Computational multi-depth single-photon imaging

This work was made openly accessible by BU Faculty. Please share how this access benefits you.
Your story matters.

| Version | Accepted manuscript |
|---------|---------------------|
| Citation (published version): | Dongeek Shin, Feihu Xu, Franco NC Wong, Jeffrey H Shapiro, Vivek K Goyal. 2016. "Computational multi-depth single-photon imaging." OPTICS EXPRESS, Volume 24, Issue 3, pp. 1873 - 1888. https://doi.org/10.1364/OE.24.001873 |

https://hdl.handle.net/2144/39191

Boston University
Computational multi-depth
single-photon imaging

Dongeek Shin,1,* Feihu Xu,1 Franco N. C. Wong,1 Jeffrey H. Shapiro,1 and Vivek K Goyal2
1Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
2Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA
*dshin@mit.edu

Abstract: We present an imaging framework that is able to accurately reconstruct multiple depths at individual pixels from single-photon observations. Our active imaging method models the single-photon detection statistics from multiple reflectors within a pixel, and it also exploits the fact that a multi-depth profile at each pixel can be expressed as a sparse signal. We interpret the multi-depth reconstruction problem as a sparse deconvolution problem using single-photon observations, create a convex problem through discretization and relaxation, and use a modified iterative shrinkage-thresholding algorithm to efficiently solve for the optimal multi-depth solution. We experimentally demonstrate that the proposed framework is able to accurately reconstruct the depth features of an object that is behind a partially-reflecting scatterer and 4 m away from the imager with root mean-square error of 11 cm, using only 19 signal photon detections per pixel in the presence of moderate background light. In terms of root mean-square error, this is a factor of 4.2 improvement over the conventional method of Gaussian-mixture fitting for multi-depth recovery.

© 2016 Optical Society of America

OCIS codes: (110.1758) Computational imaging; (110.6880) Three-dimensional image acquisition; (280.3640) Lidar.

References and links
1. K. I. Chang, K. W. Bowyer, and P. J. Flynn, “An evaluation of multimodal 2D+3D face biometrics,” IEEE Trans. Pattern Anal. Mach. Intell. 27, 619–624 (2005).
2. J. Côté, J. Widlowski, R. A. Fournier, and M. M. Verstraete, “The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial LIDAR,” Remote Sensing of Environment 113, 1067–1081 (2009).
3. C. D. Mutto, P. Zanuttigh, and G. M. Cortelazzo, Time-of-Flight Cameras and Microsoft KinectTM (Springer-Verlag, 2012).
4. R. Blahut, Theory of Remote Image Formation (Cambridge University Press, 2004).
5. J. Liang, L. Gao, P. Hai, C. Li, and L. V. Wang, “Encrypted three-dimensional dynamic imaging using snapshot time-of-flight compressed ultrafast photography,” Sci. Rep. 5, 15504 (2015).
6. C. Mallet and F. Bretar, “Full-waveform topographic lidar: State-of-the-art,” ISPRS J. Photogramm. Remote Sensing 64, 1–16 (2009).
7. F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, “Low-budget transient imaging using photonic mixer devices,” ACM Transactions on Graphics (TOG) 32, 45 (2013).
8. D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt, “SRA: Fast removal of general multipath for ToF sensors,” in European Conference on Computer Vision (ECCV), (Springer, 2014), pp. 234–249.
1. Introduction

The ability to acquire 3D structure of a scene is important in many applications, such as bio-
metrics [1], terrestrial mapping [2], and gaming [3]. Time-of-flight methods for 3D acquisition
illuminate the scene and analyze the backreflected light signal [4, 5]. However, as illustrated
in Fig. 1, scenes that include partially-reflective or partially-occluding objects have complex
patterns of light being reflected at different depths even at a single pixel. For such scenes,
one can analyze multiple light returns to fully reconstruct the multiple depths present in the
field-of-view. This is known as the problem of multi-depth reconstruction from full-waveform
measurements [6].

Conventional time-of-flight imaging sensors, such as amplitude-modulated continuous-wave
(AMCW) modules, aim to reconstruct multi-depth profiles by the methods of transient imag-
ing [7–9]. For low-power and long-range 3D imaging applications, a sensitive single-photon
detector is used instead in the active imaging setup for low-light level operations [10–13]. Re-
cent advances in single-photon imaging system design allowed experimental demonstration of
low-flux time-of-flight imaging at long ranges (∼100 m) [14, 15]. Given a pulsed illumination
source, time histogramming methods of the photon detections from the backreflected pulse
waveform have been used for pixelwise reconstruction of scene depths using single-photon im-
agers [16, 17]. For example, the high sensitivity of time-correlated single-photon imaging sys-
tems was demonstrated when full photon histogram measurements were used to track pulses of
light in flight [18].

For low-flux multi-depth imaging of scene reflectors in particular, one may choose to iden-
tify the peaks in the photon histogram by brute-force search over time bins. However, since this
leads to a large processing time (polynomial in the number of time bins, with degree equal to the
number of depths), fast algorithms using parametric deconvolution or finite-rate-of-innovation
methods have been developed [19, 20]. Assuming accurate waveform measurement, the com-
pressive depth acquisition camera (CoDAC) [21] framework also exploits parametric decon-
volution, but for estimating positions of extended planar facets rather than multiple depths per
transverse location.

The previously described multi-depth imaging methods using single-photon detectors only
give accurate results when the image acquisition time is long enough that the number of pho-
ton detections is sufficiently high to form an accurate histogram. The problem of recovering
the multi-depth information is generally difficult in low-flux scenarios due to the low signal-

(a) Partially-reflecting object

(b) Partially-occluding object

Fig. 1. Examples of active imaging scenarios in which the scene response is a sum of
responses from multiple reflectors. (a) Imaging scene with a partially-reflecting object
(shown in gray dashed line). (b) Imaging scene with a partially-occluding object.
to-noise ratio of observing only a few photon detections, and extraneous background light and detector dark counts. For moderate numbers of detected photons, a statistical approach has been used to estimate the multi-depth profile by learning a mixture of Gaussians (MoG) model that interprets the photon detection data as samples from a distribution of the full-waveform observation. In the mixture model, the mode of each mixture component corresponds to a depth value of a target. Learning of the mixture model is achieved either by using the expectation-maximization (EM) algorithm [22] or the Markov chain Monte Carlo (MCMC) method [23,24]. However, MoG-based multi-depth estimation involves a non-convex cost function and generates locally optimal solutions whose accuracy is generally poor when the number of detected photons is low. Thus, existing methods are limited in their ability to recover a scene’s multi-depth information in a photon-efficient manner.

In this paper, we demonstrate a full-waveform imaging framework that accurately recovers the multi-depth profile from single-photon observations. For example, as detailed in Section 4.2, our framework was successful in accurately reconstructing the depth features of a mannequin behind a partially-reflecting medium using only 19 signal photon detections with root mean-square (RMS) depth error of 11.4 cm. Compared to the conventional MoG-based method, which gave RMS depth error of 48.7 cm, this is an improvement by a factor of 4.2. Unlike previous works, we show that the multi-depth estimation problem from single-photon observations can be reformulated as a convex optimization problem by combining the statistics of photon detection data with sparsity of multi-depth profiles. By adapting the iterative shrinkage-thresholding algorithm (ISTA) [25,26] used for robust sparse signal pursuit to our single-photon imaging setup, we accurately solve for the global optimum of the convex optimization problem to obtain a multi-depth solution from a small number of photon detections.

2. Imaging framework

2.1. Imaging setup

Figure 2 depicts the single-photon imaging setup we used to estimate the 3D structure of scene. A focused optical source, such as a laser, periodically illuminates a patch of the scene using the pulse waveform function $s(t)$ starting at time 0. Let $T_r$ be the pulse repetition period in seconds, $N$ be the total number of independent pulse illuminations made at the image pixel, and $T_p$ be the pulsewidth defined as the RMS duration of the pulse waveform. The single-photon detector, in conjunction with a time correlator, is used to time stamp individual photon detections that are generated by the backreflected waveform from the scene plus extraneous detections arising from background light and dark counts. The recorded time of a photon detection is relative to the time of the most recent pulse illumination. We raster scan the optical source over multiple pixels in the scene to recover a spatially-resolved depth profile.

2.2. Observation model at a single pixel

We first derive the relationship between the depths of multiple reflectors in the scene and our observed data of photon detections. To avoid unnecessary notation, we characterize the light transport and detection statistics for a single pixel; the same model applies at each pixel.

Let $r(t)$ be the total optical flux that is incident on the single-photon detector. Through the linearity and time invariance of optical flux transport, we can write

$$r(t) = (h * s)(t) + B,$$  \hspace{1cm} (1)$$

where $h(t)$ is the impulse response of the scene, $B$ is the constant background light flux in photons per second, and $*$ represents convolution. Then, the rate function $\lambda(t)$ that describes
Fig. 2. (Top) Full-waveform single-photon imaging setup for estimation of depths of multiple objects. In this example, a pulsed optical source illuminates a pixel of the scene that includes a partially-reflective object occluding a target of interest. The optical flux incident at the single-photon detector combines the backreflected waveform from multiple reflectors in the scene pixel with extraneous background light. (Bottom left) The photon detections, shown as spikes, are generated by the $N\lambda(t)$ pulse rate function following an inhomogeneous Poisson process. The green and blue spikes represent photon detections from the first and second reflector, respectively; the red spikes represent the unwanted photon detections from background light and dark counts. (Bottom right) Our convex optimization processing enables accurate reconstruction of multiple depths of reflectors in the scene from a small number of photon detections.
the photon detections at the single-photon detector is

\[ \lambda(t) = \eta(h_d * r)(t) + d, \]  

(2)

where \( \eta \) is the detector’s quantum efficiency, \( d \) is the detector dark count rate, and \( h_d(t) \) is the detector’s response function. For simplicity, we assume a normalized detector response function so that \( \int h_d(t) \, dt = 1 \). Substituting Eq. (1) into Eq. (2) then gives

\[ \lambda(t) = \eta(h * s)(t) + (\eta B + d), \]  

(3)

where \( s = h_d * s \) is the effective pulse waveform after accounting for the detector response. Note that because \( \eta B + d \) is constant in \( t \) and \( \int h_d(t) \, dt = 1 \), the detector response function does not appear in the second term. Hereafter, we assume that the scene is entirely within the maximum unambiguous range of the imager, so \( h(t) = 0 \) for \( t > T_r \). We also assume that \( T_p \ll T_r \), so we may avoid complicating the exposition by assuming \( (h * s)(t) = 0 \) for \( t \not\in [0, T_r) \).

A time-correlated single-photon detector records the time-of-detection of a photon within a timing resolution of \( \Delta \) seconds. We choose a pulse repetition period that is much longer than the timing resolution (\( \Delta \ll T_r \)) and assume that \( m = T_r / \Delta \) is divisible by \( \Delta \). Then, we see that \( m = T_r / \Delta \) is the number of time bins that may contain photon detections. In other words, \( m \) is the dimension of the photon count vector \( \mathbf{y} \), where \( y_k \in \mathbb{N} \) for \( k = 1, 2, \ldots, m \). (We use \( y_k \) to indicate the scalar value at the \( k \)th index of the vector \( \mathbf{y} \).) Using the probabilistic theory of photon detections [27], after \( N \) pulse illuminations at an image pixel, the \( k \)-th bin of the observed photon count histogram is distributed as

\[ y_k \sim \text{Poisson} \left( N \int_{(k-1)\Delta}^{k\Delta} \lambda(t) \, dt \right) = \text{Poisson} \left( N \eta \int_{(k-1)\Delta}^{k\Delta} (h * s)(t) \, dt + N\Delta(\eta B + d) \right). \]  

(4)

We would like to reach an approximation in which the Poisson parameter of \( y_k \) is given by the product of a known matrix and an unknown vector.

We will approximate \( (h * s)(t) \) with a sampling period of \( \Delta' \), where we have \( \Delta' = T_r / n \) for some \( n \in \mathbb{Z}^+ \). We can approximate the first term in the Poisson parameter in Eq. (4) as

\[ \begin{align*}
N \eta \int_{(k-1)\Delta}^{k\Delta} (h * s)(t) \, dt &= \int_{(k-1)\Delta}^{k\Delta} \int_0^{T_r} N \eta h(y) s(t-y) \, dy \, dt \\
&\overset{(a)}{=} \int_{(k-1)\Delta}^{k\Delta} \sum_{j=1}^{n} \int_{(j-1)\Delta'}^{j\Delta'} N \eta h(y) s(t-y) \, dy \, dt \\
&= \sum_{j=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} h(y) N \eta s(t-y) \, dy \, dt \\
&\overset{(b)}{=} \sum_{j=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} \frac{x_j}{\Delta'} S_{k,j} \, dy \, dt \\
&= \sum_{j=1}^{n} S_{k,j} x_j, \end{align*} \]  

(5a)
where (a) follows from partitioning $[0, T_r)$ into $n$ subintervals and (b) from replacing $h(y)$ and $N\eta h(y)\delta(t-y)$ by constant approximations on $(y, t) \in ([j-1]\Delta', [j]\Delta') \times ([k-1]\Delta, k\Delta)$; specifically, we define
\[
\mathbf{x}_j = \int_{(j-1)\Delta'}^{j\Delta'} h(y) \, dy, \quad \text{for } j = 1, \ldots, n, \tag{5b}
\]
\[
\mathbf{s}_{k,j} = \frac{1}{\Delta'} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} N\eta s(t-y) \, dy \, dt, \quad \text{for } k = 1, \ldots, m, \quad j = 1, \ldots, n. \tag{5c}
\]

Note that the quality of the approximation (b) will depend on the size of $\Delta'$. Finally, using the derived approximations, the observation model of Eq. (4) can be rewritten in a concise matrix-vector form as
\[
\mathbf{y}_k \sim \text{Poisson} \left( (\mathbf{Sx} + b\mathbb{1}_m)_k \right), \quad \text{for } k = 1, 2, \ldots, m, \tag{6}
\]
where $\mathbf{x}$ is an $n \times 1$ vector, $\mathbf{S}$ is an $m \times n$ matrix, and $\mathbb{1}_m$ is an $m \times 1$ vector of ones, and
\[
b = N\Delta(\eta B + d). \tag{7}
\]

2.3. Observation likelihood expressions

Our goal of multi-depth reconstruction is to accurately estimate $\mathbf{x}$ from $\mathbf{y}$. Using Eq. (6), the photon count histogram vector $\mathbf{y}$ has the probability mass function
\[
p_{\mathbf{y}}(\mathbf{y}; \mathbf{x}, \mathbf{S}, b) = \prod_{k=1}^{m} \frac{\exp{-((\mathbf{Sx} + b\mathbb{1}_m)_k)}}{\mathbf{y}_k!}. \tag{8}
\]
We can thus write the negative log-likelihood of $\mathbf{x}$ as
\[
\mathcal{L}(\mathbf{x}; \mathbf{y}, \mathbf{S}, b) = -\log p_{\mathbf{y}}(\mathbf{y}; \mathbf{x}, \mathbf{S}, b) \stackrel{\sim}{=} \sum_{k=1}^{m} \left[ (\mathbf{Sx})_k - \mathbf{y}_k \log (\mathbf{Sx} + b\mathbb{1}_m)_k \right], \tag{9}
\]
where $\stackrel{\sim}{=}$ indicates equality up to terms independent of $\mathbf{x}$. By checking the positive-semidefiniteness of the Hessian matrix of the negative log-likelihood function, it is straightforward to prove that $\mathcal{L}(\mathbf{x}; \mathbf{y}, \mathbf{S}, b)$ is convex in $\mathbf{x}$.

2.4. Characteristics of the impulse response functions of natural scenes

It has been shown that the following $K$-reflector model is effective in describing the impulse response of a natural scene with multiple reflectors [8, 23]:
\[
h(t) = \sum_{i=1}^{K} a^{(i)} \delta \left( t - 2d^{(i)}/c \right), \quad t \in [0, T_r), \tag{10}
\]
where $a^{(i)}$ and $d^{(i)}$ are respectively the reflectivity and depth values of the $i$th reflector at an image pixel. $\delta (\cdot)$ denotes the delta function, $c$ is the speed of light, and $K$ is the number of reflectors. Let us choose the indexing rule so that $d^{(1)} < d^{(2)} < \cdots < d^{(K)}$. Then, we define the minimum separation of reflector depths as
\[
d_s = \min_{i=1, \ldots, K-1} \left| d^{(i)} - d^{(i+1)} \right|. \tag{11}
\]
We observe that, assuming the $K$-reflector model and $c\Delta' / 2 < d_s$, exactly $K$ elements of $\mathbf{x}$ are nonzero and those entries have values $\{a^{(i)}\}_{i=1}^{K}$. 
3. Novel image reconstruction algorithm

The multi-depth profile can be estimated by minimizing the negative log-likelihood function, while including a signal support constraint that the number of nonzero elements in $x$ must be equal to $K$. However, the $\ell_0$-norm constraint set, which describes the set of vectors with $K$ nonzero elements, is a non-convex set. With no additional assumptions, exactly solving an optimization problem constrained to this set is computationally infeasible, since the problem is NP-hard [28]. In order to design a computationally feasible algorithm, we apply the convex relaxation whereby the $\ell_1$-norm constraint serves as a proxy for the $\ell_0$-norm [29]. Our proposed imaging framework also includes a constraint on reflectivity values being nonnegative. Thus, we obtain the multi-depth profile estimate $\hat{x}_{\text{OPT}}$ by solving the following $\ell_1$-penalized and constrained likelihood optimization problem:

$$\begin{align*}
\text{minimize} & \quad \mathcal{L}(x; y, S, b) + \tau \|x\|_1 \\
\text{subject to} & \quad x_k \geq 0, \quad k = 1, 2, \ldots, n,
\end{align*} \tag{OPT}$$

where $\tau > 0$ is the variational parameter controlling the degree of penalizing the non-sparsity of the signal. Because $\mathcal{L}(x; y, S, b)$ and the $\ell_1$-norm are both convex functions in $x$ and the nonnegative cone is a convex set, the minimization problem given in (OPT) is a convex optimization problem.

ISTA is a celebrated algorithm for rapidly solving the $\ell_1$-penalized constrained likelihood optimization problem when the data is corrupted by additive white Gaussian noise. However, instead of using a Gaussian likelihood, (OPT) is derived based on the model for single-photon observations. Thus, we modify the first step of ISTA that takes a gradient descent in the least-squares cost to one that takes a gradient descent in the negative log likelihood obtained from photon observations given in Eq. (9). We can compute the gradient of our negative log likelihood function as

$$\nabla_x \mathcal{L}(x; y, S, b) = \sum_{k=1}^m \nabla_x \left[ (Sy)_k - y_k \log (Sx + b 1_m)_k \right] = \sum_{k=1}^m \left[ (S^T)_k - \frac{y_k}{(Sx + b 1_m)_k} (S^T)_k \right]$$

$$= S^T [1_m - \text{div}(y, Sx + b 1_m)], \tag{12}$$

where we used $(S^T)_k$ to denote the $k$th column of $S^T$ and $\text{div}(\cdot, \cdot)$ represents elementwise division of the vector in the first argument by the vector in the second one. We then modify the second step of ISTA that performs a shrinkage-thresholding operation on the gradient-descent solution to include the nonnegativity constraint of scene reflectivities. Our extra nonnegativity constraint replaces the shrinkage-thresholding operation with a shrinkage-rectification operation. The shrinkage-thresholding used in ISTA and the shrinkage-rectification used in our algorithm are illustrated in Fig. 3.

Our modified ISTA algorithm is given in Algorithm 1. After solving (OPT) using Algorithm 1, we apply post-processing on $\hat{x}_{\text{OPT}}$ that sets small residual nonzero elements to zero and groups closely-neighboring nonzero elements into an average depth. The depth grouping step is to ensure maximum sparsity of the multi-depth profile, by using the assumption that a tight cluster of depth estimates originates from a single reflector in the scene. This end-to-end processing is summarized in Algorithm 2. To illustrate the role of the post-processing, Fig. 4 shows the intermediate SPISTA output and the final output of Algorithm 2 for one pixel of our experiment detailed in Section 4.2.2 (see pixel marked in experimental setup figure).

In summary, we have developed a low-flux multi-depth imaging framework that incorporates the statistics of single-photon detections with the sparsity of the multi-depth profile at a pixel to formulate a convex optimization problem in (OPT). This is unlike existing histogram-based
Algorithm 1 Single-photon iterative shrinkage-thresholding algorithm (SPISTA)

Input: $y, S, b, \tau, \hat{x}^{(0)}, \delta$
Output: $\hat{x}^{(k)}$

Initialize $k \gets 0$;

repeat
  $k \gets k + 1$;
  $\hat{x}^{(k)} \gets \hat{x}^{(k-1)} - S^T \left[ 1_m - \text{div}(y, S \hat{x}^{(k-1)} + b 1_m) \right]$; \quad \triangleright \text{iteration update}
  $\hat{x}^{(k)} \gets \max(\hat{x}^{(k)} - \tau, 0)$; \quad \triangleright \text{gradient descent}
  until $\|\hat{x}^{(k-1)} - \hat{x}^{(k)}\|_2^2 < \delta$ \quad \triangleright \text{shrinkage-rectification}

Algorithm 2 Computational multi-depth single-photon imaging

Input: $y, S, b, \tau, \hat{x}^{(0)}, \delta, \epsilon$
Output: $\hat{x}_{\text{OPT}}$, the sparse multi-depth vector

1. Convex optimization: Obtain $\hat{x}_{\text{OPT}}$ by solving (OPT) with SPISTA($y, S, b, \tau, \hat{x}^{(0)}, \delta$).
2. Residual filtering: Identify the index set $L = \{i \in \{1, 2, \ldots, n\} : (\hat{x}_{\text{OPT}})_i < \epsilon\}$, where $\epsilon$ is a small positive number. Set $(\hat{x}_{\text{OPT}})_L = 0$ to filter out residuals.
3. Depth grouping: Identify index set in which the adjacent bins of $\hat{x}_{\text{OPT}}$ are all nonnegative. Replace them with their average depth bin.

low-flux methods, such as MoG, which solve a non-convex problem directly and do not guarantee high accuracy solutions due to local minima. Our framework modifies ISTA to include accurate photodetection statistics and the nonnegativity constraint to accurately solve (OPT). Our algorithm also employs post-processing to ensure filtering of residual signals and clustering depth estimates to maximize the level of sparsity of the final multi-depth estimate.

4. Results

4.1. Simulations

4.1.1. Performance of two-path recovery

Using simulations, we first study the multi-depth estimation performance for $K = 2$, motivated by the two-Dirac recovery scenario of second-order multipath interference from reflective surfaces [30] and looking through a transparent object [31] in conventional high light-level time-
of-flight imaging. We focus on comparing two algorithms: the MoG-based estimator using a greedy histogram-data-fitting strategy and our proposed imager using convex optimization. Let \( \{d_1, d_2\} \), with \( d_1 < d_2 \), be the set of true depths at a pixel. Also, let \( \{\hat{d}_1, \hat{d}_2\} \), with \( \hat{d}_1 < \hat{d}_2 \), be the set of identified depths obtained using either the MoG method or our proposed framework. Then, we use the pulsewidth-normalized root mean-square error (NRMSE) to quantify the recovery performance of the two-path signal:

\[
\text{NRMSE}(\{d_1, d_2\}, \{\hat{d}_1, \hat{d}_2\}) = \frac{1}{cT_p/2} \sqrt{\mathbb{E}\left[\frac{1}{2} \left( (d_1 - \hat{d}_1)^2 + (d_2 - \hat{d}_2)^2 \right) \right]},
\]

such that if NRMSE is below 1, then the imager has achieved sub-pulsewidth depth accuracy. When more than two paths were estimated by the algorithm, two depth values with highest intensities were used for NRMSE computation.

Figure 5 shows Monte Carlo simulated performance results of pixelwise two-path estimation using the MoG-based method and our method. The results are presented for low \((b = 0.1)\) and high \((b = 0.5)\) background levels. For learning MoG components given photon observa-
Fig. 5. Simulated performance of MoG-based method and proposed framework in recovering signals with $K = 2$ for two different background levels. Signal photon detections are detections originating from scene response and do not include the background-light plus dark-count detections. Note that the units of NRMSE are in meters, after being normalized by the pulsewidth; 1 NRMSE equals $c T_p / 2 = 4.5$ cm. The plots also include error bars indicating the ±1 standard errors.

In this simulation, we required an average of 85 SPISTA iterations per pixel. The average per pixel processing time of Algorithm 2 was measured to be ~0.004 seconds. Our algorithm’s processing time is short because the computational time of a SPISTA update is linear in the number of depth bins $n$, since the most costly operation in SPISTA is the size-$n$ convolution, and the post-processing step only requires a linear search over $n$ bins. The average per pixel processing time of the MoG method was measured to be ~0.019 seconds. All processing was done using a laptop with Intel(R) Core(TM) i7-4500u CPU @ 1.80 GHz.

4.1.2. Resolvability

In the problem of multi-depth estimation, it is natural to ask how small the depth separation of two adjacent reflectors can be so that the proposed algorithm can still accurately resolve two reflectors instead of one. Figure 6 shows simulation results that describe how the number of
reflectors estimated by our algorithm (number of nonzero elements in $\hat{x}_{OPT}$) varies with the distance between two reflectors and the relative reflectivities of the two reflectors. We fixed the mean number of photons coming from the first target with unit reflectivity as 100 by changing $N$, the number of illumination trials, in Eq. (5c). The reflectivity of the second target was set to be one of the values in the set \{1, 1/2, 1/4, 1/8\}. Here we had $m = n = 100$, $\Delta = 1$ cm, $\delta = 10^{-2}$, $\varepsilon = 0.1$, $\tau = 0.5$, and $b = 0$. Algorithm 2 was applied to 2000 independent simulated experiments and the results in Fig. 6 show the mean number of estimated reflectors. In Fig. 6, for all relative reflectivity settings, we observe that if the distance between two reflectors is too small (around 3$cT_p$), then our algorithm falsely recognizes two reflectors as a single reflector. Even when there is a large separation between reflectors (larger than 3$cT_p$), if the ratio between target reflectivities is too small (such as 1/8), then the number of reflectors is likely to be falsely recognized as 1 by our Algorithm 2.

4.2. Experimental results

We demonstrate the performance of the proposed multi-depth imager using an experimental single-photon imaging setup. A PicoQuant LDH series pulsed laser diode with center wavelength of 640 nm, pulsewidth $T_p = 270$ ps, and repetition period $T_r = 100$ ns was used as the illumination source. We observed that the laser spot size cast on a planar object at a 1 m distance was around 1 mm, implying that the beam solid angle was around $7.9 \times 10^{-7}$ sr. A Thorlabs GVS012 two-axis galvo was used to raster scan the scene with a field-of-view of $40^\circ \times 40^\circ$. A lensless Micro Photon Devices PDM series Geiger-mode avalanche photodiode detector with quantum efficiency $\eta = 0.35$, timing jitter less than 50 ps, and dark counts per second less than $2 \times 10^4$ was used for detection of photons. A PicoQuant model HydraHarp 400 time-correlator was used to record the detection times of individual photons. Figure 7 shows a photograph of our experimental single-photon imaging setup. We injected extraneous background light using an incandescent lamp. In this paper, we chose to use the raster-scanning setup simply due to availability of equipment; one can observe that our computational framework can be applied without modification when employing an imaging setup that includes a flood illumination source and an array of single-photon detectors.

For all experiments, we obtained the pulse waveform matrix $S$ and background-light plus
dark-count value $b$ through a calibration step prior to the scene measurements. We obtained the pulse shape by projecting the laser light at the wall, which was a reflective plane 4 meters from the imager, and collecting a photon count histogram. We time-gated an interval of the photon count histogram at around $4 \times \frac{2}{c}$ seconds, such that the interval contained a clean representation of the pulse histogram. By creating a convolution matrix using the calibrated pulse waveform shape, we got a measurement of $S$ from Eq. (5c). The value of $b$ from Eq. (7) was obtained by taking a baseline measurement of the incandescent light with laser light not present in scene. Code and data are available in [32].

4.2.1. Imaging through a partially-reflective object

Figure 8 shows experimental results of imaging through a partially-reflective object, which is the multi-depth imaging scenario in Fig. 1(a), using the MoG-based and proposed multi-depth estimation methods with an average of 46 photon detections at each pixel. We used a stack of plastic sheets enclosed in a plexiglass case as the partially-reflective object, with an average reflectivity of $\sim 50\%$. Through calibration we found that the probabilities of a photon coming from the scatterer, the scene behind the scatterer, and background light or dark counts were equal to 0.44, 0.42, and 0.14, respectively. These numbers were calibrated using a single histogram with 10000 photons gathered from all pixels, where each pixel contributed a single photon to the aggregate histogram. Thus, the number of photons that originated from the scene-of-interest behind the scatterer is calculated to be $46 \times 0.42 \approx 19$. The raster-scanning resolution was set to be $100 \times 100$ in this experiment.

We see that the existing MoG-based method fails to recover useful depth features of the mannequin, but our method successfully does so. For example, in the side view of the reconstructed depth, the result from our method differentiates the longitudinal locations of the face and the torso of the mannequin, unlike the result from the MoG method. We were able to form a ground truth depth map of the mannequin by using the log-matched filtering solution [27] on a larger dataset of 500 photons per pixel, after time-gating the photon arrivals at around 2.6 ns such that...
remaining photons only describe the mannequin scene placed at around 4 m. With respect to this ground truth, the conventional MoG solution gave 48.7 cm of RMS depth error and ours gave 11.4 cm. The RMS depth error was computed as the square root of the average of squared depth errors over all pixels. Our framework thus gave an improvement in reducing erroneous pixel depth by a factor of 4.2, compared to the MoG method, for the task of imaging a scene with partially-reflective object.

In this experiment of imaging through a partially-reflecting object, we required an average of 98 SPISTA iterations per pixel. The average per pixel processing time of Algorithm 2 was measured to be \( \sim 0.036 \) seconds and its total processing time for the spatially-resolved multi-depth reconstruction was \( \sim 6 \) minutes. The average per pixel processing time of the MoG method was measured to be \( \sim 0.003 \) seconds and its total processing time was \( \sim 30 \) seconds.

### 4.2.2. Imaging a partially-occluding object

For experimental validation of multi-depth estimation for scenes with partially-occluding objects, we used a photon detection dataset that was collected by Dheera Venkatraman for the first-photon imaging project [11] with the same raster-scanning setup. Our experimental scene consisted of a sunflower and the background wall as shown in Fig. 9. This experiment models the multi-depth imaging scenario in Fig. 1(b). Here we have a higher raster-scanning resolution of \( 300 \times 300 \), such that many pixels are at the depth boundaries of the two reflective objects: the sunflower and the wall. There are multiple returns at such pixels, where the sunflower petals partially occlude the wall behind it. This artifact is also known as one of mixed pixels in the context of time-of-flight imaging [33]. In our data, the probabilities of a photon originating from the scene and from background light or dark counts are 0.8 and 0.2, respectively. These numbers were calibrated using a single histogram with 90000 photons gathered from all pixels, where each pixel contributed to a single photon to the aggregate histogram.

Figure 10 shows how the proposed imager compares to the MoG estimator for the sunflower
and wall scene. The mean number of photon detections over all pixels was measured to be 26 for this experiment. In the figure, we observe that our proposed multi-depth imager successfully distinguishes the sunflower’s petals from its leaves and the wall behind it, even though there exist mixed-depth pixels at boundaries and high background light plus dark counts. This is most visible in the side view, where we see that the noisy depth estimates present in the MoG results are absent when using our method. Similar to the previous imaging experiment, we were able to form a ground truth depth map of the sunflower by using the log-matched filtering solution on a larger dataset of 500 photons per pixel, after time-gating the photon arrivals at around 1.6 ns such that remaining photons only describe the sunflower placed at around 2.5 m, and not the wall behind it. Then, we computed that while the conventional MoG solution gave 46.5 cm of RMS depth error, ours gave 4.3 cm. Our framework thus gave an improvement in RMS depth error by a factor of 10.8, compared to the MoG method, for the task of imaging a scene with partially-occluded object.

In this experiment of imaging a partially-occluding object, we required an average of 7 SPISTA iterations per pixel. The average per pixel processing time of Algorithm 2 was measured to be ~0.0014 seconds and its total processing time for the spatially-resolved multi-depth reconstruction was ~2 minutes. The average per pixel processing time of the MoG method was measured to be ~0.0057 seconds and its total processing time was ~8.5 minutes.

5. Conclusion and future work

In this paper, we presented a robust framework for reconstruction of multi-depth profiles of a scene using a single-photon detector. Our novel imaging method accurately models the single-
Fig. 10. Experimental results of depth reconstruction of sunflower occluding a wall, given that our imaging setup is at $z = 0$. Using our imaging framework, the mixed-pixel artifacts at the depth boundary of the flower and background light plus dark count noise are suppressed. Our multi-depth results were generated using the parameters $\tau = 0.2$, $\delta = 10^{-4}$, $\varepsilon = 0.1$, and $\hat{x}(0) = S\hat{y}$.

photon detection statistics from multiple reflectors in the scene, while exploiting the fact that multi-depth profiles can be expressed as sparse signals. Using our signal model the multi-depth estimation problem is a sparse deconvolution problem. We designed an algorithm inspired by ISTA that reaches the globally optimal solution of the convex relaxation of the sparse deconvolution problem with high computational efficiency, demonstrated by the sub-second per-pixel processing time in our simulations and experiments. Using both simulations and experiments for scenes including partially-reflecting and partially-occluding objects, we demonstrated that our imaging framework outperforms the existing MoG-based method for multi-depth estimation in the presence of background light and dark counts.

Unlike the parameter-free MoG-based multi-depth imaging method, our framework introduces a number of free parameters, such as the regularization parameter. For our experiments, we used the heuristic of choosing the parameters based on the calibrated background level. In practice, when the signal-to-background ratio varies over multiple imaging experiments, one can employ cross-validation techniques to learn the scalar parameters from multiple experiments [34].

For future work, it is of interest to study how applying other post-processing methods, such as 3D point cloud filtering, can improve multi-depth recovery performance. Also, optoelectronic techniques, such as range gating and narrowband optical filtering, can be incorporated to reject background counts at the data acquisition level to have a more accurate multi-depth reconstruction.

Acknowledgments

This material is based upon work supported in part by a Samsung Scholarship, the US National Science Foundation under Grant No. 1422034, and the MIT Lincoln Laboratory Advanced Concepts Committee. We thank Dheera Venkatraman for his assistance with the experiments.