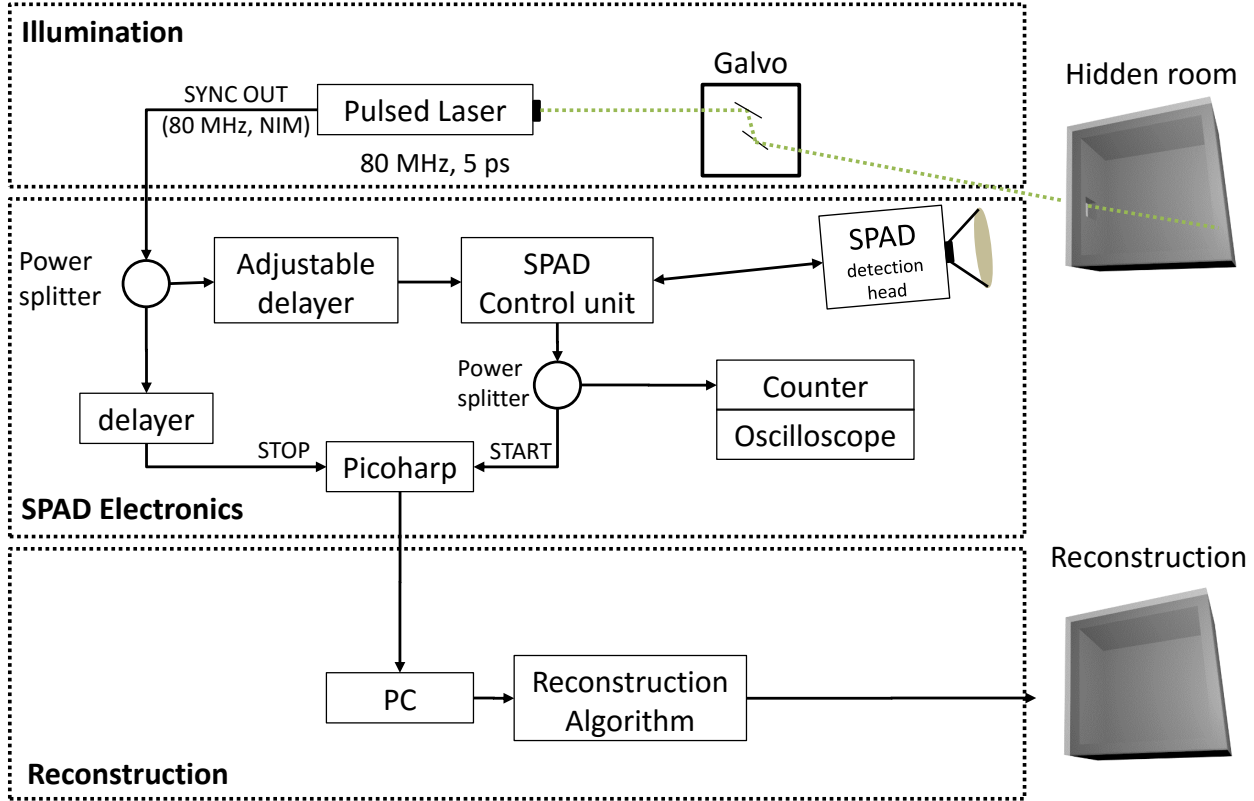
Figure 6. **Block diagram of experimental setup:** The three core components of our setup are illumination hardware, SPAD electronics, and reconstruction algorithm. The illumination hardware consists of a pulsed laser, which sends periodic pulses of short duration and a galvo

using a calibrated point grey camera. By summing these two delays, we compute the shift we need to incorporate to each histogram. We shift the histogram measured for each VS position appropriately and the resulting histograms corresponds to the travel time between VS to object to VD. We then use the reconstruction algorithm discussed in Section 5 for measuring plane parameters.

7. Experiments

In this section, we systematically evaluated the performance of plane reconstruction and room reconstruction.

7.1. Single plane

As mentioned in Section 5, we are interested in the computation of plane parameters of the finite plane and not the finiteness parameters as the later can be computed by the intersection of walls of the room. We first analyzed the reconstruction accuracy of single plane by varying the size, orientation and distance of the plane. The lab was almost sealed off to avoid stray light reaching the detector. However, there was always some amount of background light reaching the SPAD, also due to inter-reflections, and resulting SBR was around 4. Table 4 summaries the reconstruction accuracies.
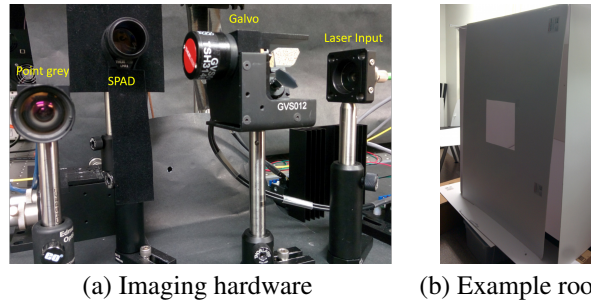

(a) Imaging hardware   (b) Example room

Figure 7. **Hardware setup:** The image hardware is madeup of fiber optic laser input, galvo, and SPAD. We also used a point grey camera for calibration purposes. We built the room with white foam boards. Though they are weak and sag on their own weight, foam boards offer the flexibility of reconfiguring the room shapes. The hole in the front foam board is the peephole of our setup.

The intercept, azimuth, and elevation angles have an approximate error of 50 mm, 5° and 9° respectively.

Here we also introduce a metric inspired by iterative closest point algorithm which is used in vision literature to compute the distance between two surfaces [3, 40, 7].

Dimensions of the	Ground truth			Reconstruction			Error			AICP
plane (mm × mm)	Intercept (mm)	$\theta°$	$\phi°$	Intercept (mm)	$\theta°$	$\phi°$	Intercept (mm)	$\theta°$	$\phi°$	error (mm)
1000 × 2000	570	13.3	0	630	18	354	60	4.7	6	46
1000 × 2000	490	13.3	90	540	12	96	50	1.3	6	34
1000 × 2000	410	13.3	180	450	18	186	40	4.7	6	46
1000 × 2000	500	13.3	270	540	18	246	40	4.7	24	50
1000 × 2000	600	21.1	0	600	12	348	0	9.1	12	199
1000 × 2000	500	21.1	90	540	18	96	40	3.1	6	79
1000 × 2000	300	21.1	180	360	18	180	60	3.1	0	51
1000 × 2000	460	21.1	270	510	24	252	50	2.9	18	30
750 × 750	600	21.1	0	690	30	354	90	8.9	6	46
750 × 750	490	21.1	90	510	18	102	20	3.1	12	74
750 × 750	300	21.1	180	330	30	186	30	8.9	6	115
750 × 750	470	21.1	270	540	24	264	70	2.9	6	28
Average error	-	-	-	-	-	-	45.83	4.78	9	66.5

Table 4. **Real data (Single plane reconstruction):** A single plane is placed at different orientations and positions. We measured the transient photon flux with the SPAD + TCSPC hardware.

We refer to this metric as Average Iterative Closest Point (AICP) error, and define it as the average distance between the points in the original finite plane and the reconstructed plane. AICP is independent of the co-ordinate system and is one number that encompasses the error in intercept, θ, ϕ, and the finiteness of the plane. The average AICP for the planes we have reconstructed is around 7 cm.

7.2. Rooms

We constructed rooms using foamboards as shown in Figure 7. Foamboards sag on their own weight when placed vertically but offer the ease in reconfiguring room. As the single plane reconstructions do not have a lot of inter reflections, we have first placed only three hidden walls, without roof and the front wall. This configuration has significant inter-reflections compared to a single wall. We have observed an average ICP error of 84 mm. Notice the increase in the ICP error compared to reconstruction of single wall.

We then proceeded with all five hidden walls that will make a closed room. We show the reconstruction results in Figure 8. The average ICP error for the walls of the room is 106 mm. The average wall length is 1.1m. Hence, the percentage ICP error is 9.18%.

Jaccard index [27] is used to compare the similarity of sets. We extend it to compare the rooms we have reconstructed with the original rooms and redefine it to quantify the reconstruction accuracy of the rooms as

$$ J = \frac{\mu(O \cap R)}{\mu(O \cup R)}, $$

where O is the set of voxels inside original room, R is the set of voxels inside the reconstructed room μ is the cardinality of the set. The Jaccard similarity for both the rooms are 0.63 (higher is better).
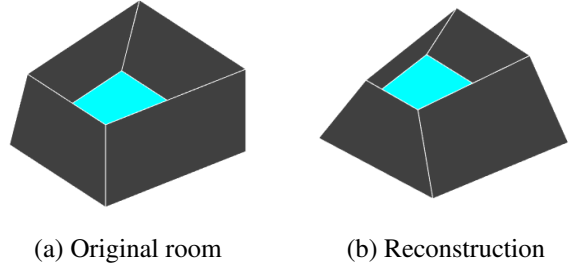

(a) Original room   (b) Reconstruction

Figure 8. **Real data (Room reconstruction):** (a) A hidden sample roomis used for testing the proposed algorithm. The cyan color indicates the visible wall. (b) The resulting wall reconstructions have an average iterative closest point error of 106 mm (9.18%) and jaccard similarity of 0.63 (higher is better).

8. Conclusions

We have proposed a plane based model for looking around corner. We showed how symmetries exist in the transient light echoes of a plane and demonstrated practical techniques to break the symmetry. We then performed systematic analysis of various parameters on the reconstruction of a plane. We used this formulation for the reconstruction of hidden rooms. We built a prototype experimental platform composed of pulsed laser, SPAD, and a time correlated single photon counting module. We used the prototype for reconstruction of single hidden planes and a room. Using additional prior,s such as the right angle between the adjacent walls, average size of the room, can further improve the accuracy of the reconstruction.

References

[1] N. Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, 1978.

[2] W. Becker. *Advanced time-correlated single photon counting techniques*, volume 81. Springer Science & Business Media, 2005.

[3] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[4] M. Buttafava, G. Boso, A. Ruggeri, A. Dalla Mora, and A. Tosi. Time-gated single-photon detection module with 110 ps transition time and up to 80 mhz repetition rate. *Review of scientific instruments*, 85(8):083114, 2014.

[5] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015.

[6] T. R. Chandrupatla and T. J. Osler. The perimeter of an ellipse. *Math. Sci*, 35:122–131, 2010.

[7] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[8] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996.

[9] J. Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu. com, 2010.

[10] I. Dokmanić, Y. M. Lu, and M. Vetterli. Can one hear the shape of a room: The 2-d polygonal case. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 321–324. IEEE, 2011.

[11] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, 2013.

[12] D. Gallup, J.-M. Frahm, and M. Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1418–1425. IEEE, 2010.

[13] L. Gao, J. Liang, C. Li, and L. V. Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014.

[14] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio. Single-photon sensitive light-in-fight imaging. *Nature communications*, 6, 2015.

[15] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 2015.

[16] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20(17):19096–19108, 2012.

[17] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)*, 32(4):45, 2013.

[18] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014.

[19] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics (ToG)*, 32(6):167, 2013.

[20] A. Kadambi, H. Zhao, B. Shi, and R. Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics (TOG)*, 35(2):15, 2016.

[21] O. Katz, E. Small, and Y. Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature photonics*, 6(8):549–553, 2012.

[22] A. Kirmani, A. Colaço, F. N. Wong, and V. K. Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Optics Express*, 19(22):21485–21507, 2011.

[23] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009.

[24] A. Kirmani, H. Jeelani, V. Montazerhodjat, and V. K. Goyal. Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing. *IEEE Signal Processing Letters*, 19(1):31–34, 2012.

[25] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. Wong, J. H. Shapiro, and V. K. Goyal. First-photon imaging. *Science*, 343(6166):58–61, 2014.

[26] J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B. Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports*, 6, 2016.

[27] M. Levandowsky and D. Winter. Distance between sets. *Nature*, 234(5323):34–35, 1971.

[28] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai. Fourier analysis on transient imaging with a multifrequency time-of-flight camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3230–3237, 2014.

[29] D. O'Connor. *Time-correlated single photon counting*. Academic Press, 2012.

[30] A. K. Pediredla, N. Matsuda, O. Cossairt, and A. Veeraraghavan. linear systems approach to identifying performance bounds in indirect imaging. In *To appear in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.

[31] P. V. Rangarajan and M. P. Christensen. Imaging hidden objects by transforming scattering surfaces into computational holographic sensors. In *Computational Optical Sensing and Imaging*, pages CTh4B–4. Optical Society of America, 2016.

[32] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro. Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors. *IEEE Transactions on Computational Imaging*, 1(2):112–125, 2015.

[33] R. Tadano, A. Kumar Pediredla, and A. Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3595–3603, 2015.

[34] R. Tadano, A. K. Pediredla, K. Mitra, and A. Veeraraghavan. Spatial phase-sweep: Increasing temporal resolution of transient imaging using a light source array. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1564–1568, Sept 2016.

[35] 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 *CVPR*, 2017.

[36] A. Velten, E. Lawson, A. Bardagjy, M. Bawendi, and R. Raskar. Slow art with a trillion frames per second camera. In *ACM SIGGRAPH 2011 Talks*, page 44. ACM, 2011.

[37] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3:745, 2012.

[38] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (TOG)*, 32(4):44, 2013.

[39] F. Villa, D. Bronzi, Y. Zou, C. Scarcella, G. Boso, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, et al. Cmos spads with up to 500 μm diameter and 55% detection efficiency at 420 nm. *Journal of Modern Optics*, 61(2):102–115, 2014.

[40] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994.

[41] Z. Zhou, H. Jin, and Y. Ma. Robust plane-based structure from motion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1482–1489. IEEE, 2012.